Algebras with three anticommuting elements. I. Spinors and quaternions.

Yehiel Ilamed
Soreq Nuclear Research Center, Yavne, Israel
Nikos Salingaros
Physics Department, The University of Crete, Iraklion, Crete, Greece
Physics Department, University of Massachusetts, Boston, Massachusetts 02125

(Received 13 February 1981; accepted for publication 29 May 1981)

A general construction of alternative algebras with three anticommuting elements and a unit is given. As an exhaustive result over the real and complex fields, we obtain the Clifford algebras H (quaternions), N_1 (dihedral Clifford algebra which is related to real 2-spinors), and S_1 (algebra of Pauli matrices which is related to complex 2-spinors). What is important is that the algebras N_1 and S_1 possess inverses everywhere except on a region akin to the light cone of the Minkowski space, while the quaternion algebra H has inverses everywhere except at the zero element. We discuss the reasons why the three algebras N_1, H, and S_1 are so difficult to distinguish in the representation space of 2×2 complex matrices.

PACS numbers: 02.10.Tq

I. INTRODUCTION

The aim of this paper is to discuss the structure of all possible algebras with three anticommuting elements over the real and complex fields $\mathbb{R}$ and $\mathbb{C}$ that can arise in physical descriptions.

We present a general construction of alternative algebras with three anticommuting elements and a unit. Associativity is a result of the construction, and three nonisomorphic algebras are obtained. This is, moreover, an exhaustive result over the fields $\mathbb{R}$ and $\mathbb{C}$. We are able to utilize a previously published classification of all Clifford algebras,[1,2] in order to identify the three algebras obtained by the general construction of this paper with known Clifford algebras. They are (i) the quaternions H, (ii) the dihedral Clifford algebra N_1, which is related to real 2-spinors, and (iii) the algebra of Pauli matrices S_1, which is related to complex 2-spinors.

Of additional interest is the demonstration that all three algebras possess unique two-sided inverses. The algebras N_1 and S_1 possess inverses for all elements except those on a region akin to the light cone of the Minkowski space. The quaternion algebra H is free of such a singular region and is therefore a division algebra, in the usual sense.[3,4] This important property is a feature of all Clifford algebras up to dimension eight (see Ref. 5). In order to distinguish between the divisibility properties of the Clifford algebras up to dimension eight and those of the larger Clifford algebras which possess the more restricted "sectional divisibility" (see Refs. 1 and 2), it is convenient to introduce a special terminology. Those algebras which possess well-defined inverses except on a region akin to the Minkowski light cone can be referred to as 'singular division algebras"; those algebras that do not possess this singular region are simply "division algebras" in the traditional sense.

These results are fully consistent with the Frobenius theorem,[3,4] which presupposes the absence of singular regions such as those discussed here. In addition, the algebra S_1 is an algebra over the complex field $\mathbb{C}$, and the Frobenius theorem does not apply since it classifies division algebras over the real field $\mathbb{R}$.

The algebraic construction is detailed in Secs. II and III. In Sec. IV we use the classification of finite groups corresponding to Clifford algebras given in Refs. 1, 2 in order to identify the algebras obtained. The structure is summarized in the multiplication tables (Table II). In Sec. V, we explicitly construct the inverses in each algebra. For the case of S_1, we employ the realization of S_1 as a Clifford algebra that was given in Ref. 1.

From group theoretical methods we obtain the relationship $S_1 = H \otimes \mathbb{C} = N_1 \otimes \mathbb{C}$, which demonstrates the isomorphism between the algebra of Pauli matrices and the algebra of complex quaternions. Using this relation as a starting point, and following an alternate procedure from that outlined in this paper, one can obtain the three algebras N_1, H, S_1 as subalgebras of the complex quaternions.

In Sec. VI, the matrix representations of the algebras H, N_1, and S_1 are given in terms of the Pauli matrices (Table III), and the relationship of the algebra N_1 to real 2-spinors is indicated explicitly. The relation $S_1 = N_1 \otimes \mathbb{C}$ shows the connection between complex and real 2-spinors in a very simple way. A point worth noting is that the three algebras H, N_1, and S_1 are almost impossible to distinguish in the representation space of 2×2 complex matrices.

II. THE ALGEBRAIC CONSTRUCTION

Consider a set of three elements and a scalar unit, $\{1, e_1, e_2, e_3\}$. These elements are defined to anticommute. Then, the general multiplication of these elements can be written as a linear combination of the other elements, with constant scalar coefficients. (The following analysis will determine those cases when they are real and those when they are complex.) The multiplication of the elements can be written in the most general manner using nine scalar coefficients, α_i, β_i, γ_i, $i = 1, 2, 3$, as follows:

 0022-2488/81/102091-05$01.00

$$e_1 e_2 = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3, \quad (1a)$$

$$e_2 e_3 = \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3, \quad (1b)$$

$$e_3 e_1 = \gamma_1 e_1 + \gamma_2 e_2 + \gamma_3 e_3. \quad (1c)$$

We define a scalar square for each element,

$$e_1 e_1 = a_1, \quad e_2 e_2 = a_2, \quad e_3 e_3 = a_3, \quad (2)$$

where a_1, a_2, and a_3 are each equal to either $+1$ or -1. We leave them in this general form for the moment, since the distinct cases will in fact determine the structure of the corresponding algebras.

We impose left and right alternativity as the only condition other than anticommutativity. This condition enables us to obtain relations between the scalar coefficients $\alpha_i, \beta_i, \gamma_i$ in the following manner. Using left alternativity and (2), we have the identities

$$e_1 (e_1 e_2) = (e_1 e_1) e_2 = a_1 e_2, \quad (3a)$$

$$e_2 (e_2 e_3) = (e_2 e_2) e_3 = a_2 e_3, \quad (3b)$$

$$e_3 (e_3 e_1) = (e_3 e_3) e_1 = a_3 e_1. \quad (3c)$$

Similarly, right alternativity gives the identities

$$(e_1 e_2) e_2 = a_2 e_1, \quad (4a)$$

$$(e_2 e_3) e_3 = a_3 e_2, \quad (4b)$$

$$(e_3 e_1) e_1 = a_1 e_3. \quad (4c)$$

Our procedure is to rewrite the alternativity conditions (3) and (4) in terms of the general prescription (1). In this way we obtain relations between the coefficients α_i, β_i, and γ_i.

Left multiplying (1a) by e_1 and using (3a) and (2), we obtain the following equation:

$$a_1 e_2 = \alpha_1 a_1 + \alpha_2 e_1 e_2 + \alpha_3 e_1 e_3. \quad (5)$$

Substituting (1a) and (1c) on the right-hand side of (5) and using anticommutation gives us a linear equation involving only the e_i with scalar coefficients:

$$a_1 e_2 = \alpha_1 a_1 + (\alpha_1 \alpha_2 - \alpha_3 \gamma_1) e_1 + (\alpha_2^2 - \alpha_3 \gamma_2) e_2 + (\alpha_2 \alpha_3 - \alpha_3 \gamma_3) e_3. \quad (6)$$

From Eq. (6), we can find a set of relations between the coefficients α_i, β_i, and γ_i by equating scalar coefficients of each distinct e_1. These are

$$\alpha_1 a_1 = 0, \quad \alpha_1 \alpha_2 - \alpha_3 \gamma_1 = 0,$$
$$\alpha_2^2 - \alpha_3 \gamma_2 - a_1 = 0, \quad \alpha_2 \alpha_3 - \alpha_3 \gamma_3 = 0. \quad (7)$$

In a similar manner, all the other equations in (1) can be treated using first left and then right alternativity, to obtain six distinct sets of equations relating the scalar coefficients α_i, β_i, and γ_i. These are collected in Table I.

III. ANALYSIS

From Table I, we see that, since $a_i \neq 0$, then

$$\alpha_1 = \alpha_2 = \beta_2 = \beta_3 = \gamma_1 = \gamma_3 = 0. \quad (8)$$

Hence, we are left with the following very simple algebraic structure

$$e_1 e_2 = \alpha_3 e_3, \quad (9a)$$

$$e_2 e_3 = \beta_1 e_1, \quad (9b)$$

$$e_3 e_1 = \gamma_2 e_2. \quad (9c)$$

TABLE I. Equations between the scalar coefficients.

$\alpha_1 a_1 = 0$	$\alpha_2 a_2 = 0$
$\alpha_1 \alpha_2 - \alpha_3 \gamma_1 = 0$	$\alpha_1^2 - \alpha_3 \beta_1 = a_2$
$\alpha_2^2 - \alpha_3 \gamma_2 = a_1$	$\alpha_1 \alpha_2 - \alpha_3 \beta_2 = 0$
$\alpha_2 \alpha_3 - \alpha_3 \gamma_3 = 0$	$\alpha_1 \alpha_3 - \alpha_3 \beta_3 = 0$
$\beta_2 a_2 = 0$	$\beta_3 a_3 = 0$
$-\beta_1 \alpha_1 + \beta_3 \beta_1 = 0$	$-\beta_1 \gamma_1 + \beta_1 \beta_2 = 0$
$\beta_2 \beta_3 - \beta_1 a_2 = 0$	$\beta_2^2 - \beta_1 \gamma_2 = a_3$
$\beta_3^2 - \alpha_3 \beta_1 = a_2$	$\beta_2 \beta_3 - \beta_1 \gamma_3 = 0$
$\gamma_3 a_3 = 0$	$\gamma_1 a_1 = 0$
$\gamma_1^2 - \gamma_2 \beta_1 = a_3$	$\gamma_1 \gamma_3 - \alpha_1 \gamma_2 = 0$
$\gamma_1 \gamma_2 - \beta_2 \gamma_2 = 0$	$\gamma_2 \gamma_3 - \alpha_2 \gamma_2 = 0$
$\gamma_1 \gamma_3 - \gamma_2 \beta_3 = 0$	$\gamma_3^2 - \gamma_2 \alpha_3 = a_1$

From Table I, substituting (8), the remaining equations are

$$a_1 = -\alpha_3 \gamma_2,$$
$$a_2 = -\alpha_3 \beta_1, \quad (10)$$
$$a_3 = -\gamma_2 \beta_1.$$

The a_i can be expressed by eliminating the α_3, β_1, and γ_2 constants. The a_i are equal to either $+1$ or -1. Since they can be permuted into each other, it is sufficient to consider the following four distinct cases:

$$a_1 = 1, \quad a_2 = 1, \quad a_3 = 1, \quad (11a)$$

$$a_1 = 1, \quad a_2 = 1, \quad a_3 = -1, \quad (11b)$$

$$a_1 = 1, \quad a_2 = -1, \quad a_3 = -1, \quad (11c)$$

$$a_1 = -1, \quad a_2 = -1, \quad a_3 = -1. \quad (11d)$$

The four separate cases give the following values for the multiplication coefficients α_3, β_1, and γ_2; using Eq. (10),

$$\text{case (a):} \quad \alpha_3 = \beta_1 = \gamma_2 = i \ \text{or} \ -i, \quad (12a)$$

$$\text{case (b):} \quad -\alpha_3 = \beta_1 = \gamma_2 = 1 \ \text{or} \ -1, \quad (12b)$$

$$\text{case (c):} \quad \alpha_3 = \gamma_2 = -\beta_1 = i \ \text{or} \ -i, \quad (12c)$$

$$\text{case (d):} \quad \alpha_3 = \beta_1 = \gamma_2 = 1 \ \text{or} \ -1. \quad (12d)$$

We have been forced in cases (12a) and (12c) to find solutions of Eq. (10) over the complex field. This is very important in the analysis that follows, since it in fact doubles the size of the corresponding algebra.

An interesting point is the fact that the four algebras obtained are associative. This can be checked directly using (9) and (12). Associativity was not assumed in the analysis; hence, it appears as a consequence of our construction.

IV. IDENTIFICATION WITH CLIFFORD ALGEBRAS

In general, the structure of an associative algebra is determined by the structure of the corresponding finite group defined by its generators.[1,2] For small dimensions, the order of the group elements is sufficient to define the algebra.

A construction and classification of all Clifford algebras in terms of their corresponding finite groups was given in Refs. 1 and 2. We can apply those results in order to identify the associative algebras obtained above with Clifford algebras. Note that all associative algebras are not necessarily Clifford algebras; those under discussion happen to be.

We proceed to find the order of the elements in each of the algebras defined by (9) and (12). The order of each element is defined as the power to which that element has to be raised to obtain the unit one. For the group structure, we must consider the negative elements as separate group elements, even though this distinction is not necessary when considering the algebra.

For example, we compute the group order of case (12d). We have, from (11d),

$$(e_k)^2 = -1 \Rightarrow (e_k)^4 = 1, \quad k = 1, 2, 3. \tag{13}$$

This gives us three elements of order 4. By including $\{-1, -e_1, -e_2, -e_3\}$ separately, we have one element of order 2, and three more elements of order 4. In total, we have the unit (order 1), the negative of the unit (order 2), and the three generators plus their negatives (order 4). This can be written as

$$\text{order} = 1^1 2^1 4^6 \quad \text{or} \quad (1, 1, 6). \tag{14}$$

The orders of the groups defined by the other three cases (11) and (12) are calculated in the same manner. It is important to observe that in cases (12a) and (12c), where we have been forced to go to the complex number field, the order of the group is doubled, since all combinations of the elements with i must be counted separately.

The orders of the finite groups defined by the multiplication (9) in the four distinct cases (12a), (12b), (12c), (12d) are presented below:

$$\text{case (a):} \quad \text{order} = (1, 7, 8), \tag{15a}$$

$$\text{case (b):} \quad \text{order} = (1, 5, 2), \tag{15b}$$

$$\text{case (c):} \quad \text{order} = (1, 7, 8), \tag{15c}$$

$$\text{case (d):} \quad \text{order} = (1, 1, 6). \tag{15d}$$

Note that case (15a) is isomorphic to case (15c), yet the multiplication table is not identical (Table II).

In Refs. 1 and 2, the finite groups corresponding to each Clifford algebra were constructed and classified. We can use a table giving the orders of each group (Table IV[1,2]) to identify the algebras obtained in this paper with the Clifford algebras constructed in Refs 1 and 2 simply by comparing the orders. Since these algebras are of small order, equivalence of the orders implies an isomorphism between the algebras.

A comparison with the tables of the orders of the "vee-groups" shows that the group defined by (14) with order (1, 1,6) is isomorphic to the quaternion group Q_4. Therefore, the algebra of example (14) is isomorphic to the quaternion algebra H.

In the same manner, by comparing cases (15a) and (15b) to the group orders in Refs. 1 and 2, we can identify the other two groups with the known finite groups:

$$(1, 1, 6) = \text{quaternion group } Q_4, \tag{16a}$$

$$(1, 5, 2) = \text{dihedral group } D_4, \tag{16b}$$

$$(1, 7, 8) = \text{Pauli spinor group} = Z_4 \otimes Z_2 \otimes Z_2. \tag{16c}$$

The Clifford algebras corresponding to the finite groups (16) are obtained from Refs. 1 and 2, and are the following:

$$Q_4 \leftrightarrow \text{quaternion algebra } H, \tag{17a}$$

$$D_4 \leftrightarrow \text{dihedral Clifford algebra } N_1, \tag{17b}$$

$$Z_4 \otimes Z_2 \otimes Z_2 \leftrightarrow \text{algebra of Pauli matrices } S_1. \tag{17c}$$

The results of the preceding discussion can now be summarized by giving the multiplication tables of the three algebras. From equations (2), (9), (11), (12), (15), and (16), we can construct the multiplication tables, which are presented in Table II.

We recall the relation: group = algebra $\otimes Z_2$, where Z_2 is the cyclic group of order 2 composed of the two elements $\{1, -1\}$.[1,2] Hence, the *group* multiplication tables can be obtained from Table II after including the negative bases as distinct elements.

The quaternions H and the dihedral Clifford algebra N_1 are both of order 4 over $\mathbb{R}$; the algebra of Pauli materices S_1 is of order 4 over $\mathbb{C}$ (or as a Clifford algebra of order 8 over $\mathbb{R}$).[1,2,5]

This completes the identification of all the algebras constructed by the general prescription (1).

TABLE II. Multiplication tables of the algebras N_1, H, and S_1.

			e_1	e_2	e_3
Quaternion algebra H	(d)	e_1	-1	e_3	$-e_2$
		e_2	$-e_3$	-1	e_1
		e_3	e_2	$-e_1$	-1
			e_1	e_2	e_3
Clifford algebra N_1	(b)	e_1	1	$-e_3$	$-e_2$
		e_2	e_3	1	e_1
		e_3	e_2	$-e_1$	-1
			e_1	e_2	e_3
Algebra of Pauli Matrices S_1	(a)	e_1	1	ie_3	$-ie_2$
		e_2	$-ie_3$	1	ie_1
		e_3	ie_2	$-ie_1$	1
			e_1	e_2	e_3
	(c)	e_1	1	ie_3	$-ie_2$
		e_2	$-ie_3$	-1	$-ie_1$
		e_3	ie_2	ie_1	-1

V. DIVISIBILITY AND NORMS

It is important to stress that we have assumed neither normality nor divisibility in our construction. Therefore, a key result of this paper is the demonstration that all three algebras obtained possess two-sided inverses. This is shown by explicitly constructing the inverse elements in each case.

The Frobenius theorem[3,4] gives all associative division algebras over $\mathbb{R}$ without divisors of zero (except for the single point corresponding to the zero element) as the algebras $\mathbb{R}$, $\mathbb{C}$, and H. The algebra N_1 circumvents this theorem since N_1 has a string of singular inverses, i.e., divisors of zero.

The inverse of an element in the algebra N_1 is easily obtained. Any element in N_1 has four real components; we denote the expansion of the three components on the algebraic bases by the familiar vector notation:

$$u = u^0 + u^1 e_1 + u^2 e_2 + u^3 e_3 = u^0 + \mathbf{u}. \tag{18a}$$

Define a conjugate element $\tilde{u}$ as follows:

$$\tilde{u} = u^0 - u^1 e_1 - u^2 e_2 - u^3 e_3 = u^0 - \mathbf{u}. \tag{18b}$$

A scalar product can be defined as the algebraic product (denoted by the symbol $\vee$) of u and its conjugate,

$$u \vee \tilde{u} = \tilde{u} \vee u = (u^0)^2 - (u^1)^2 - (u^2)^2 + (u^3)^2. \tag{19}$$

This scalar is not positive definite; hence one cannot define a real norm in the usual manner, as $\|u\| = (u \vee \tilde{u})^{1/2}$.

The inverse of u is simply obtained from (19) as

$$u^{-1} = \tilde{u}/(\tilde{u} \vee u). \tag{20}$$

This can be verified directly. The inverse (20) possesses a singularity at the point where $\tilde{u} \vee u = 0$, which corresponds to the null cone of the vector u. This is a result of the noncompactness of the metric induced by the conjugation (18b). This type of singularity is a common feature of field theory in Minkowski spacetime, and poses no serious calculational problem.

The algebra S_1 is also a singular division algebra. This algebra was constructed as a Clifford algebra of order 8 over $\mathbb{R}$ in Ref. 1, and the products of fields were given. The connection with this paper is obtained by identifying the complex unit $i = \sqrt{-1}$ in $\mathbb{C}$ with the volume element in Euclidean 3-space $\omega^3 = dx^1 \wedge dx^2 \wedge dx^3$. Recall from Ref. 1 that ω^3 commutes with all elements in S_1 and has square equal to -1. S_1 is therefore generated by the elements $\{1, e_k, ie_k, i\}$, $k = 1, 2, 3$, and has center equal to $\mathbb{C}$.[2] Utilizing these results here, we can explicitly give the inverse of every element α in S_1. Define a conjugate $\tilde{\alpha}$ as follows:

$$\alpha = u^0 + \mathbf{u} + i\mathbf{v} + iv^0, \tag{21a}$$

$$\tilde{\alpha} = u^0 - \mathbf{u} - i\mathbf{v} + iv^0. \tag{21b}$$

A complex scalar product can be defined as

$$\tilde{\alpha} \vee \alpha = [(u^0)^2 - |\mathbf{u}|^2 + |\mathbf{v}|^2 - (v^0)^2] + 2i(u^0 v^0 - (\mathbf{u}\cdot\mathbf{v})). \tag{22}$$

It is obvious that we cannot define a real norm in this algebra.

The inverse of each element α is easily obtained by finding the inverse of the complex scalar product (22). Denote by x and y respectively, the real and complex parts of (22):

$$x = (u^0)^2 - |\mathbf{u}|^2 + |\mathbf{v}|^2 - (v^0)^2,$$

$$y = 2[u^0 v^0 - (\mathbf{u}\cdot\mathbf{v})]. \tag{23}$$

The inverse of (22) is simply the inverse of a complex number,

$$\frac{1}{\tilde{\alpha} \vee \alpha} = \frac{1}{x + iy} = \frac{x - iy}{x^2 + y^2}. \tag{24}$$

Hence, the inverse of any element α in S_1 is just

$$\alpha^{-1} = \frac{\tilde{\alpha}}{\tilde{\alpha} \vee \alpha} = \frac{\tilde{\alpha}(x - iy)}{x^2 + y^2}. \tag{25}$$

The identity $\tilde{\alpha} \vee \alpha = \alpha \vee \tilde{\alpha}$ ensures that this inverse is two-sided.

For completeness, we recall the well-known corresponding results for the quaternion algebra H. Equations (18a) and (18b) are the same, and expression (20) for the inverse is valid, with the difference that the scalar product is now positive-definite and is given by

$$\tilde{u} \vee u = (u^0)^2 + (u^1)^2 + (u^2)^2 + (u^3)^2 = (u^0)^2 + |\mathbf{u}|^2. \tag{26}$$

Hence, in the case of the quaternion algebra H, one can indeed define a real norm as $\|u\| = (\tilde{u} \vee u)^{1/2}$.

On the question of normed algebras,[4,6,7], it is important to note that the expression $\|u\|^2 = \tilde{u} \vee u$ satisfies the normed identity $\|u\|^2 \|v\|^2 = \|\tilde{u} \vee v\|^2$, in both the algebras N_1 and H. In the case of H, $\|u\|$ is a real norm (26), while, in the case of N_1, $\|u\|$ is either real or purely imaginary, (19). (On this point, see the Conclusion.)

VI. MATRIX REPRESENTATIONS

We include a discussion of the matrix representations of the algebras H, N_1, and S_1 for the important reason that it is nearly impossible to distinguish these algebras in representation space. From the multiplication tables (Table II) it is easy to find representations corresponding to the four cases (a), (b), (c), and (d) in terms of the familiar Pauli matrices. These are given in Table III:

$$\tau_1, \tau_2, \tau_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{27}$$

The representation space of the algebras of interest has been determined in Refs. 8 and 9 in the manner outlined in Ref. 10. Recalling those results for the algebras N_1 and S_1, we have

$$N_1 \leftrightarrow \mathbb{R}(2), \quad 2\times 2 \text{ real matrices}, \tag{28a}$$

$$S_1 \leftrightarrow \mathbb{C}(2), \quad 2\times 2 \text{ complex matrices}. \tag{28b}$$

The reality of the representation space of N_1 is of fundamental nature (see Table III). The case of S_1 is, of course, well known. Any complex representation of H spans only one-half of the space $\mathbb{C}(2)$, as can be verified directly. We note that any other matrix representations of these algebras are related via a similarity transformation to those given here. This follows from the universality of the Clifford algebras.[8]

The relationship between the algebras H, N_1, and S_1 is obtained by relating their corresponding group orders (16). By multiplying the group orders following the procedure in Refs. 1 and 2, we obtain the following relationships. Here, Z_4 is the complex group, isomorphic to the cyclic group of order 4,

$$Q_4 \otimes Z_4 = D_4 \otimes Z_4 = Z_4 \otimes Z_2 \otimes Z_2 \otimes Z_2. \tag{29}$$

Recalling the discussion at the end of Sec. IV, we can obtain from (17) the following relation between the algebras:

$$H \otimes \mathbb{C} = N_1 \otimes \mathbb{C} = S_1. \tag{30}$$

This simple relationship is confirmed by the matrix re-

TABLE III. Matrix representations of the algebras N_1, H, and S_1, in terms of the Pauli matrices.

Basis	H	N_1	S_1 case (a)	S_1 case (c)
e_1	$i\tau_1$	τ_1	τ_1	τ_1
e_2	$i\tau_2$	τ_3	τ_2	$i\tau_2$
e_3	$-i\tau_3$	$i\tau_2$	τ_3	$i\tau_3$

presentations (Table III).

We now proceed with an outline of the relation between the algebra N_1 and the formalism of 2-spinors used in physics.[11-14] Recall from Refs. 1 and 2 the geometrical properties of N_1 as a Clifford algebra. The product of two vector fields u and v in two-dimensional space $A^{2,0} = N_1$ is

$$u \vee v = (u^1 e_1 + u^2 e_2) \vee (v^1 e_1 + v^2 e_2)$$
$$= (u^1 v^1 + u^2 v^2) + e_3(u^1 v^2 - u^2 v^1) \tag{31a}$$

or

$$u \vee v = (u, v) + e_3(u \times v). \tag{31b}$$

The antisymmetric part of the product is used in defining the invariant product of 2-spinors as

$$(u \times v) = \sum_{A,B=1}^{2} u^A v^B \epsilon^{AB} = \sum_{A=1}^{2} u^A v_A, \quad v_A = \sum_{B=1}^{2} v^B \epsilon^{AB}. \tag{32}$$

Here, ϵ^{AB} is the symplectic matrix representing e_3, $\epsilon^{AB} = (i\tau_2)^{AB}$ (Table III). A spinor rotation by a real parameter θ is described as follows:

$$u' = \exp(\theta e_3) u = (\cos\theta + e_3 \sin\theta)\mathrm{u}. \tag{33}$$

The rotation operator in the matrix representation is simply

$$\exp(\theta e_3) \leftrightarrow \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{34}$$

which has determinant equal to $+1$. Alternately, one can describe Lorentz transformations of the vector field w in two-dimensional spacetime $A^{1,1} = N_1$ as

$$w = w^2 e_2 + w^3 e_3, \quad (w)^2 = (w^2)^2 - (w^3)^2, \tag{35a}$$

$$w' = \exp(\theta e_1) w = (\cosh\theta + e_1 \sinh\theta) w. \tag{35b}$$

The transformation operator in the matrix representation again has determinant 1, as is required in the Lorentz group:

$$\exp(\theta e_1) \leftrightarrow \begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix}. \tag{36}$$

We have used the identity that the Clifford algebras $A^{2,0}$ and $A^{1,1}$ are both isomorphic to N_1.[1,2,5]

For the construction of higher rank spinors as irreducible representations of the rotation and Lorentz groups and their relationship to Clifford algebras, see Refs. 11 and 14.

This completes the discussion on the matrix representations of the algebras H, N_1, and S_1, and their relation to the formalism of 2-spinors.

VII. CONCLUSION

We have provided a simple method of generating algebras with three anticommuting elements over the fields $\mathbb{R}$ and $\mathbb{C}$. As an exhaustive result, we obtained the Clifford algebras: (i) the quaternions H, (ii) the dihedral Clifford algebra N_1, which is related to real 2-spinors, and (iii) the algebra of Pauli matrices S_1, which is related to complex 2-spinors. The algebras H and N_1 are of dimension 4 over $\mathbb{R}$, while the algebra S_1 is of dimension 4 over $\mathbb{C}$. [We note that the algebra N_1 is often referred to in terms of its representation space as $\mathbb{R}(2)$.]

The three algebras obtained were shown to possess inverses which were explicitly constructed. The algebra N_1 circumvents the Frobenius theorem on real division algebras,[3,4] since N_1 possesses singular inverses other than that of the zero element. Because those singular inverses lie just on a region akin to the light cone of the Minkowski space, we can treat the algebra N_1 as a division algebra everywhere outside this singular line.

Of the three algebras N_1, H, and S_1, only the quaternion algebra H is a normed algebra, in keeping with the Hurwitz–Albert theorem.[4,6,7] However, it is important to note that all elements in the algebra N_1 satisfy the normed relation $\|u\|\,\|v\| = \|\tilde{u} \vee v\|$, even though $\|u\|^2 = (u^0)^2 - (u^1)^2 - (u^2)^2 + (u^3)^2$ does not define a norm in the usual sense. The familiarity of working in noncompact spaces which has developed since the classic work of Frobenius[3] and Hurwitz[6] allows us to consider this, along with the divisibility properties, as natural extensions of the traditional algebraic properties.

Via group-theoretical methods, we determined the relationship $S_1 = H \otimes \mathbb{C} = N_1 \otimes \mathbb{C}$, which demonstrates that these algebras can be considered as subalgebras of the complex Quaternions. Also, the connection of the algebra N_1 to the formalism of real 2-spinors was explicitly outlined, and representations of the two-dimensional rotation and Lorentz groups were given.

We believe that this discussion clarifies the structure and properties of real and complex algebras with three anticommuting elements which can arise in physics.

[1] N. Salingaros, in *Proceedings of the Second Workshop on Lie-Admissible Formulations*, Harvard University, 1979, Hadronic J. **3**, 339 (1979).
[2] N. Salingaros, J. Math. Phys. **22**, 226 (1981).
[3] F. G. Frobenius, J. Reine Angew. Math. **84**, 59 (1877).
[4] A. G. Kurosh, *Lectures on General Algebra* (Chelsea, New York, 1963).
[5] N. Salingaros, "Algebras With Three Anticommuting Elements. II," J. Math. Phys. **22**, 2096 (1981).
[6] A. Hurwitz, Nachr. Ges. Wiss. Göttingen, 309 (1898).
[7] C. W. Curtis, in *Studies in Modern Algebra*, edited by A. A. Albert (Prentice-Hall, Englewood Cliffs, N. J., 1963).
[8] M. F. Atiyah, R. Bott, and A. Shapiro, Topology **3**, (suppl. 1), 3 (1964).
[9] M. Karoubi, *K-Theory* (Springer-Verlag, Berlin, 1979).
[10] N. Salingaros, in *Proceedings of the Third Workshop on Lie-Admissible Formulations*, University of Massachusetts at Boston, 1980, Hadronic J. **4**, 949 (1981).
[11] E. Cartan, *The Theory of Spinors* (Hermann, Paris, 1966).
[12] B. L. van der Waerden, Nachr. Ges. Wiss. Göttingen, 100 (1929).
[13] O. Laporte and G. Uhlenbeck, Phys. Rev. **37**, 1380 (1931).
[14] R. Brauer and H. Weyl, Amer. J. Math. **57**, 425 (1935).

Algebras with three anticommuting elements. II. Two algebras over a singular field

Nikos Salingaros
Physics Department, The University of Crete, Iraklion, Crete, Greece
Physics Department, University of Massachusetts, Boston, Massachusetts 02125

(Received 25 February 1981; accepted for publication 29 May 1981)

The Clifford algebra Ω generated by the elements $\{1,\omega\}$ with $(\omega)^2 = +1$, is an Abelian ring of dimension two with properties analogous to the complex field $\mathbb{C}$. The ring Ω has a string of singular inverses, and may be regarded as a "singular field" which circumvents both the fundamental theorem of algebra and the Frobenius theorem. We construct two associative algebras of dimension four over Ω: the Clifford algebra Ω_1 and the biquaternions of Clifford Ω_2, and demonstrate that both algebras possess inverses everywhere except on a singular region akin to the light cone of the Minkowski space. Matrix representations are discussed, as well as the importance of the algebras Ω_1 and Ω_2, in the description of physical vector fields.

PACS numbers: 02.10.Tq

I. INTRODUCTION

This paper extends the construction of algebras with three anticommuting elements over the real and complex fields $\mathbb{R}$ and $\mathbb{C}$ given in Ref. 1, by considering algebras over the ring Ω.

Motivated by the appearance of the Clifford algebra Ω (elsewhere denoted $\mathbb{R} \oplus \mathbb{R}$) as the only abelian Clifford algebra other than $\mathbb{R}$ and $\mathbb{C}$,[2–5] we here interpret it as a field extension of $\mathbb{R}$ of dimension two, analogous to the complex field $\mathbb{C}$. The key difference is that the ring Ω possesses a string of singular inverses. For this reason Ω circumvents both the Fundamental theorem of algebra[6,7] and the Frobenius theorem.[8,9]

It is possible to extend the construction of algebras with three anticommuting elements and a unit given in Ref. 1 to the case where the underlying field can be the ring Ω. Two new algebras are obtained, which are identified by group-theoretical methods to be isomorphic to the Clifford algebra Ω_1 [elsewhere denoted $N_1 \oplus N_1$ or $\mathbb{R}(2) \oplus \mathbb{R}(2)$] and the biquaternions of Clifford Ω_2 (elsewhere denoted $H \oplus H$).[4,5,10] We then show explicitly that the algebras Ω_1 and Ω_2 possess inverses everywhere except on a singular region akin to the light cone of the Minkowski space. Therefore, the algebras Ω_1 and Ω_2 are division algebras with zero divisors, or "singular division algebras" of dimension four over the ring Ω.

The analysis of Ref. 1 and the present paper, combined with the results of Refs. 4 and 5, leads to a natural identification of the Clifford algebras of dimensions one, two, four, and eight (Table III) as associative singular division algebras. Of these eight algebras, the three familiar algebras $\mathbb{R}$, $\mathbb{C}$, and H are the associative division algebras without zero divisors other than the zero element.

II. THE SINGULAR FIELD Ω

We construct Ω following the procedure of Refs. 4 and 5 as follows. Consider a one-dimensional flat Riemannian space $M^{1,0}$ defined by the differential 1-form $\omega = dx$ and the metric $g^{11} = \omega \vee \omega = (\omega,\omega) = +1$. We consider the set $\{1,\omega,-\omega,-1\}$. These elements along with the "vee" product $\vee$ define a finite group isomorphic to the Gauss–Klein "Veergruppe" $Z_2 \otimes Z_2$, which is isomorphic to the dihedral group D_2.

From this group we can construct an associative Abelian ring over $\mathbb{R}$, which we denote Ω. (Note that in Refs. 2–5 Ω was denoted as $\mathbb{R} \oplus \mathbb{R}$; the reason for this is discussed in Sec. IV.) An element α of the ring Ω can be written as

$$\alpha = x + \omega y, \quad x,y \in \mathbb{R}. \tag{1}$$

Define a conjugate of α, denoted by a tilde (as in $\mathbb{C}$),

$$\tilde{\alpha} = x - \omega y. \tag{2}$$

Then a scalar product is

$$\tilde{\alpha} \vee \alpha = \alpha \vee \tilde{\alpha} = (x + \omega y)(x - \omega y) = x^2 - y^2. \tag{3}$$

Clearly, the inverse α^{-1} in Ω is given by

$$\alpha^{-1} = \frac{\tilde{\alpha}}{\tilde{\alpha} \vee \alpha} = \frac{x - \omega y}{x^2 - y^2}. \tag{4}$$

There is a string of singularities at the points where $x = \pm y$; hence the ring Ω has more than one divisor of zero (Fig. 1). This property poses no great difficulty in physics, where singularities of this type are a common feature of field theory in Minkowski space. We may therefore manipulate the elements in Ω with the same ease as complex numbers in $\mathbb{C}$, except along the singular lines. Since Ω is an Abelian associative division ring with zero divisors, we may refer to Ω as a "singular field," for brevity.

Note in particular the case of $x = y = 1$ in (1) and (4). The elements $(1 + \omega)$ and $(1 - \omega)$ have singular inverses.

The ring Ω circumvents the Frobenius theorem on the construction of associative division algebras over $\mathbb{R}$, which does not permit singular points other than the point corresponding to the zero element [i.e., the point (0,0) in Fig. 1].[8,9] In the Frobenius construction, one uses the lemma that when a product of terms is zero, then at least one of those terms is zero.[8,9] Clearly this is *not* true for elements of Ω, and this is the reason why Ω is not well known as a division algebra in the usual sense.

The reason why Ω is not identified as a field is that fields are usually constructed via the fundamental theorem of algebra.[6,7] We note, however, that, in this context, the equation

$$x^2 - 1 = 0 \tag{5}$$

 0022-2488/81/102096-05$01.00

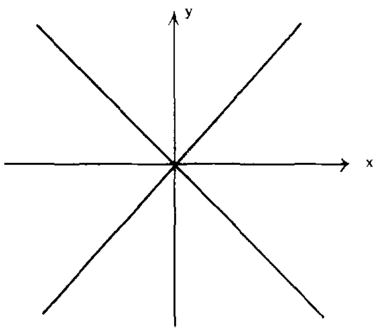


FIG. 1. Singular regions of the singular field $\mathbf{\Omega}$.

has a solution both in $\mathbb{R}$ and in $\mathbf{\Omega}$, as follows:

$$x = \pm 1 \in \mathbb{R}, \tag{6a}$$

$$x = \pm \omega \in \mathbf{\Omega}. \tag{6b}$$

Because a solution of (5) already exists in $\mathbb{R}$, the usual field construction as defined by the roots of a real polynomial overlooks the existence of the ring $\mathbf{\Omega}$ entirely.

Note that factoring Eq. (5) with (6b) to obtain $(1+\omega)(1-\omega)=0$ *does not* imply that $\omega = \pm 1$, since both factors $(1+\omega)$ and $(1-\omega)$ have singular inverses, and one cannot divide by either to obtain $(1 \pm \omega) = 0$. (This is the point that is important in the proof of the Frobenius theorem.)

This completes the description of the ring $\mathbf{\Omega}$. We have paid particular attention to the qualities of $\mathbf{\Omega}$, which make it a useful algebraic tool.

III. TWO ALGEBRAS OVER THE SINGULAR FIELD Ω

We recall the construction of algebras with three anticommuting elements and a unit which was presented in Ref. 1. The general multiplication of the three elements e_1, e_2, e_3, can be written using scalar coefficients $\alpha_k, \beta_k, \gamma_k$, $k = 1,2,3$ which are elements of some field:

$$\begin{aligned} e_1e_2 &= \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3, \\ e_2e_3 &= \beta_1 e_1 + \beta_2 e_2 + \beta_3 e_3, \\ e_3e_1 &= \gamma_1 e_1 + \gamma_2 e_2 + \gamma_3 e_3. \end{aligned} \tag{7}$$

We define a scalar square for each element as

$$(e_k)^2 = a_k, \quad a_k = \pm 1, \quad k = 1,2,3. \tag{8}$$

Imposing alternativity reduces the system of equations (7) to the following (see Ref. 1):

$$\begin{aligned} e_1e_2 &= \alpha_3 e_3, \\ e_2e_3 &= \beta_1 e_1, \\ e_3e_1 &= \gamma_2 e_2. \end{aligned} \tag{9}$$

In addition, alternativity implies the following relations between the scalar coefficients $\alpha_k, \beta_k, \gamma_k$ and a_k[1]:

$$\begin{aligned} a_1 &= -\alpha_3\gamma_2, \\ a_2 &= -\alpha_3\beta_1, \\ a_3 &= -\beta_1\gamma_2. \end{aligned} \tag{10}$$

In order to determine the structure of the algebra defined by relations (9), it is necessary to solve relations (10) over a field. Since the constants a_k are either 1 or -1 (8), it is sufficient to consider the following four distinct cases.

$$\begin{aligned} \{a_1,a_2,a_3\} &= \{1, 1, 1\} && \text{(11a)} \\ &= \{1, 1, -1\} && \text{(11b)} \\ &= \{1, -1, -1\} && \text{(11c)} \\ &= \{-1, -1, -1\}. && \text{(11d)} \end{aligned}$$

We proceed to solve Eq. (10) in each case (11) over a field of dimension two, i.e., either $\mathbb{C}$ or $\mathbf{\Omega}$. We obtain the following solutions, which are easily checked:

$$\alpha_3 = \beta_1 = \gamma_2 = \pm i, \tag{12a}$$

$$-\alpha_3 = \beta_1 = \gamma_2 = \pm \omega, \tag{12b}$$

$$\alpha_3 = -\beta_1 = \gamma_2 = \pm i, \tag{12c}$$

$$\alpha_3 = \beta_1 = \gamma_2 = \pm \omega. \tag{12d}$$

Cases (12a) and (12c) were identified with the algebra of Pauli matrices S_1 in Ref. 1. Cases (11b) and (11d) have already been factored over $\mathbb{R}$ in Ref. 1, and they defined two associative Clifford algebras of dimension four over $\mathbb{R}$: the Hamilton quaternions H and the dihedral Clifford algebra N_1 [elsewhere denoted $\mathbb{R}(2)$].

We now show that cases (12b) and (12d) factored over the singular field $\mathbf{\Omega}$ define the Clifford algebra Ω_1 [elsewhere denoted $N_1 \oplus N_1$ or $\mathbb{R}(2) \oplus \mathbb{R}(2)$], and the biquaternions of Clifford Ω_2 (elsewhere denoted $H \oplus H$).

The identification is obtained by computing the order of the finite group corresponding to each case (12b) and (12d). Each group consists of the elements $\{\pm 1, \pm e_k, \pm \omega e_k, \pm \omega\}$, $k = 1,2,3$. Following the procedure of Refs. 1, 4, and 5 in computing the order of each element, we obtain the group order corresponding to cases (12b) and 12d), using (11) and the property $(\omega)^2 = +1$:

$$\text{case (b):} \quad \text{order} = 1^1 2^{11} 4^4 = (1,11,4), \tag{13a}$$

$$\text{case (d):} \quad \text{order} = 1^1 2^3 4^{12} = (1,3,12). \tag{13b}$$

We recall from Refs. 4 and 5 that every Clifford algebra has a corresponding finite group. The correspondence is listed in Table I, for algebras up to order 8. Here, Z_n are the cyclic groups, D_n are the dihedral groups, and Q_4 is the quaternion group. For details on the construction of Table I, see Refs. 4 and 5. Using the fact that, for small order, equivalence of the orders implies a group isomorphism, we find that case (b) is isomorphic to the group $D_4 \otimes Z_2$, and case (d)

TABLE I. Finite group corresponding to Clifford algebras.

Dimension	Algebra	Finite group	Order
1	$\mathbb{R}$	Z_2	(1,1,0)
2	$\mathbb{C}$	Z_4	(1,1,2)
	$\mathbf{\Omega}$	$D_2 = Z_2 \otimes Z_2$	(1,3,0)
4	N_1	D_4	(1,5,2)
	$N_2 = H$	Q_4	(1,1,6)
8	S_1	$Z_4 \otimes D_2$	(1,7,8)
	Ω_1	$D_4 \otimes Z_2$	(1,11,4)
	Ω_2	$Q_4 \otimes Z_2$	(1,3,12)

TABLE II. Multiplication tables of the algebras Ω_1 and Ω_2.

	$\vee$	e_1	e_2	e_3
Ω_1	e_1	1	$-\omega e_3$	$-\omega e_2$
	e_2	ωe_2	1	ωe_1
	e_3	ωe_2	$-\omega e_1$	-1
	$\vee$	e_1	e_2	e_3
Ω_2	e_1	-1	$-\omega e_3$	ωe_2
	e_2	ωe_3	-1	$-\omega e_1$
	e_3	$-\omega e_2$	ωe_1	-1

is isomorphic to the group $Q_4 \otimes Z_2$. The corresponding Clifford algebras are also obtained from Table I, and are the following:

$$\text{case (b):}\quad \text{order } (1,11,4) \leftrightarrow D_4 \otimes Z_2 \leftrightarrow \Omega_1, \tag{14a}$$

$$\text{case (d):}\quad \text{order } (1,3,12) \leftrightarrow Q_4 \otimes Z_2 \leftrightarrow \Omega_2. \tag{14b}$$

The algebra Ω_2 was first explicitly constructed by Clifford himself, and named the "biquaternions."[10] The algebra Ω_1 is not well known.

A useful summary of the algebraic construction of the algebras Ω_1 and Ω_2 is obtained by giving their multiplication tables (Table II). This completes the construction of the algebras Ω_1 and Ω_2 via an extension of the method introduced in Ref. 1 to the singular field $\mathbf{\Omega}$.

IV. DISCUSSION

We recall the construction of Clifford algebras given in Refs. 4 and 5. We were able to construct all Clifford algebras as three distinct types of algebras N, S, and Ω (See Table I).

The abelian cases correspond to the "fields" $\mathbb{R}$, $\mathbb{C}$, and $\mathbf{\Omega}$. The algebras with anticommuting elements are defined by N_k, S_k, and Ω_k, $k = 1,2,\cdots$. Of the N algebras, $N_2 = H$, the quaternions of Hamilton. Elsewhere, we have identified N_3 as the algebra of the Majorana matrices.[11]

Of the S algebras, S_1 is the familiar algebra of Pauli matrices while $S_2 = D$ is the Dirac algebra. Of the Ω algebras, only Ω_2 is well known, and was discovered by Clifford as the "biquaternions."[10]

Using group-theoretical methods,[1,4,5,11] it is possible to derive the following key relations between the algebras.

Theorem 1:

$$S_k = N_{2k} \otimes \mathbb{C} = N_{2k-1} \otimes \mathbb{C}, \tag{15a}$$

$$\Omega_k = N_k \otimes \mathbf{\Omega}. \tag{15b}$$

In particular, these relations (15) clarify the following point: The biquaternion algebra Ω_2 is frequently confused with the algebra of complex quaternions which is isomorphic to the algebra of Pauli matrices S_1. From (15), these algebras are related to the quaternions H in the following simple manner (Table I):

$$\Omega_2 = H \otimes \mathbf{\Omega}, \quad S_1 = H \otimes \mathbb{C}. \tag{16}$$

This relationship illustrates the distinction between the biquaternions and the complex quaternions.

Following the analysis of Sec. II, it is possible to consider the fields $\mathbb{C}$ and $\mathbf{\Omega}$ as field extensions of $\mathbb{R}$, by the elements i and ω, respectively. These can be written as

$$\mathbb{C} = \mathbb{R} \oplus i\mathbb{R}, \tag{17a}$$

$$\mathbf{\Omega} = \mathbb{R} \oplus \omega\mathbb{R}. \tag{17b}$$

Using (17) in Theorem 1 [Eq. (15)] gives us the following relations:

$$S_k = N_{2k} \oplus iN_{2k} = N_{2k-1} \oplus iN_{2k-1}, \tag{18a}$$

$$\Omega_k = N_k \oplus \omega N_k. \tag{18b}$$

These relations form the basis for constructing and relating the higher-order Clifford algebras.

We now clarify a point which is an important distinction to previous work. It is clear that the vector space corresponding to the ring $\mathbf{\Omega}$ is a plane $\mathbb{R}^2$ which is isomorphic to the sum of two real lines $\mathbb{R} \oplus \mathbb{R}$. This vector space isomorphism has previously been employed in discussing the structure of Clifford algebras,[2–5] and is responsible for the notation $N_k \oplus N_k$ for Ω_k, instead of $N_k \oplus \omega N_k$, (18b). In particular, the biquaternions Ω_2 were identified with $H \oplus H$ instead of $H \oplus \omega H = H \otimes \mathbf{\Omega}$, (16). This interpretation has the serious disadvantage that it does not give the correct finite group corresponding to each Clifford algebra (it gives a group one-half the size), and also that it effectively hides the singular field $\mathbf{\Omega}$.

An important structural result is the determination of the center of each algebra. This was obtained in Refs. 4 and 5, and follows in part as a corollary of Theorem 1 [Eq. (15)]:

Theorem 2:

$$\text{The algebras } \begin{cases} N_k \\ S_k \\ \Omega_k \end{cases} \text{ have center } \begin{cases} \mathbb{R}, & (19a) \\ \mathbb{C}, & (19b) \\ \mathbf{\Omega}. & (19c) \end{cases}$$

Following the notation introduced in Ref. 4 and 5, we can consider the "fields" $\mathbb{R}$, $\mathbb{C}$, and $\mathbf{\Omega}$ as the zeroth algebras of type N, S, and Ω as follows: $\mathbb{R} = N_0$, $\mathbb{C} = S_0$, $\mathbf{\Omega} = \Omega_0$.

These results summarize many of the properties of the Clifford algebras. We conclude with a classification of the Clifford algebras up to dimension eight, which is given in Table III. Recall from Refs. 4 and 5 that each Clifford algebra is generated by the basis forms of a flat Riemannian space $M^{p,q}$ of dimension $n = p + q$. The metric is the scalar

TABLE III. Clifford algebras corresponding to Riemannian spaces.

							dimension over $\mathbb{R}$
			$\mathbb{R}$				1
		$\mathbf{\Omega}$		$\mathbb{C}$			2
	N_1		N_1		$N_2 = H$		4
S_1		Ω_1		S_1		Ω_2	8
							dimension over $\mathbb{R}$
			$A^{0,0}$				1
		$A^{1,0}$		$A^{0,1}$			2
	$A^{2,0}$		$A^{1,1}$		$A^{0,2}$		4
$A^{3,0}$		$A^{2,1}$		$A^{1,2}$		$A^{0,3}$	8

form

$$g^{\alpha\alpha} = ((+1, +1, ..., -1, -1, ...), \quad \alpha = 1,...,n, \tag{20}$$

where there are p plus ones, and q minus ones (the order of signs is immaterial).

One considers the set of 2^n p-forms of all ranks.[12] These basis forms, when endowed with the "vee" product,[4,5] define an associative ring of dimension 2^n over $\mathbb{R}$, isomorphic to the Clifford algebra $A^{p,q}$.

Our construction in terms of the N, S, and Ω algebras indentifies those $A^{p,q}$ which are isomorphic for distinct metrics, (20) (Table III). For example, for the Clifford algebras up to order 8 we can read off the following isomorphisms from Table III: $A^{2,0} \approx A^{1,1}$ and $A^{3,0} \approx A^{1,2}$.

V. DIVISIBILITY AND INVERSES

The demonstration that the algebras Ω_1 and Ω_2 are singular division algebras can be given in two different ways. First a simple algebraic proof based on Theorem 1 and the results of Ref. 1.

Theorem 3: $\Omega_1 = N_1 \otimes \mathbf{\Omega}$ and $\Omega_2 = H \otimes \mathbf{\Omega}$ are singular division algebras of dimension four over the ring $\mathbf{\Omega}$.

Proof: Since H and N_1 are singular division algebras of dimension four over $\mathbb{R}$,[1] it follows that their Kroenecker products with the ring $\mathbf{\Omega}$ are singular division algebras of the same dimension over the ring $\mathbf{\Omega}$.

For the practical interests of physics, it is not enough merely to show the existence of inverses; one should actually display them. We proceed to explicity construct the inverses in each algebra, utilising the "vee" product formalism from Refs. 4 and 5, and the multiplications of Table II.

An element α of Ω_2 can be written with eight real coefficients u^0, u^k, v^k, v^0, $k = 1,2,3$ as follows. The vector notation denotes expansion on the three basis elements.

$$\alpha = u^0 + \mathbf{u} + \omega\mathbf{v} + \omega v^0, \tag{21a}$$

$$\mathbf{u} = \sum_{k=1}^{3} u^k e_k, \quad \mathbf{v} \sum_{k=1}^{3} v^k e_k \,. \tag{21b}$$

The product in this algebra is obtained using the vector product rule, given here in terms of the familiar dot and cross products (see Refs. 4 and 5 for the role of the duality theorem in the product):

$$\mathbf{u} \vee \mathbf{v} = -(\mathbf{u}\cdot\mathbf{v}) - \omega\mathbf{u}\mathbf{x}\mathbf{v}. \tag{22}$$

Following the discussion of Sec. IV, we have identified the element ω, which is in the center of Ω_2, (19). (This corresponds to the volume element ω^3 in the geometrical construction of Refs. 4 and 5.) Define a conjugate for α as follows:

$$\tilde{\alpha} = u^0 - \mathbf{u} - \omega\mathbf{v} + \omega v^0. \tag{23}$$

Note that the conjugate does *not* mean that just $\tilde{\omega} = -\omega$; it is a definition on all the basis elements.

Using (21)–(23), we obtain the product $\tilde{\alpha} \vee \alpha$, which is a quantity in the singular field $\mathbf{\Omega}$:

$$\tilde{\alpha} \vee \alpha = (u^0)^2 + |\mathbf{u}|^2 + |\mathbf{v}|^2 + (v^0)^2 + 2\omega[u^0v^0 + (\mathbf{u}\cdot\mathbf{v})] \,. \tag{24}$$

We separate (24) into a real part, and the coefficient of ω (which in the case of $\mathbb{C}$ would correspond to the complex part):

$$x = (u^0)^2 + |\mathbf{u}|^2 + |\mathbf{v}|^2 + (v^0)^2,$$

$$y = 2(u^0v^0 + \mathbf{u}\cdot\mathbf{v}). \tag{25}$$

The inverse α^{-1} in Ω_2 is now obtained from (1), (4), and (25) as follows:

$$\alpha^{-1} = \frac{\tilde{\alpha}}{\tilde{\alpha} \vee \alpha} = \frac{\tilde{\alpha}(x - \omega y)}{x^2 - y^2} \,. \tag{26}$$

Note the existence of a singularity at the point where $x = \pm y$ (Fig. 1). The identity $\tilde{\alpha} \vee \alpha = \alpha \vee \tilde{\alpha}$ ensures that the inverse is two-sided. This demonstrates explicitly that the algebra Ω_2 is a division algebra with zero divisors.

Using a identical method and Table II, we can give the inverse in the algebra Ω_1 as formula (26), with the following substitutions:

$$x = (u^0)^2 + (v^0)^2 - (u^1)^2 - (u^2)^2 + (u^3)^2 - (v^1)^2 - (v^2)^2 + (v^3)^2, \tag{27a}$$

$$y = 2(u^0v^0 - u^1v^1 - u^2v^2 + u^3v^3). \tag{27b}$$

This completes the explicit construction of the inverses in the algebras Ω_1 and Ω_2.

VI. MATRIX REPRESENTATIONS

It is easy to find a 2×2 matrix representation of the algebras Ω_1 and Ω_2. Using Theorem 1 [Eq. (15)], we have the relations

$$\Omega_1 = N_1 \otimes \mathbf{\Omega}, \tag{28a}$$

$$\Omega_2 = N_2 \otimes \mathbf{\Omega} = H \otimes \mathbf{\Omega}. \tag{28b}$$

Recall from Ref. 1 that the representations of N_1 are 2×2 matrices in $\mathbb{R}$; those of H are 2×2 matrices in $\mathbb{C}$, which, however, do not span the space $\mathbb{C}(2)$. Using (28), we obtain the representation space of Ω_1 as 2×2 matrices over $\mathbf{\Omega}$. The representations of Ω_2 are 2×2 matrices over $\mathbb{C} \otimes \mathbf{\Omega}$. Once we know the representation space, it is not difficult to find a set of matrices satisfying the products in Table II. These are given in Table IV.

For those preferring not to work with the algebraic unit ω, it is possible to construct 4×4 matrix representations of Ω_1 and Ω_2 over $\mathbb{R}$ and $\mathbb{C}$, respectively. This is achieved by employing the following 2×2 matrix representation of the field $\mathbf{\Omega}$ over $\mathbb{R}$:

$$1 \leftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \omega \leftrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{29}$$

Substituting (29) in Table IV, one obtains two sets of 4×4 matrices: a real representation of Ω_1 and a complex

TABLE IV. Matrix representations of the algebras Ω_1 and Ω_2.

	Ω_1	Ω_2
e_1	$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$
e_2	$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$
e_3	$\begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$

representation of Ω_2. These matrices can be used to reproduce the representation-free manipulations of the previous section.

VII. CONCLUSION

In conclusion, we have been able to demonstrate some interesting properties for the Clifford algebras up to order 8 (Tables I and III). The algebras of dimension one and two are Abelian algebras: $\mathbb{R}$ and $\mathbb{C}$ are the well-known fields, while the singular field $\mathbf{\Omega}$ is not as well known, and has the interesting feature that it possesses a string of singular inverses, i.e., divisors of zero.

In this paper, we showed that the Clifford algebras Ω_1 and Ω_2 are singular division algebras of dimension four over the singular field $\mathbf{\Omega}$. In Ref. 1, we did the same thing for the Clifford algebras N_1, H (Hamilton's quaternions), and S_1 (the algebra of Pauli matrices) as singular division algebras over the fields $\mathbb{R}$ and $\mathbb{C}$. All of these algebras, except H, possess strings of singular inverses.

Usually, when one considers associative division algebras with three anticommuting elements, one thinks only of the quaternions H. The present discussion demonstrates that the singular division algebras with three anticommuting elements N_1, S_1, Ω_1, and Ω_2, can be considered just as useful as the quaternions H. The connection of some of these algebras with physically relevant structures, i.e., N_1 with the real 2-spinors; S_1 with the complex 2-spinors; and Ω_2 with the familiar vector algebra[4,10,13] demonstrates the importance of these algebras in physical description.

Another point is that we provide an answer to the question of generalizing the quaternions H while retaining associativity and divisibility. Following the discussion of this paper, both the algebra of Pauli matrices S_1 and the biquaternions of Clifford Ω_2 are associative generalizations of the quaternions H which maintain divisibility, but introduce strings of singular inverses.

If the Clifford algebras are considered, as is implied by our discussion, as a useful framework for physical description, then there are several unique advantages to this scheme. First, there are an infinite number of Clifford algebras (which by definition are all associative). Second, the algebras of dimension 16 and 32 include the Dirac, Majorana, and Clifford algebras in Minkowski spacetime,[4,5,11,14] which are of fundamental importance in field theory. Third, all the algebras of dimensions larger than 8 (over $\mathbb{R}$), while not division algebras, possess the more restricted property of "sectional divisibility."[4,5] (This means that any antisymmetric tensor field has a unique two-sided inverse in the algebra, a fact which simplifies many manipulations.) The Clifford algebras that are division algebras, or singular division algebras, are just those which have three anticommuting elements. Hence the concept of an associative division algebra, with or without zero divisors, appears to be a property peculiar and unique to the algebras with three anticommuting elements.

We believe that this paper, along with Ref. 1, has clarified the structure of associative algebras with three anticommuting elements that are applicable to the description of physical systems.

[1]Y. Ilamed and N. Salingaros, "Algebras with Three Anticommuting Elements. I," J. Math. Phys. **22**, 2091 (1981).

[2]M. F. Atiyah, R. Bott, and A. Shapiro, Topology **3**, (Suppl. 1), 3 (1964).

[3]M. Karoubi, *K-Theory* (Springer-Verlag, Berlin, 1979).

[4]N. Salingaros, in *Proceedings of the Second Workshop on Lie-Admissible Formulations*, Harvard University, 1979, Hadronic J. **3**, 339 (1979).

[5]N. Salingaros, J. Math. Phys. **22**, 226 (1981).

[6]G. Birhoff and S. MacLane, *A Survey of Modern Algebra*, 4th ed. (MacMillan, New York, 1977).

[7]V. I. Arnold, *Ordinary Differential Equations* (MIT Press, Cambridge, Mass., 1978).

[8]F. G. Frobenius, J. Reine Angew. Math. **84**, 59 (1877).

[9]A. G. Kurosh, *Lectures on General Algebra* (Chelsea, New York, 1963).

[10]W. K. Clifford, Amer. J. Math. **1**, 350 (1878).

[11]N. Salingaros, in *Proceeedings of the Third Workshop on Lie-Admissible Formulations*, University of Massachusetts at Boston, 1980, Hadronic J. **4**, 949 (1981).

[12]H. Flanders, *Differential Forms* (Academic, New York, 1963).

[13]*The Scientific Papers of J. W. Gibbs* (Dover, New York, 1961).

[14]N. Salingaros and M. Dresden, Phys. Rev. Lett. **43**, 1 (1979).

A unified theory of the point groups and their general irreducible representations

Shoon K. Kim

Physical Chemistry Laboratory, Oxford University, South Parks Road, Oxford OX13QZ, England and Department of Chemistry, [a] *Temple University, Philadelphia, Pennsylvania 19122*

(Received April 1981; accepted for publication 18 May 1981)

Point groups and their general irreducible (vector and projective) representations are characterized by the subgroup conditions for SU(2) in a unified scheme: these conditions are given by simple polynomial equations imposed on the matrix elements of the 2×2 unitary matrices of SU(2). The general irreducible representations for the point groups D_∞, D_n (with arbitrary integer $n \geqslant 2$), O, and T are given by four simple and effective tables.

PACS numbers: 02.20. − a

1. INTRODUCTION

Let G_R be a point group in three-dimensional Euclidean space and G_S be the 2×2 double-valued spinor representation of G_R.[1,2] As usual, G_S may be simply referred to as the double group of G_R. Evidently, double group G_S is a subgroup of SU(2), the group of all 2×2 unitary matrices with determinant 1. However, the direct relation between a double group G_S and SU(2) has never been investigated systematically. In the present work, we shall show that one can define a subgroup G_S of SU(2) by imposing a set of polynomial equations on the matrix elements of the 2×2 unitary matrices of SU(2). For simplicity, such a set of equations may be called a subgroup condition for SU(2). The effectiveness of this approach lies in the fact that the subgroup condition for a subgroup G_S provides the necessary and sufficient condition for reducing all irreducible representations of SU(2) into those of G_S. This is simply due to the fact that all the irreducible representations of SU(2) are described by the matrix element of $U \in \mathrm{SU}(2)$. The irreducible representations of G_S thus obtained are *general representations* in the sense that they hold for any element of G_S, each element being characterized by a set of roots of the polynomial equations which define the subgroup. This kind of general representation for the point groups of finite orders has never been reported previously (for the most comprehensive recent work on the point groups see Ref. 3).

We shall outline the content of the work. In Sec. 2, we shall first discuss the group requirements which should be satisfied by a subgroup condition of SU(2). Then, we shall discuss how the subgroup condition should be used to construct the point groups as well as their irreducible representations. As it has been proposed by many authors,[4,5] the irreducible representations of the double groups can be presented very effectively through the projective representations of the point groups. In the present work we shall follow this proposition.

In the remaining sections, we shall apply the general theory developed in Sec. 2 to individual subgroups of SU(2), i.e., the double groups of the point groups D_∞, D_n, C_∞, C_n, O, and T. There exists one more proper point group of finite order known as the icosahedral group I.[6] Since its analysis is much more involved and it is less important for applications, it may well be discussed elsewhere. In the final section, we shall briefly discuss how to extend the present results to improper point groups.

2. GENERAL DISCUSSION

An element of the group SU(2) is given by a 2×2 unitary matrix written as

$$U(a,b) = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix}, \tag{2.1a}$$

where a,b range over all complex numbers satisfying

$$|a|^2 + |b|^2 = 1. \tag{2.1b}$$

A pair of numbers (a,b) may be regarded as a unit vector (or spinor) in a two dimensional space $S^{(2)}$ over the complex field. Then, by definition there exists one-to-one correspondence between the unit vectors in $S^{(2)}$ and $U(a,b)$ in SU(2). One may define a subset $G_S = \{U(a,b)\}$ of SU(2) by imposing a set of polynomial equations

$$f_i(a,b) = 0, \quad 0 \leqslant i \leqslant 2 \tag{2.2}$$

on the matrix elements a and b. In order for the subset G_S to form a subgroup, the set must satisfy the group requirements:

$$\begin{aligned} &\text{(i)}\, f_i(1,0) = 0, \\ &\text{(ii)}\, f_i(a^*, -b) = 0, \quad \text{if } f_i(a,b) = 0, \\ &\text{(iii)}\, f_i(a_1a_2 - b_1b_2^*, a_1b_2 + b_1a_2^*) = 0, \\ &\qquad\qquad \text{if } f(a_1,b_1) = f(a_2,b_2) = 0. \end{aligned} \tag{2.3}$$

These follow since $U(1,0)$ is the identity, $U(a,b)^{-1} = U(a^*, -b)$, and $U(a_1,b_1)\,U(a_2,b_2) = U(a_1a_2 - b_1b_2^*, a_1b_2 + b_1a_2^*)$. These conditions may also be regarded as the requirements for a set of unit vectors $\{(a,b)\}$ in $S^{(2)}$ to be transitive with respect to a subgroup of SU(2). In order for G_S to be a double group of a point group, it is necessary to impose one more additional property,

$$\text{(iv)}\, f_i(-a, -b) = 0, \quad \text{if } f_i(a,b) = 0 \tag{2.4}$$

which means that $U(-a,-b) = -U(a,b)$ is contained in G_S if $U(a,b) \in G_S$. Hereafter, we may include (iv) in the subgroup condition of SU(2) unless otherwise specified. The

[a] Present address.

simplest subgroup condition for SU(2) is

$$b = 0, \tag{2.5}$$

which means that $|a| = 1$. It characterizes the spinor representation of SO(2). Another simple subgroup condition is given by

$$ab = 0, \tag{2.6}$$

which means that either $|a| = 1$ or $|b| = 1$. This condition characterizes the double group D'_∞ of the dihedral group D_∞. This group will be discussed in detail as a prototype example. A less obvious subgroup condition is given by

$$a^8 = 1 \quad \text{or} \quad b^8 = 1 \quad \text{or} \quad a^4 = b^4 = \pm\tfrac{1}{4}, \tag{2.7}$$

which will be shown to characterize the double group O' of the octahedral group O. It is well known[6] that there exists only a *finite number* of point groups if one excludes the uniaxial groups C_n and the dihedral groups D_n. This means that there exist only a finite number of different subgroup conditions for SU(2), unless $ab = 0$.

We shall now discuss how to construct the irreducible representations of a subgroup $G_S \in \mathrm{SU}(2)$ from those of SU(2) by means of a subgroup condition. Let us introduce the basis set belonging to the $(2j+1)\times(2j+1)$ irreducible representation $D^{(j)}(U)$ of $U(a,b)\in\mathrm{SU}(2)$. It is defined by a set of $(2j+1)$ monomials of two variables ξ_1 and ξ_2

$$\phi(j,m;\xi_1,\xi_2) = \xi_1^{j+m}\xi_2^{j-m}/[(j+m)!(j-m)!]^{1/2}, \quad m = j, j-1, \ldots, -j, \tag{2.8}$$

where j is an integer or half-integer. The set transforms according to

$$\phi(j,m;\xi'_1,\xi'_2) = \sum_{m'} \phi(j,m';\xi_1,\xi_2) D^{(j)}(U(a,b))_{m',m} \tag{2.9a}$$

under the transformation of the variables

$$(\xi'_1,\xi'_2) = (\xi_1,\xi_2)U(a,b), \tag{2.9b}$$

where (ξ_1,ξ_2) is regarded as a raw vector. It should be noted that $D^{(1/2)}(U) = U$. The general expression of $D^{(j)}(U)$ is well known as a function of a and b,[1] but it is not needed in the present work. In the reduction of $D^{(j)}(U)$ we shall incorporate the subgroup condition directly into the left-hand side of (2.9a) and obtain the reduced representations by constructing the invariant subspaces spanned by appropriate linear combinations of the spinor components. As we shall see in the actual examples, it requires a trivial amount of the algebraic manipulations and provides the general expression for each irreducible representation of G_S, which holds for any $U(a,b)\in G_S$. The reason is that the subgroup condition does not contain any individual characteristic of each group element.

Now, we shall discuss how to characterize the individual group element of a subgroup G_S defined by a subgroup condition of SU(2). For this purpose, we shall introduce the parameter space of SO(3) and SU(2); a vector $\boldsymbol{\alpha}$ in the space defines a rotation $R(\boldsymbol{\alpha})\in\mathrm{SO}(3)$ as well as the double valued spinor representation $\pm S(\boldsymbol{\alpha})\in\mathrm{SU}(2)$ (see, for example, Ref. 5). The vector $\boldsymbol{\alpha}$ is called the rotation vector and is related to the axis of rotation $\mathbf{n}$ and the angle of rotation α about $\mathbf{n}$ (in the clockwise direction along $\mathbf{n}$) by

$$\boldsymbol{\alpha} = \alpha\mathbf{n}, \quad |\mathbf{n}| = 1. \tag{2.10}$$

We set

$$U(a,b) = \exp[-i\boldsymbol{\alpha}\cdot\boldsymbol{\sigma}] = S(\boldsymbol{\alpha}), \tag{2.11}$$

where $\boldsymbol{\sigma}$ is the Pauli spin. Then, we obtain the so-called Euler–Olinde–Rodrigues parametrization of a and b,

$$a(\boldsymbol{\alpha}) = \cos\frac{\alpha}{2} - in_z \sin\frac{\alpha}{2}, \qquad b(\boldsymbol{\alpha}) = -(in_x + n_y)\sin\frac{\alpha}{2}. \tag{2.12}$$

Here, $0\leqslant|\boldsymbol{\alpha}|\leqslant 2\pi$, which will be reduced to $0\leqslant|\boldsymbol{\alpha}|\leqslant\pi$ later by introducing the projective representation. Substitution of (2.12) into the subgroup condition (2.2) yields a set of possible values of the rotation vectors $\boldsymbol{\alpha}$'s, which then defines $G_S = \{S(\boldsymbol{\alpha})\}$ as well as $G_R = \{R(\boldsymbol{\alpha})\}$. The correspondence between G_S and G_R is obviously two-to-one since $S(\boldsymbol{\alpha} + 2\pi\mathbf{n}) = -S(\boldsymbol{\alpha})$ while $R(\boldsymbol{\alpha}+2\pi\mathbf{n}) = R(\boldsymbol{\alpha})$.

In practical applications, it is more convenient to introduce a projective (or ray) representation $\check{G}_S$, which makes one-to-one correspondence with the point group G_R. It has been established by the author[5] that the projective set $\check{G}_S = \{S(\boldsymbol{\alpha})\}$ of $G_R = \{R(\boldsymbol{\alpha})\}$, which preserves the class structure of G_R, is given by the correspondence

$$U(a,b) = S(\boldsymbol{\alpha}) \leftrightarrow R(\boldsymbol{\alpha}), \quad 0\leqslant|\boldsymbol{\alpha}|\leqslant\pi, \tag{2.13}$$

where $\boldsymbol{\alpha}$'s are chosen equivalently for equivalent $R(\boldsymbol{\alpha})$'s if there exist more than one $\boldsymbol{\alpha}$ corresponding to a given R. In fact, this is the case for a binary rotation on account of the cyclic boundary condition $R(\pi\mathbf{n}) = R(-\pi\mathbf{n})$. If, however, the binary rotation is bilateral,[7] i.e., $\pi\mathbf{n}$ and $-\pi\mathbf{n}$ are equivalent, one may choose any one of them. Frequently, it is convenient to choose the one on the positive hemisphere, a domain defined by

$$\pi n_3 > 0 \quad \text{or} \quad \pi n_3 = 0, \pi n_1 > 0, \quad \text{or} \quad \pi n_3 = \pi n_1 = 0, \pi n_2 > 0. \tag{2.14}$$

The correspondence (2.13) defines the projective factor system as follows. Let $S_i = S(\boldsymbol{\alpha}_i)$ and $R_i = R(\boldsymbol{\alpha}_i)$, then

$$S_1S_2 = f_{12}S_3 \quad \text{if } R_1R_2 = R_3, \tag{2.15a}$$

where

$$f_{12} = \begin{cases} 1, & \text{if } S_1S_2\in\check{G}_S, \\ -1, & \text{otherwise.}\end{cases} \tag{2.15b}$$

In terms of the projective set $\check{G}_S = \{S(\boldsymbol{\alpha})\}$, the double group of $G_R = \{R(\boldsymbol{\alpha})\}$ is given by $G_S = \{\pm S(\boldsymbol{\alpha})\}$. It is well established that all irreducible representations (integral and half-integral) of G_R are constructed by the irreducible representations of G_S. Let $P^{(j)}(S)$ be an irreducible representation of G_S, with j being integral or half-integral. Then, corresponding to (2.15), we have

$$P^{(j)}(-S) = (-1)^{2j}P^{(j)}(S). \tag{2.16}$$

Accordingly, when j is integral $P^{(j)}(S)$ gives a vector representation of G_R; when j is half-integral $P^{(j)}(S)$ gives a projective representation of G_R belonging to the factor system (2.15). It is also evident from (2.16) that all irreducible representations of the double group G_S follow from those of $\check{G}_S$.

Based on the general discussion given in this section we shall discuss the point groups D_∞, D_n, C_∞, C_n, O, and T in the following sections.

3. THE GROUP D_∞

The subgroup condition for the double group D'_∞ of the group D_∞ is given by

$$ab = 0. \tag{3.1a}$$

It is evident that this condition satisfies the group requirements (2.3) and (2.4). It is equivalent to

$$|a| = 1 \quad \text{or} \quad |b| = 1. \tag{3.1b}$$

Thus, there exist two types of the general elements,

$$U(a,0) = \begin{bmatrix} a & 0 \\ 0 & a^* \end{bmatrix}, \quad U(0,b) = \begin{bmatrix} 0 & b \\ -b^* & 0 \end{bmatrix}. \tag{3.2}$$

The set $\{U(a,0)\}$ forms a normal subgroup of D'_∞, which is the double group of SO(2). The element $U(0,b)$ is involutional in the sense that

$$[U(0,b)]^2 = -\mathbf{1}, \tag{3.3}$$

where $\mathbf{1}$ is the 2×2 unit matrix. This property ensures that $U(0,b)$ represents a binary rotation, as will be seen shortly.

In terms of the Euler–Olinde–Rodrigues parametrization (2.12), we have

$$a = a(\alpha) = \exp[-(i/2)\alpha],$$
$$b = b(\beta) = -i\exp[-(i/2)\beta], \tag{3.4}$$

where $-2\pi < \alpha,\beta \leqslant 2\pi$ for the double group. The one-to-one correspondence between the projective set $\check{G}_S = \{S(\boldsymbol{\alpha})\}$ and $G_R = \{R(\boldsymbol{\alpha})\}$ introduced by (2.13) yields

$$U(a,0) = \exp[-(i/2)\,\alpha\sigma_z] \longleftrightarrow C_z(\alpha),$$
$$U(0,b) = -i\boldsymbol{\sigma}\cdot\mathbf{h}_\beta \leftrightarrow C'_{2,\beta}, \tag{3.5}$$

where $-\pi < \alpha, \beta \leqslant \pi$, $\mathbf{h}_\beta = (\cos\frac{1}{2}\beta, \sin\frac{1}{2}\beta, 0)$, and $C_z(\alpha)$ denotes the rotation about the z axis through an angle α and $C'_{2,\beta}$ denotes a binary rotation about the vector $\mathbf{h}_\beta$. Since these binary rotations are bilateral, their rotation vector $\pi\mathbf{h}_\beta$ are placed in the positive hemisphere in accordance with (2.14). The correspondence (3.5) proves that the subgroup condition (3.1) indeed defines the double group D'_∞.

We shall now construct all the irreducible (vector and projective) representations of D_∞. From (2.9a) one can show that the subgroup condition (3.1) makes invariant the spinor spaces spanned by $\phi(j,0)$ if j is an integer, and by $\{\phi(j,m),\phi(j,-m)\}$ with $m = j, j-1,\ldots,1$ or $\frac{1}{2}$, respectively, under all $U(a,b) \in D'_\infty$. The corresponding 2×2 representations

$$\bar{M}^j_m[U(a,0)] = \begin{bmatrix} a^{2m} & 0 \\ 0 & a^{*2m} \end{bmatrix},$$
$$\bar{M}^j_m[U(0,b)] = (-i)^{2j}\begin{bmatrix} 0 & (ib)^{2m} \\ (ib)^{*2m} & 0 \end{bmatrix}, \tag{3.6}$$

are irreducible since it is impossible to satisfy $a^{2m} = a^{*2m}$ for all a's on the unit circle in the complex plane. In many practical applications, it is more convenient to introduce the following bases:

$$\phi_+(j,m) = 2^{-1/2}(\phi(j,m) + \phi(j,-m)),$$
$$\phi_-(j,m) = -i2^{-1/2}(\phi(j,m) - \phi(j,-m)). \tag{3.7}$$

Then, the corresponding representations are given by

$$M^j_m[U(a,0)] = \begin{bmatrix} \mathrm{Re}a^{2m} & \mathrm{Im}a^{2m} \\ -\mathrm{Im}a^{2m} & \mathrm{Re}a^{2m} \end{bmatrix},$$
$$M^j_m[U(0,b)] = (-i)^{2j}\begin{bmatrix} \mathrm{Re}(ib)^{2m} & -\mathrm{Im}(ib)^{2m} \\ -\mathrm{Im}(ib)^{2m} & -\mathrm{Re}(ib)^{2m} \end{bmatrix}, \tag{3.8}$$

where Re and Im denote the real and imaginary parts, respectively. These satisfy the equivalence relations

$$M^j_{-m} = \sigma_z M^j_m \sigma_z, \quad M^{j_e}_m = \sigma_y M^{j_o}_m \sigma_y, \tag{3.9}$$

where $j_e(j_0)$ is a j with an even (odd) integral part and σ_z,σ_y are the Pauli spin matrices. On account of these equivalences, for a given m different j's do not give any new respresentations except for the constraint $j \geqslant m$. Thus, inequivalent 2×2 irreducible representations are determined by m alone, which can be taken to be

$$m = \tfrac{1}{2}, 1, \tfrac{3}{2}, \cdots. \tag{3.10}$$

The one dimensional representations M^j_0 on $\phi(j,0)$ occur when j is integral and is given by

$$M^j_0[U(a,0)] = 1, \quad M^j_0[U(0,b)] = (-1)^j, \tag{3.11}$$

which splits into two inequivalent representations depending on $j = j_e$ or j_0. The final results for the irreducible representations of D_∞ are given in Table I, where use has been made of the Euler–Olinde–Rodrigues parametrization (3.4) with $-\pi < \alpha,\beta \leqslant \pi$. It is noted that if one takes

TABLE I. The irreducible (vector and projective) representations of D_∞.

D_∞	$C_z(\alpha)$	$C'_{2,\beta}$	Bases
A_1	1	1	$\phi(j_e,0)$
A_2	1	-1	$\phi(j_0,0)$
E_m	$\begin{bmatrix} \cos(m\alpha) & -\sin(m\alpha) \\ \sin(m\alpha) & \cos(m\alpha) \end{bmatrix}$	$(-i)^{2j}\begin{bmatrix} \cos(m\beta) & \sin(m\beta) \\ \sin(m\beta) & -\cos(m\beta) \end{bmatrix}$	$(\phi_+(j,m),\phi_-(j,m))$

(1) $m = 1,2,3,\ldots$, for vector representations, $m = \frac{1}{2},\frac{3}{2},\ldots$, for projective representations.

(2) $-\pi < \alpha,\beta \leqslant \pi$.

$-2\pi < \alpha,\beta \leqslant 2\pi$, then Table I gives the irreducible representation for the double group D'_∞.

4. THE GROUP D_n

The subgroup condition for the double group D'_n of the dihedral group D_n is given by

$$a^{2n} = 1 \quad \text{or} \quad (ib)^{2n} = 1, \quad n \geqslant 2. \tag{4.1}$$

Since D'_n is a subgroup of D'_∞, one can use the previous results of D'_∞ for this case with due modifications. Firstly, the subgroup condition (4.1) limits the allowed values of the angles α and β of (3.4) to the following values:

$$\alpha_k = 2\pi k/n, \quad \beta_q = 2\pi q/n, \tag{4.2}$$

where $k,q = 0, \pm 1,..., \pm (n-1),n$. The corresponding values of a and b are given by

$$a_k = \exp(-i\pi k/n), \quad ib_q = \exp(-i\pi q/n). \tag{4.3}$$

From (3.5), the correspondence between the projective set $\check{G}_S$ and G_R is given by

$$\begin{aligned} &U(a_k,0) = \exp(-i\sigma_z \pi k/n) \leftrightarrow C_n^k, \\ &U(0,b) = -i(\boldsymbol{\sigma}\cdot\mathbf{h}_q) \leftrightarrow C'_{2,q}, \end{aligned} \tag{4.4a}$$

where $\mathbf{h}_q = (\cos(\pi q/n), \sin(\pi q/n), 0)$, and C_n is the n-fold axis of rotation in the z direction and $C'_{2,q}$ is a binary rotation about $\mathbf{h}_q$. It is noted here that the equivalent sets of the binary rotation vectors are $\{\pi \mathbf{h}_q; q = 0, \pm \text{ even}\}$ and $\{\pi \mathbf{h}_q; q = \pm \text{ odd}\}$ and also that when n is even all $C'_{2,q}$ are bilateral while when n is odd all $C'_{2,q}$ are unilateral. Thus, for the projective representations which preserve the class structure of D_n, we take

$$\begin{aligned} &k,q = 0, \pm 1,..., \pm (n/2 - 1),n/2, \quad \text{for an even } n \\ &2k,q = 0, \pm 2,..., \pm (n-1), \quad \text{for an odd } n \end{aligned} \tag{4.4b}$$

consistent with (2.13) and (2.14).

Now all the irreducible representations of D_n follow from those of D_∞ with slight modifications. Substituting α_k and β_q of (4.2) into α and β in E_m of Table I, we obtain E_m as given in Table II, which satisfy

$$E_{m+n} = E_m. \tag{4.5}$$

From this and the equivalence relations (3.9), the inequivalent 2×2 representations are limited to

$$0 < m < n/2. \tag{4.6}$$

When $m = n/2, E_{n/2}$ of Table I is reduced into two one dimensional representations denoted as B_1 and B_2 of Table II. Accordingly, there exists a total of $(n+3)$ irreducible representations for D'_n, of which four are one dimensional and $(n-1)$ are two dimensional, satisfying

$$4n = 1^2 + 1^2 + 1^2 + 1^2 + (n-1)2^2. \tag{4.7}$$

This is consistent with the class structure of the double group $D'_n\{U(a,b)\}$ given by

$$\begin{aligned} &E,\bar{E} = U(-1,0), \{U(0,b_q); q = 0, \pm \text{ even}\}, \\ &\{U(0,b_q); q = \pm \text{ odd}\}, \{U(a_k,0), U(a_{-k},0)\}, \end{aligned} \tag{4.8}$$

where $k,q = 0, \pm 1,..., \pm (n-1),n$.

The final results of the general irreducible (vector and projective) representations of the dihedral group D_n are presented in Table II. It is noted that when n is even, all four one dimensional representations A_1,A_2 and B_1,B_2 are vector representations while when n is odd, A_1 and A_2 are vector representations, and B_1 and B_2 are projective representations. This is due to the fact that when n is even, all the binary rotations in D_n are bilateral while when n is odd all the binary rotations in D_n are unilateral. It is also noted that in Table II all the spinor bases of the proper rotation group SO(3) are classified according to the irreducible representations of D_n. Thus, one can read off the rotation group compatibility relations with D_n from the table without any further calculations. Since the table is completely general for any D_n, the notations introduced for the group elements as well as for the irreducible representations seem most rational choices. In particular, the integral or half-integral suffixes m for E_m distinguish the vector and projective representations most effectively.

5. THE GROUPS C_∞ AND C_n

The general representations for these two abelian groups are completely known. However, for the sake of completeness, we shall discuss these groups in the present line of

TABLE II. The irreducible (vector and projective) representations of D_n.

D_n	$C_n^{\ k}$	$C'_{2,q}$	Bases
A_1	1	1	$\phi(j_e,0),\phi_+(j_e,n),\phi_-(j_0,n)$
A_2	1	-1	$\phi(j_0,0),\phi_+(j_0,n),\phi_-(j_e,n)$
B_1	$(-1)^k$	$(-1)^q i^{\theta(n)}$	$\phi_+(j_0,n/2),\phi_-(j_e,n/2)$
B_2	$(-1)^k$	$(-1)^{q+1} i^{\theta(n)}$	$\phi_+(j_e,n/2),\phi_-(j_0,n/2)$
E_m	$\begin{bmatrix} \cos(2\pi mk/n) & -\sin(2\pi mk/n) \\ \sin(2\pi mk/n) & \cos(2\pi mk/n) \end{bmatrix}$	$(-i)^{2j} \begin{bmatrix} \cos(2\pi mq/n) & \sin(2\pi mq/n) \\ \sin(2\pi mq/n) & -\cos(2\pi mq/n) \end{bmatrix}$	$(\phi_+(j,m),\phi_-(j,m))$

(1) $\theta(n, \text{odd}) = 1$; $\theta(n, \text{even}) = 2$.
(2) $k,q = 0, \pm 1,..., \pm (n/2 - 1),n/2$; for an even n, $2k,q = 0, \pm 2,..., \pm (n-1)$; for an odd n.
(3) $m = \frac{1}{2},1,...,\frac{1}{2}(n-1)$.
(4) $\phi(j,m)$ and $\phi(j,m+n)$ belong to the same representation for $m \neq 0$.

approach. The spinor representation of the uniaxial group C_∞ [or SO(2)] is characterized by the subgroup condition

$$b = 0, \tag{5.1}$$

which is equivalent to $|a| = 1$ on account of (2.1b). Thus, from the Euler–Olinde–Rodrigues parametrization (3.4), the projective correspondence (3.5), and the representation (3.6) we have the following irreducible (vector and projective) representation for C_∞,

$$M_m[C_z(\alpha)] = \exp(-im\alpha), \quad -\pi < \alpha \leqslant \pi, \quad m = 0, \pm\tfrac{1}{2}, \pm 1, \cdots \tag{5.2}$$

with the basis set $\phi(j,m)$.

The subgroup condition for the double group C'_n of the group C_n is given by

$$a^{2n} = 1. \tag{5.3}$$

Thus, there exist $2n$ irreducible representations given by

$$M_m(C_n^k) = \exp(-i2\pi mk/n) \tag{5.4}$$

with the bases $\phi(j,m), \phi(j,m \pm n)$, where m are integers or half-integers in the range

$$-\tfrac{1}{2}n < m \leqslant \tfrac{1}{2}n \tag{5.5}$$

and

$$k = \begin{cases} 0, \pm 1, \ldots, \pm(n/2 - 1), n/2, & \text{for an even } n \\ 0, \pm 1, \ldots \pm(n-1)/2, & \text{for an odd } n \end{cases} \tag{5.6}$$

6. THE GROUP O

The subgroup condition for the double group O' of the octahedral group O is given by

$$a^8 = 1 \quad \text{or } b^8 = 1 \quad \text{or } a^4 = b^4 = \pm\tfrac{1}{4}. \tag{6.1}$$

In the Appendix, we shall show that these conditions do satisfy the group requirements (2.3) and (2.4). It is also a trivial matter to show that the order of the group defined by the subgroup condition is 48.

We shall next show by using (2.13) that the subgroup condition (6.1) indeed leads to the projective set $\check{O}$ of the group O. For this purpose, let us assume for the spinor representations $S(\boldsymbol{\alpha})$ of (2.13) to represent binary or threefold or fourfold rotations. Then we obtain the following results:

(a) For binary rotations, $S(\pi\mathbf{n}) = U(-in_3, -in_1 - n_2)$, we have

$$\begin{gathered} n_1^2 = 1 \quad \text{or } n_2^2 = 1 \quad \text{or } n_3^2 = 1 \quad \text{or} \\ n_1^2 = n_2^2 = \tfrac{1}{2} \quad \text{or } n_2^2 = n_3^2 = \tfrac{1}{2} \quad \text{or } n_3^2 = n_1^2 = \tfrac{1}{2}, \end{gathered} \tag{6.2}$$

which define 9 binary rotations, for $R(\pi\mathbf{n}) = R(-\pi\mathbf{n})$. Since these are all bilateral, one may place their rotation vectors $\pi\mathbf{n}$'s in the positive hemisphere.

(b) For threefold rotations, $S((2\pi/3)\mathbf{n}) = U(\tfrac{1}{2}(1 - i3^{1/2}n_3), -\tfrac{3}{2}(in_1 + n_2))$, we have

$$n_1^2 = n_2^2 = n_3^2 = \tfrac{1}{3}, \tag{6.3}$$

which define 8 threefold rotations.

(c) For fourfold rotations, $S((2\pi/4)\mathbf{n}) = U(2^{-1/2}(1 - in_3), -2^{-1/2}(in_1 + n_2))$, we have

$$n_1^2 = 1 \quad \text{or } n_2^2 = 1 \quad \text{or } n_2^2 = 1, \tag{6.4}$$

which define 6 fourfold rotations.

These axes of rotations together with the unit element E indeed define a total of 24 elements of the group $O = \{R(\boldsymbol{\alpha})\}$ and its projective set $\check{O} = \{S(\boldsymbol{\alpha})\}$. The projective set given in Table III is in accordance with (2.13) so that it preserves the class structure of O. The double group O' is obviously given by $O' = \{\pm S(\boldsymbol{\alpha})\}$.

Now, we shall construct the irreducible representations of the group O. Again, the subgroup conditions (6.1) is very effective in obtaining all the general irreducible (vector and projective) representations as given in Table III. For example, E and T_2 of Table III are obtained by showing that the

TABLE III. The irreducible (vector and projective) representations of O.

O	$E, 9S(\pi\mathbf{n}), 8S((2\pi/3)\mathbf{n}), 6S((\pi/2)\mathbf{n})$	Bases
Γ_1, A_1	1	1
Γ_2, A_2	$(-1)^\kappa \operatorname{sgn}(a^4 + b^4)$	$\phi_-(3,2)$ or xyz
Γ_3, E	$\begin{bmatrix} (-1)^\kappa(\lvert a\rvert^4 + \lvert b\rvert^4) & 2(3)^{1/2}a^2b^{*2} \\ 2(3)^{1/2}a^2b^2 & a^4 + b^4 \end{bmatrix}$	$[\phi(2,0), \phi_+(2,2)]$, or $[3z^2 - r^2, 3^{1/2}(x^2 - y^2)]$
Γ_4, T_1	$\begin{bmatrix} \mathrm{Re}(a^2 - b^2) & \mathrm{Im}(a^2 + b^2) & -2\,\mathrm{Re}(ab) \\ -\mathrm{Im}(a^2 - b^2) & \mathrm{Re}(a^2 + b^2) & 2\,\mathrm{Im}(ab) \\ 2\,\mathrm{Re}(ab^*) & 2\,\mathrm{Im}(ab^*) & aa^* - bb^* \end{bmatrix}$	$[-i\phi_-(1,1), i\phi_+(1,1), \phi(1,0)]$ or (x,y,z)
Γ_5, T_2	$\begin{bmatrix} (-1)^\kappa\,\mathrm{Re}(a^2 - b^2) & -(-1)^\kappa\,\mathrm{Im}(a^2 + b^2) & 4\,\mathrm{Re}(ab^{*3}) \\ (-1)^\kappa\,\mathrm{Im}(a^2 - b^2) & (-1)^\kappa\,\mathrm{Re}(a^2 + b^2) & -4\,\mathrm{Im}(ab^{*3}) \\ -4\,\mathrm{Re}(ab^3) & -4\,\mathrm{Im}(ab^3) & (a^4 - b^4) \end{bmatrix}$	$[i\phi_+(2,1) i\phi_-(2,1), \phi_-(2,2)]$ or (yz,zx,xy)
$\Gamma_6, E_{1/2}$	$U(a,b)$	$[\phi(\tfrac{1}{2},\tfrac{1}{2}), \phi(\tfrac{1}{2},-\tfrac{1}{2})]$
$\Gamma_7, E'_{1/2}$	$A_2 \times U(a,b)$	$\phi_-(3,2)[\phi(\tfrac{1}{2},\tfrac{1}{2}), \phi(\tfrac{1}{2},-\tfrac{1}{2})]$
Γ_8, Q	$E \times U(a,b) \cong E_{3/2}$	$[\phi(\tfrac{3}{2},\tfrac{3}{2}), \phi(\tfrac{3}{2},\tfrac{1}{2}), \phi(\tfrac{3}{2},-\tfrac{1}{2}), \phi(\tfrac{3}{2},-\tfrac{3}{2})]$

(1) $\kappa = 2|ab|$: $\kappa = 0$ if $ab = 0$; $\kappa = 1$ if $ab \neq 0$.
(2) $S(\pi\mathbf{n}) = U(a = -in_3, b = -in_1 - n_2)$: $n_1 = 1$ or $n_2 = 1$ or $n_3 = 1$; $n_1 = \pm n_2 = 2^{-1/2}$ or $n_3 = \pm n_2 = 2^{-1/2}$, or $n_3 = \pm n_1 = 2^{-1/2}$.
(3) $S((2\pi/3)\mathbf{n}) = U(a = \tfrac{1}{2}(1 - i(\sqrt{3})n_3), b = -\tfrac{3}{2}(in_1 + n_2))$: $n_1^2 = n_2^2 = n_3^2 = \tfrac{1}{3}$.
(4) $S((2\pi/4)\mathbf{n}) = U(a = 2^{-1/2}(1 - in_3), b = -2^{-1/2}(in_1 + n_2))$: $n_1^2 = 1$ or $n_2^2 = 1$ or $n_3^2 = 1$.
(5) $\Gamma_1 \sim \Gamma_8$, the notations used by Koster *et al.* (Ref. 9).

TABLE IV. The irreducible (vector and projective) representations of T.

T	$E,3S(\pi\mathbf{n}),8S((2\pi/3)\mathbf{n})$	Bases
Γ_1,A	1	1 or $\phi_-(3,2)$ or xyz
Γ_2,B_1	$(-\frac{1}{2}+i2(3)^{1/2}a^2b^2)^\kappa$	$2^{-1/2}[\phi(2,0)+i\phi_+(2,2)]$ or $2^{-1/2}(u+iv)$
Γ_3,B_2	$(-\frac{1}{2}-i2(3)^{1/2}a^2b^2)^\kappa$	$2^{-1/2}[\phi(2,0)-i\phi_+(2,2)]$ or $2^{-1/2}(u-iv)$
Γ_4,T	T_1 of 0	$(-i\phi_-(1,1),i\phi_+(1,1),\phi(1,0))$ or $(i\phi_+(2,1),i\phi_-(2,1),\phi_-(2,2))$ or (x,y,z) or (yz,zx,xy)
$\Gamma_5,E_{1/2}$	$U(a,b)$	$(\phi(\frac{1}{2},\frac{1}{2}),\phi(\frac{1}{2},-\frac{1}{2}))$
$\Gamma_6,E'_{1/2}$	$B_1\times U(a,b)$	$2^{-1/2}(u+iv)\times(\phi(\frac{1}{2},\frac{1}{2}),\phi(\frac{1}{2},-\frac{1}{2}))$
$\Gamma_7,E''_{1/2}$	$B_2\times U(a,b)$	$2^{-1/2}(u-iv)\times(\phi(\frac{1}{2},\frac{1}{2}),\phi(\frac{1}{2},-\frac{1}{2}])$

(1) $\kappa=2|ab|$.
(2) $u=3z^2-r^{-2},v=3^{1/2}(x^2-y^2)$.
(3) $S(\pi\mathbf{n})=U(a=-in_3,\ b=-in_1-n_2)$; $n_1=1$ or $n_2=1$ or $n_3=1$.
(4) $S((2\pi/3)\mathbf{n})=U(a=\frac{1}{2}(1-i3^{1/2}n_3),b=-\frac{3}{2}(in_1+n_2));n_1^2=n_2^2=n_3^2=\frac{1}{3}$.
(5) $\Gamma_1\sim\Gamma_7$; the notations used by Koster *et al.* (Ref. 9).

condition (6.1) reduces the five-dimensional spinor space spanned by $\{\phi(2,m);m=2,1,0,-1,-2\}$ into two- and three-dimensional invariant subspaces. The bases sets given in terms of x, y, and z in Table III are spherical harmonics belonging to exactly the same representations as the spinor bases (not just up to a similarity transformation). We have given only the most elementary basis sets for each irreducible representation. If necessary, additional basis sets based on spherical harmonics can easily be obtained by the method recently introduced by the author.[8] This method is different from the ordinary projection operator method. It uses the elementary basis sets instead of their irreducible representations.

7. THE GROUP T

This group is a subgroup of O. Proceeding as in the case of the spinor representation of O, we can show that the double group T' of the tetrahedral group T is characterized by the subgroup condition,

$$a^4=1 \quad \text{or } b^4=1 \quad \text{or } a^4=b^4=-\tfrac{1}{4}, \tag{7.1}$$

imposed on $U(a,b)\in SU(2)$. The projective set $\check{T}=\{S(\alpha)\}$ of $T=\{R(\alpha)\}$ is given by the unit element $U(1,0)$ and
(a) three $S(\pi\mathbf{n})=U(-in_3,-in_1-n_2)$'s with

$$n_1=1 \quad \text{or } n_2=1 \quad \text{or } n_3=1, \tag{7.2}$$

(b) eight $S((2\pi/3)\mathbf{n})=U(\frac{1}{2}(1-i3^{1/2}n_3)-\frac{3}{2}(in_1+n_2))$'s with

$$n_1{}^2=n_2{}^2=n_3{}^2=\tfrac{1}{3}. \tag{7.3}$$

The irreducible representations of T given in Table IV are obtained by reducing those of O given in Table III. In fact, by using the subgroup conditions (7.1) one can easily show that

$$\begin{aligned} &A_1,A_2\to A,\ E\to B_1+B_2,\\ &T_1,T_2\to T,\\ &E_{3/2}\cong E\times E_{1/2}\to B_1\times E_{1/2}+B_2\times E_{1/2}. \end{aligned} \tag{7.4}$$

8. CONCLUDING REMARKS

We have constructed the proper point groups D_∞, D_n, C_∞, C_n, O, and T and their general irreducible (vector and projective) representations by imposing simple polynomial equations on the matrix elements of 2×2 unitary matrices $U(a,b)$ with determinant 1. In the forthcoming paper we shall discuss the icosahedral group I which is the only finite proper point group that is not discussed in the present work.

If we know all the proper finite point groups G_R's, one can also construct all the improper finite point groups in a simple and well-defined manner on account of a well known theorem (see Ref. 6, p. 341). According to this theorem, an improper finite point group $\overline{G}_R$ takes one of the following two forms:

(a) $\overline{G}_R=\{P,JP'\}$, where J is the inversion, P is a subgroup of index 2 in a proper point group G_R, and $P'=G_R-P$.

(b) $\overline{G}_R=\{G_R,JG_R\}$, the direct product of the group of inversion $G_J=\{E,J\}$ and a proper point group G_R.

In case (a), $\overline{G}_R$ is isomorphic to G_R. Hence, the irreducible representations of $\overline{G}_R$ and their spinor bases are the same as those of G_R, since the inversion J leaves the spinor bases invariant. In case (b) the projective set of $\overline{G}_R$ be taken as $\check{G}_R=\{\check{G}_S,J\check{G}_S\}$ where $\check{G}_S$ is the projective set of G_R. Thus, all the irreducible (vector and projective) representations of $\check{G}_R$ follow from those of $\check{G}_R$ and those of J (see Refs. 2 and 5).

One may also state that the present results are more than sufficient to produce all the irreducible representations of the 32 point groups.

ACKNOWLEDGMENTS

The author wishes to thank Professor J. S. Rowlinson for his hospitality during which time this paper was written.

APPENDIX

We shall show that the set of condition (6.1) satisfies the group requirements (i)–(iv) given by (2.3) and (2.4). For this purpose, it is convenient to restate (6.1) as follows, combining with (2.1b),

$$\text{if } ab=0,\ a^8=1 \quad \text{or } b^8=1, \tag{A1}$$

$$\text{if } ab\neq0,\ a^4=b^4=\pm\tfrac{1}{4}. \tag{A2}$$

The proof will be given only for (iii) of (2.3) since the proof for the rest is trivial. It is to show that when (a_1,b_1) and (a_2,b_2) satisfy (A1) or (A2), so does the set (a,b) defined by

$$a=a_1a_2-b_1b_2^*,\quad b=a_1b_2+b_1a_2^*. \tag{A3}$$

When any one of the two sets (a_1,b_1) and (a_2,b_2) satisfies (A1), one can easily show that (a,b) satisfies (A1) or (A2). Thus we are left with the case

$$a_1{}^4=b_1{}^4=a_2{}^4=b_2{}^4=\pm\tfrac{1}{4}. \tag{A4}$$

Let $x=a_1a_2/(b_1b_2^*)$ then $x^4=1$. Since one can write

$$a^4=a_1{}^4a_2{}^4(x-1)^4,\quad b^4=a_1{}^4a_2{}^4(x+1)^4, \tag{A5}$$

we have $a^8=1$ or $b^8=1$ or $a^4=b^4=\pm\frac{1}{4}$ (Q.E.D.).

[1]M. Hamermesh, *Group Theory and Its Application to Physical Problems* (Addison–Wesley, Reading, MA, 1964); see also E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).
[2]S. K. Kim, J. Math. Phys. **21**, 1299 (1980).

[3]C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972), and references therein.
[4]E. Brown, Am. J. Phys. **38**, 704 (1970); N. B. Backhouse, Physica **70**, 505 (1973); S. L. Altmann, Mol. Phys. **38**, 489 (1979); S. L. Altmann and F. Palacio, Mol. Phys. **38**, 513 (1979).
[5]S. K. Kim, Mol. Phys. (in press).
[6]L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967), Appendix, p. 341.
[7]If the rotation vectors α and $-\alpha$ are equivalent with respect to the point group G_R, the rotation $R(\alpha)$ is called bilateral, otherwise it is called unilateral.
[8]S. K. Kim, Physica A **105**, 577 (1981); **106**, 521 (1981).
[9]G. F. Koster, J. O. Dimmock, R G. Wheeler, and H. Stetz, *Properties of the Thirty-two Point Groups* (M.I.T., Cambridge, 1963).

On the parametrization of certain finite groups and their representations

Anton Bovier[a)] and Daniel Wyler
Physikalisches Institut, Universität Bonn, D-5300, Bonn, West Germany

(Received 24 April 1981; accepted for publication 29 May 1981)

We discuss a convenient way of parametrizing certain finite groups. They comprise groups with abelian normal subgroups and central extensions thereof; they include the groups D_n, T, O, and their double groups, as well as the SU(3) subgroups T_n, $\Delta(3n^2)$, $\Delta(6n^2)$, $\Sigma(72)$, $\Sigma(216)$, $\Sigma(36)$ and their Z_3-extensions. The method allows for a rapid calculation of representations and coupling coefficients; in particular, it solves the "multiplicity problem" of nonsimply reducible groups. We illustrate the method treating the Hessian group $\Sigma(216)$ and its central extension.

PACS numbers: 02.20. + b

I. INTRODUCTION

For physical applications of finite groups it is important to have a good way of parametrizing their elements, which allows for convenient formulae for representations, Clebsch–Gordan coefficients, and other quantities. In recent papers[1–4] we have developed a convenient procedure for handling certain finite groups. As main results we derived analytic formulae for the group product rule, matrix elements of representations, and, most interesting, Clebsch–Gordan coefficients. In particular, this procedure led to a natural solution of the "multiplicity problem," that is, of the question how to label the several equivalent irreducible representations occurring in the Clebsch–Gordan series of a nonsimply reducible group. The groups investigated so far are the groups $\mathscr{D}(n)$ and $Q(n)$ (Z_n-central extensions of $Z_n \otimes Z_n$),[1] the Z-metacyclic groups,[2] including the dihedral groups D_n as SU(2) subgroups and the "trihedral" groups T_n as SU(3) subgroups,[3] the double dihedral groups $D_n{}^*$, the sequences $\Delta(3n^2)$ and $\Delta(6n^2)$ as SU(3) subgroups,[3,5] including the tetrahedral group T and the octhedral group $O \simeq S_4$ and their double groups T^* and O^*, respectively.[4]

In this paper we wish to explain our method in more detail and indicate what types of finite groups can be treated this way. We will show that the groups best treatable are those which have an abelian normal subgroups and are semidirect (or "almost" semidirect; see Sec. II for an explanation) products of this normal subgroup with its factor group. Then the method works straightforwardly, if representations, etc. of the factor group are known. Of course, it is desirable to treat the factor group by the same method.

At the expense of less transparent results (and more work), the above requirements can be considerably weakened. In particular, the method is also applicable to central extensions of the groups mentioned above, for instance, if the normal subgroup is a central extension of an abelian group. As an illuminating example for both cases, we will apply the method to the Hessian group[6] of order 216 (and its subgroups of order 72 and 36, respectively) and to their Z_3-central extensions. All these groups are SU(3) subgroups[5,7]; in fact, all subgroups of SU(3) are either covered by the present method or are simple.

[a)]Supported by Studienstiftung des deutschen Volkes.

We hasten to add that we do not involve new mathematical concepts here. Rather, we show that known mathematical procedures lead to eminently practical ways for handling many finite groups.

The remainder of this paper is organized as follows. In Sec. II we explain how the groups are parametrized and how one finds representations. Section III contains a detailed treatment of the Hessian group and its subgroups; in Sec. IV we treat the Z_3-extension.

II. DESCRIPTION OF THE METHOD

The basic idea in our treatment of finite groups is to parametrize the group in terms of abelian subgroups. This means we label any element of the group by indices which take values in some abelian group and express the product of the group elements as a function of their indices:

$$g_{\alpha_1,\dots,\alpha_n} \cdot g_{\alpha'_1,\dots,\alpha'_n} = g_{f_1(\alpha_1,\dots,\alpha'_n),\dots,f_n(\alpha_1,\dots,\alpha'_n)} . \tag{2.1}$$

As long as the functions f_i are reasonable [necessary conditions such as the associativity of (2.1) and the existence of an inverse element restrict the f_i considerably], this way of expressing the product offers sizeable advantages over group tables in most calculations. It is even more useful if one is able to express the representations and their matrix elements in a similar way as functions of abelian group parameters:

$$[T^{[i_1,\dots,i_k]}(g_{\alpha_1,\dots,\alpha_n})]_{r'_1,\dots,r'_l;\, r_1,\dots,r_l} = F(i_1,\dots,i_k, r_1,\dots, r'_l, \alpha_1,\dots,\alpha_n), \tag{2.2}$$

where the parameters i_γ label the representations and the r_i, r'_i the basis vectors; and they take values from the abelian groups used to parametrize the group. Equation (2.2) is especially suitable for evaluating Clebsch–Gordan coefficients or similar constructions. To do this, one has to perform sums over the group manifold, and they are far easier done as multiple sums of the respective parameter ranges. In all groups treated so far this way, the multiplicity problem of nonsimply reducible groups is solved automatically as a consequence of the parametrization.[1–4]

How does one find the most suitable abelian parametrization for a given group? The first thing to do is to find normal subgroups (NSG) and corresponding factor groups (FG) (if a group has no proper NSG, e.g., is simple, the method is not applicable). There may be several NSG's, and we will, in principle, first look at the smallest one. In practice,

 0022-2488/81/102108-05$01.00

however, it is not necessary to start with the smallest one. Let the starting NSG be N and the corresponding factor group G/N be F.

The simplest possibility is (beside trivial factorization) that $G = N \,\mathrm{s}\, F$, s denoting the semidirect product. In this case we can label the group elements by

$$G \ni g^{\alpha}_{\beta};\quad \alpha\in F,\ \beta\in N$$

with the group rule

$$g^{\alpha}_{\beta},\ g^{\alpha'}_{\beta'} = g^{\alpha\cdot\alpha'}_{\beta\cdot M(\alpha)\beta'},\tag{2.3}$$

where M is a mapping $F\to\mathrm{Aut}(N)$ and can be regarded as a representation of F over the group N.

The method works best (almost trivially) if N is abelian; among the cases considered so far are $N = Z_n$, $N = Z_n \otimes Z_n$ (and its central extensions, see below).[1,4] Generalizations to $N = Z_n \otimes Z_n \cdots \otimes Z_n$ (k times) and central extensions thereof are straightforward. In this case one identifies the group algebra of N with the K-dimensional module over the ring Z_n and gives $M(\alpha)$ in matrix form. Further simplifications always occur when n is prime, for then Z_n is a field and $Z_n{}^k$ a vector space. Easy cases are given if F acts transitively on this vector space (transitivity means that there are only two orbits of F in N, namely $\{0\}$, consisting of the identity of N and one other orbit containing all other elements). The orbit $\{\beta\}$ of F in N defined as the set of elements β' for which $\beta' = M(\alpha)\beta$ for some $\alpha\in F$. If N is a central extension of the group $Z_n \otimes Z_n \otimes \cdots \otimes Z_n$ (k times) and if G is an "almost" semidirect product of N with F (see below for a definition), one can still apply the procedure. Let us write $N = Z_l \circ ((Z_n)^k)$ for the l-central extension of $Z_n{}^k$, and parametrize the elements of N by

$$n_{\boldsymbol{\beta},i},\quad \boldsymbol{\beta}\in(Z_n)^k,\quad i\in Z_l.$$

The group rule for N can now be written as

$$n_{\boldsymbol{\beta},i}\cdot n_{\boldsymbol{\beta}',i'} = n_{\boldsymbol{\beta}+\boldsymbol{\beta}',i+i'+(\boldsymbol{\beta},\boldsymbol{\beta}')},\tag{2.4}$$

where $(\ ,\ )$ is some quadratic form in $Z_n{}^k$. Now, if $M(\alpha)$ is unitary, i.e., $(M(\alpha)\boldsymbol{\beta}, M(\alpha)\boldsymbol{\beta}') = (\boldsymbol{\beta},\boldsymbol{\beta}')$ for all $\alpha\in F$, the rule for the entire group G takes the form

$$g^{\alpha}_{\boldsymbol{\beta},i}\cdot g^{\alpha'}_{\boldsymbol{\beta}';i'} = g^{\alpha\cdot\alpha'}_{\boldsymbol{\beta}+M(\alpha)\boldsymbol{\beta}',i+i'+(\boldsymbol{\beta},M(\alpha)\boldsymbol{\beta}')+f(\alpha,\alpha')},\tag{2.5}$$

where $f(\alpha,\alpha')$ is a function which may be necessary in order to obtain the correct group rule. We will call departures from the semidirect product rule such as the additional term $f(\alpha,\alpha')$ almost semidirect products. They produce no further difficulties and result in additional phases in the repreentations only.[4] If $M(\alpha)$ is not unitary under $(\ ,\ )$, (2.5) is not associative. To render it so, one might reparametrize N such that the new form $(\ ,\ )^{\sim}$ is invariant under M or add correction terms; an example is O^*.[4]

To take full advantage of the method, one analyzes the FG F in the same way as G, and so on. At the end of the procedure one has the decomposition (provided the group allows for it!)

$$G = N_1\circ(N_2\circ\cdots\circ N_l\circ F)\cdots),$$

where the N_i are abelian, or central extensions of abelian groups and F simple. (As mentioned before, it is not always practical to go all the way in the decomposition of G; often several NSG's or big FG's are already known and can be used.) G is now parametrized by indices β_i from the N_i and $\alpha\in F$, and the product rule is

$$g^{\alpha}_{\beta_1,\ldots,\beta_l}\cdot g^{\alpha'}_{\beta'_1,\ldots,\beta'_l}$$
$$= g^{\alpha\cdot\alpha'}_{\beta_1+M_1(\alpha)\beta'_1,\ \beta_2+M_2(\alpha,\beta_1)\beta'_2,\ldots,\ \beta_l+M_l(\alpha,\beta,\ldots,\beta_{l-1})\beta'_l}.\tag{2.6}$$

Here we have written only a semidirect product, but generalizations of the type indicated in Eq. (2.5) can be included. At this stage, G is fully described.

Equation (2.6) is now very helpful in deriving the representations. Take first a group G with abelian NSG N (such as $Z_n{}^k$) and assume the representations of $F = G/N$ are known. Equation (2.6) now reduces to Eq. (2.3); the representations of G are now found trivially, using the method of induced representations.[8] In this formulation of G it is extremely easy to use, and one is, in fact, led to it automatically.

We label the basis vectors of a representation space by indices taken from N. It is straightforward to see that

$$T(g^{\alpha}_{\boldsymbol{\beta}}):\quad T(g^{\alpha}_{\boldsymbol{\beta}})e_{\mathbf{b}} = \omega^{\mathbf{b}\cdot M^{-1}(\alpha)\boldsymbol{\beta}}e_{\bar{M}^{-1}(\alpha)\mathbf{b}}\tag{2.7}$$

(where $\omega = e^{2\pi i/n}$, $\boldsymbol{\beta},\mathbf{b}\in Z^{\,n}_{k}$; $\bar{M}$ denotes the transpose of M and the dot the ordinary scalar product) is a representation of G. Such a T is reducible and not a complete system of representations. To reduce it, we must find the minimal invariant subspaces of $M(\alpha)$ in N, which are, of course, the orbits of F in N. Once the orbits are calculated, a set of representatives of each orbit may be selected to label the representations of the form (2.7). Let R be this set. The basis vectors of the representation $T^{\mathbf{r}}$, $\mathbf{r}\in R$, can then be labelled by $e_{\bar{M}(\alpha)\mathbf{r}}$. Obviously, the number of basis vectors in the representation $T^{\mathbf{r}}$ (its dimension) is equal to the number of elements in the orbit $\{\mathbf{r}\}$.

All representations of G are now found by taking the tensor products of the representations $T^{\mathbf{r}}$ and the representations of the little group of $\{\mathbf{r}\}$ (the little group of $\{\mathbf{r}\}$ is that subgroup of F which leaves $\mathbf{r}$ fixed).

If N is not abelian (and G is not just a semidirect product), but a central extension of an abelian group, with a product rule of the form (2.5), we cannot give such an easy prescription. In this case, one must first find the representations of N.

These representations are then used to construct one faithful representations of G of the same dimension as that of the representation of N. This is the only difficult step. All other representations are then given as tensor products of these representations and representations of F. The procedure will be exemplified in Sec. IV.

From the irreducible representations obtained this way, one may calculate Clebsch–Gordan coefficients using standard methods.[9] Due to our parametrization, performing sums over the group is simplified. The resulting expressions are very transparent; in all cases considered we have always been able to label several identical representations occurring in a Clebsch–Gordan series (multiplicity problem) in a natural and self-suggesting way.[10] We consider this an important indication of the advantages of our method.

III. THE HESSIAN GROUP $\Sigma(216)$

The intention of this and the following section is twofold. First we will use a rather complicatedly structured group as an illustration of the method outlined in the preceeding section. Second, we will treat a group which is possibly important for applications. We will thereby complete our treatment of the finite SU(3) subgroups[3] (the remaining subgroups are simple).[7]

The Hessian group is generated by the two elements T and U satisfying[6]

$$T^3 = U^3 = (TU)^4 = 1, \quad (TUT)^2U = U(TUT)^2. \tag{3.1}$$

As is known,[5] this group possesses an abelian normal subgroup of order 9 and factor group T^* of order 24. Starting from this, one finds that $\Sigma(216)$ can be decomposed as

$$\Sigma(216) = (Z_3 \otimes Z_3) \text{ s } T^* = (Z_3 \otimes Z_3) \text{ s } (Q \text{ s } Z_3)$$
$$= (Z_3 \otimes Z_3) \text{ s } (Z_2 \circ (Z_2 \otimes Z_2) \text{ s } Z_3).$$

Here, $\otimes$ is the direct, s is the semidirect product, $\circ$ what we termed almost semidirect product [see the previous section, after Eq. (2.5)], Q is the quaterion group, and we have used the fact that $T^* = Q \text{ s } Z_3$.[4] Following Sec. II, we label the elements of the Hessian group by six indices:

$$g_{\boldsymbol\tau}^{\epsilon,\boldsymbol\alpha,\gamma}, \quad \boldsymbol\alpha = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \quad \boldsymbol\tau = \begin{pmatrix}\tau_1\\ \tau_2\end{pmatrix}, \quad \begin{cases}\alpha,\beta,\gamma \in Z_2,\\ \epsilon,\tau_1,\tau_2 \in Z_3.\end{cases} \tag{3.2}$$

The indices ϵ, $\boldsymbol\alpha$, γ label the factor group T^*, for which the product rule is[4]

$$g^{\epsilon,\boldsymbol\alpha,\gamma}\cdot g^{\epsilon',\boldsymbol\alpha',\gamma'} = g^{\epsilon+\epsilon',\boldsymbol\alpha+M_2^{\epsilon}\boldsymbol\alpha',\gamma+\gamma'+(\boldsymbol\alpha,M_2^{\epsilon}\boldsymbol\alpha')}, \tag{3.3}$$

with $M_2 = \left(\begin{smallmatrix}1 & 1\\ 1 & 0\end{smallmatrix}\right)$ and $(\boldsymbol\alpha,\boldsymbol\alpha') = \alpha\alpha' + \beta\beta' + \alpha\beta'$.

In order to find the rule for the whole group, we have to find a two-dimensional representation of T^* over Z_3. Some inspection shows that the following works:

$$R(\epsilon,\boldsymbol\alpha,\gamma) = P^{\beta}Q^{\alpha}(-1)^{\gamma}M_3^{\epsilon},$$

with

$$Q = \begin{pmatrix}-1 & 1\\ 1 & 1\end{pmatrix}, \quad P = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}, \quad M_3 = \begin{pmatrix}-1 & -1\\ 1 & 0\end{pmatrix}. \tag{3.4}$$

As matrices with entries in Z_3 they satisfy

$$Q^2 = P^2 = -\mathbb{R}, \quad M^3 = \mathbb{R}, \quad PQ = -QP,$$
$$MQ = PQM, \quad MP = QM, \tag{3.5}$$

which reflects the product rule (3.2) of T^*. Observe, that because $\alpha, \beta \in Z_2$, we must keep in mind that

$$Q^{\alpha}Q^{\alpha'} = Q^{\alpha+\alpha'}\cdot(-1)^{\alpha\alpha'}, \quad \text{etc.}$$

The rule for $\Sigma(216)$ is now

$$g_{\boldsymbol\tau}^{\epsilon,\boldsymbol\alpha,\gamma}\cdot g_{\boldsymbol\tau'}^{\epsilon',\boldsymbol\alpha',\gamma'} = g_{\boldsymbol\tau+P^{\beta}Q^{\alpha}(-1)^{\gamma}M^{\epsilon}\boldsymbol\tau'}^{\epsilon+\epsilon',\boldsymbol\alpha+M_2^{\epsilon}\boldsymbol\alpha',\gamma+\gamma'+(\boldsymbol\alpha,M_2^{\epsilon}\boldsymbol\alpha')}. \tag{3.6}$$

Setting

$$T = g_{\binom{0}{0}}^{2\binom{0}{0}0}, \quad U = g_{\binom{1}{1}}^{1\binom{1}{1}1}$$

and using (3.6) one indeed recovers (3.1).

From (3.6) one easily obtains the rules for the subgroups $\Sigma(72)$ and $\Sigma(36)$:

$$\Sigma(72) = (Z_3 \otimes Z_3) \text{ s } Q \quad (\epsilon \equiv 0),$$
$$g_{\boldsymbol\tau}^{\boldsymbol\alpha,\gamma}\cdot g_{\boldsymbol\tau'}^{\boldsymbol\alpha',\gamma'} = g_{\boldsymbol\tau+P^{\beta}Q^{\alpha}(-1)^{\gamma}\boldsymbol\tau'}^{\boldsymbol\alpha+\boldsymbol\alpha',\gamma+\gamma'+(\boldsymbol\alpha,\boldsymbol\alpha')}, \tag{3.7}$$
$$\Sigma(36) = (Z_3 \otimes Z_3) \text{ s } Z_4, \quad (\epsilon,\alpha \equiv 0),$$
$$g_{\boldsymbol\tau}^{\beta,\gamma}\cdot g_{\boldsymbol\tau'}^{\beta',\gamma'} = g_{\boldsymbol\tau+P^{\beta}(-1)^{\gamma}\boldsymbol\tau'}^{\beta+\beta',\gamma+\gamma'+\beta\beta'}. \tag{3.8}$$

The next step of reduction leads to $\Sigma(18) = (Z_3 \otimes Z_3) \text{ s } Z_2$, which is isomorphic to $D_3 \otimes Z_3$ and has been studied before.[2]

Let us now derive the representations of the Hessian group. As we already know those of T^*, the induced representation can be written down immediately.

T^* acts transitively on $Z_3 \otimes Z_3$. Thus, there are only two orbits, $\{\binom{0}{0}\}$ of length 1 and $\{\binom{1}{1}\}$ of length 8 (the "length" of an orbit is the number of its elements). The little group of $\{\binom{0}{0}\}$ is, of course, T^*, while that one of $\{\binom{1}{1}\}$ is Z_3. Thus, $\Sigma(216)$ has as its representation those of T^* and three eight-dimensional ones. The first ones are

$$[T^{[0,\mathbf{m},i,\kappa]}(g_{\boldsymbol\tau}^{\epsilon,\boldsymbol\alpha,\gamma})]_{\bar M_2^{\alpha'}\mathbf{m},\kappa\rho;\bar M_2^{\alpha}\mathbf{m},\kappa\rho}$$
$$= (-1)^{\kappa\gamma}\omega^{i\epsilon}(-1)^{\bar M^{\alpha}\mathbf{m}\cdot M^{\gamma}\boldsymbol\alpha}\delta_{\bar M^{\alpha'}\mathbf{m},\bar M^{\alpha}{}^{\gamma}\mathbf{m}}$$
$$\times\left(\delta_{\kappa,1}\left[\begin{pmatrix}0 & i\\ i & 0\end{pmatrix}^{\beta}\begin{pmatrix}-i & 0\\ 0 & i\end{pmatrix}^{\alpha}\Gamma^{\epsilon}\right]_{\rho\rho'} + \delta_{\kappa,0}\right), \tag{3.9}$$

where $\mathbf{m} = \binom{0}{0}$ or $\binom{1}{1}$, $i \in Z_3$, $\kappa \in Z_2$ and

$$\kappa = 0, \quad i = 0 \quad \text{if } \mathbf{m} = \tbinom{1}{1},$$
$$\kappa = 0,1, \quad i = 0,1,2, \quad \text{if } \mathbf{m} = \tbinom{0}{0}, \tag{3.10}$$
$$\Gamma = \frac{e^{i\pi/4}}{\sqrt{2}}\begin{pmatrix}-1 & 1\\ i & i\end{pmatrix},$$
$$\omega = e^{2\pi i/3}.$$

The eight-dimensional representations are labelled by the representative $\binom{1}{1}$ and by the Z_3-indices k of its little group. They are

$$[T^{[\binom{1}{1},k]}(g_{\boldsymbol\tau}^{\epsilon,\boldsymbol\alpha,\gamma})]_{\bar P^{b'}Q^{a'}(-1)^{c'}\binom{1}{1},\bar P^{b}Q^{a}(-1)^{c}\binom{1}{1}}$$
$$= \omega^{k\epsilon}\omega^{\bar P^{b}Q^{a}(-1)^{c}\binom{1}{1}\cdot(-1)^{\gamma}M^{\epsilon}\hat Q^{\alpha}\hat P^{\beta}\boldsymbol\tau}$$
$$\times\delta_{\bar P^{b'}Q^{a'}(-1)^{c'}\binom{1}{1},\hat{\bar P}^{\beta}\hat Q^{\alpha}(-1)^{\gamma}M^{\epsilon}P^{b}Q^{a}(-1)^{c}\binom{1}{1}}, \tag{3.11}$$

where $\bar A$ denotes the transpose of A and $\hat A$ its inverse. Notice that $\hat Q^{\alpha} = (-1)^{\alpha}Q^{\alpha}$. Thus (3.11) is equal to

$$T = \omega^{k\epsilon}\omega^{\bar P^{b}Q^{a}(-1)^{c}\binom{1}{1}\cdot(-1)^{\gamma+\alpha+\beta}M^{\epsilon}Q^{\alpha}P^{\beta}\boldsymbol\tau}$$
$$\times\delta_{\bar P^{b'}Q^{a'}(-1)^{c'}\binom{1}{1},\bar P^{(\boldsymbol\alpha+M^{\epsilon}\mathbf{a})_2}Q^{(\boldsymbol\alpha+M^{\epsilon}\mathbf{a})_1}(-1)^{\gamma+\alpha+\beta+c+(\boldsymbol\alpha,M^{\epsilon}\mathbf{a})}\binom{1}{1}} \tag{3.12}$$

in an obvious notation.

The corresponding formulae for $\Sigma(72)$ and $\Sigma(36)$ follow easily. For $\Sigma(72)$ the eight-dimensional representation remains irreducible; it is given by (3.12) with $k \equiv 0$ and $\epsilon \equiv 0$. The representations are just those of Q and are given by[1]

$$[T^{[i,j,k]}(g_{\boldsymbol\tau}^{\boldsymbol\alpha,\gamma})]_{a,b} = \delta_{ka,k(b+\beta)}$$
$$\times(-1)^{k(\gamma+\alpha b-(\alpha^2+\beta^2)/2)+i\alpha+j\beta}, \tag{3.13}$$

where $k = 0,1$, and $i,j = 0$ for $k = 1$ and $i,j = 0,1$ for $k = 0$.

$\Sigma(72)$ is the smallest group with an eight-dimensional irreducible representation.[11]

For $\Sigma(36)$, the eight-dimensional representation is no longer irreducible, but splits into two four-dimensional ones. They correspond to the orbits

$$\left\{\begin{pmatrix}1\\ 1\end{pmatrix}\right\} = \left\{\begin{pmatrix}1\\ 1\end{pmatrix}\begin{pmatrix}-1\\ -1\end{pmatrix}\begin{pmatrix}1\\ -1\end{pmatrix}\begin{pmatrix}-1\\ 1\end{pmatrix}\right\},$$
$$\left\{\begin{pmatrix}1\\ 0\end{pmatrix}\right\} = \left\{\begin{pmatrix}1\\ 0\end{pmatrix}\begin{pmatrix}-1\\ 0\end{pmatrix}\begin{pmatrix}0\\ -1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix}\right\}$$

of Z_4 in $Z_3 \otimes Z_3$. Thus we have

$$[T^{[\mathbf{t}]}(g_{\boldsymbol{\tau}}^{\beta\gamma})]_{\bar{P}^{b'}(-1)^{c'}\mathbf{t},\bar{P}^{b}(-1)^{c}\mathbf{t}} = \omega^{\bar{P}^{b}(-1)^{c}\mathbf{t}\cdot(-1)^{\gamma+\beta}P^{\beta}\boldsymbol{\tau}} \cdot \delta_{\bar{P}^{b'}(-1)^{c'}\mathbf{t},\bar{P}^{b+\beta}(-1)^{c+\gamma+\beta+b\beta}\mathbf{t}}$$

(four-dimensional) (3.14)

$$T^{[k,l]}(g_{\boldsymbol{\tau}}^{\beta,\gamma}) = i^{2\beta l + k(2\gamma + \beta^2)} \quad \text{(one-dimensional)},$$

where $\mathbf{t} = \binom{1}{1}, \binom{1}{0}$, $k,l = 0,1$.

Since the characters are often important in applications, we given them in an appendix explicitly.

IV. THE EXTENDED HESSIAN GROUP (Z_3 EXTENSION)

The Hessian group $\Sigma(216)$ is not only a subgroup of SU(3) but also of SU(3)/Z_3, because the Z_3-center is not contained in it. One can make it a subgroup of SU(3) only, by adding this Z_3-center through a central extension with Z_3. The resulting group will be denoted by $\Sigma^*(216)$ and is of order 648. The same construction can be done with its subgroups. As shown in Sec. III, $\Sigma(216)$ is a semidirect product of T^* with the abelian normal subgroup $Z_3 \otimes Z_3$. In order to extend it, we first extend $Z_3 \otimes Z_3$ to $Q(3)$.[1] This group is of the form $Z_3 \circ (Z_3 \otimes Z_3)$, where $\circ$ is the almost semidirect product described before. The elements of $Q(3)$ are labelled by $g_{\boldsymbol{\tau},k}$ with $\boldsymbol{\tau} \in Z_3^2$, $k \in Z_3$, and the rule is

$$g_{\boldsymbol{\tau},k} \cdot g_{\boldsymbol{\tau}',k'} = g_{\boldsymbol{\tau}+\boldsymbol{\tau}',k+k'+\langle\boldsymbol{\tau},\boldsymbol{\tau}'\rangle}, \tag{4.1}$$

where $\langle\boldsymbol{\tau},\boldsymbol{\tau}'\rangle = \tau_1\tau_2' - \tau_2\tau_1'$.[12] The elements $g_{0,k}$ form the center.

$Q(3)$ is the normal subgroup of the extended Hessian group, with factor group T^*. To obtain the multiplication rule, first restrict T^* to Q [that is, consider the extension of $\Sigma(72)$]. Since the orders of Q and $Q(3)$ have no common prime factor, the extended $\Sigma(72)$ is a semidirect product of Q with $Q(3)$ (theorem of Zassenhaus).[13] Using the matrices P and Q (and, later on, also M) from the previous sections, the product rule is

$$g_{\boldsymbol{\tau},k}^{\boldsymbol{\alpha},\gamma} \cdot g_{\boldsymbol{\tau}',k'}^{\boldsymbol{\alpha}',\gamma'} = g_{\boldsymbol{\tau}+P^{\beta}Q^{\alpha}(-1)^{\gamma}\boldsymbol{\tau}',k+k'+\langle\boldsymbol{\tau},P^{\beta}Q^{\alpha}(-1)^{\gamma}\boldsymbol{\tau}'\rangle}^{\boldsymbol{\alpha}+\boldsymbol{\alpha}',\gamma+\gamma'+(\boldsymbol{\alpha},\boldsymbol{\alpha}')} \cdot \tag{4.2}$$

(4.2) is associative, because P and Q (and also M) are unitary with respect to $\langle\ ,\ \rangle$.

Next we include ϵ. Obviously, the $\langle\ ,\ \rangle$ is just changed to $\langle\ ,P^{\beta}Q^{\alpha}(-1)^{\gamma}M^{\epsilon}\rangle$, but one can add ϵ-dependent terms in various ways. We will consider two possibilities, defining two different extensions of $\Sigma(216)$:

$$g_{\boldsymbol{\tau},k}^{\epsilon,\boldsymbol{\alpha},\gamma} \cdot g_{\boldsymbol{\tau}',k'}^{\epsilon',\boldsymbol{\alpha}',\gamma'} = g_{\boldsymbol{\tau}+P^{\beta}Q^{\alpha}(-1)^{\gamma}M^{\epsilon}\boldsymbol{\tau}',k+k'+\langle\boldsymbol{\tau},P^{\beta}Q^{\alpha}(-1)^{\gamma}M^{\epsilon}\boldsymbol{\tau}'\rangle+\kappa\epsilon\epsilon'(\epsilon+\epsilon'+1)}^{\epsilon+\epsilon'+\boldsymbol{\alpha}+M_2^{\epsilon}\boldsymbol{\alpha}',\gamma+\gamma'+(\boldsymbol{\alpha},M_2^{\epsilon}\boldsymbol{\alpha}')} \cdot \tag{4.3}$$

with $\kappa = 0,1$.

The representations are constructed as sketched at the end of Sec. II. $Q(3)$ is not abelian; it has two three-dimensional and nine one-dimensional representations. They are labelled by

3-dim: $[\mathbf{0},k]$, $k = 1,2$,

1-dim: $[\mathbf{t},0]$, $\mathbf{t} \in Z_3^2$.

The one-dimensional representations are those of $Z_3 \otimes Z_3$ and give rise to the ordinary representations of $\Sigma(216)$, given in (3.9)–(3.12); they are not faithful, because all elements differing just by the index k have the same representations.

In order to find all other representations, we must find the faithful three-dimensional representations based on the three-dimensional ones of Q. Then, the representations not given as representations of $\Sigma(216)$ are just the tensor products of those of T^* with the faithful three-dimensional ones. The representations of T^* have the dimensions one (three times), two (three times), and three (once). Thus there will be six three-dimensional, six six-dimensional and two nine-dimensional representations. They are labelled by the usual T^* indices and a number $k = 1,2$.

Let us denote the representation matrices of the faithful three-dimensional representation ($k = 1$, say) as follows:

$$\begin{aligned}
&T(g_{\binom{0}{0}0}^{1\binom{0}{0}0}) = E, \quad T = (g_{\binom{0}{0}0}^{0\binom{0}{0}1}) = G,\\
&T = (g_{\binom{0}{0}0}^{0\binom{1}{0}0}) = A_1, \quad T(g_{\binom{1}{0}0}^{0\binom{0}{0}0}) = T_1.\\
&T(g_{\binom{0}{0}0}^{0\binom{0}{1}0}) = A_2, \quad T(g_{\binom{0}{1}0}^{0\binom{0}{0}0}) = T_2,\\
&T(g_{\binom{0}{0}k}^{0\binom{0}{0}0}) = \omega^k \mathbb{1}, \quad \omega = e^{2\pi i/3},
\end{aligned} \tag{4.4}$$

One must determine these matrices such that they satisfy the product rule (4.3); for example

$$\begin{aligned}
&E^3 = 1 \quad (\kappa = 0), \quad E^9 = 1 \quad (\kappa = 1),\\
&A_1^2 = A_2^2 = G, \quad G^2 = 1, \quad T_1^3 = T_2^3 = 1,\\
&A_1A_2 = GA_2A_1, \quad EA_2 = A_1E, \quad T_1T_2 = \omega^2 T_2T_1.
\end{aligned}$$

However, some care is required when considering the representation of elements such as $g_{\binom{1}{1}0}^{0\binom{0}{0}0}$, since they are not just products of the form $g_{\binom{1}{0}0}^{0\binom{0}{0}0} \cdot g_{\binom{0}{1}0}^{0\binom{0}{0}0}$, but are of the form

$$g_{\binom{1}{0}0}^{0\binom{0}{0}0} \cdot g_{\binom{0}{1}0}^{0\binom{0}{0}0} \cdot g_{\binom{0}{0}2}^{0\binom{0}{0}0}.$$

It is easy to see that the representation of $g_{\binom{\alpha}{\beta}0}^{0\binom{0}{0}0}$ is given by $T_1^{\alpha} T_2^{\beta} \omega^{-\alpha\beta}$. As for the Ω indices, no such problem arises. Using the known representations of $Q(3)$,[1] the above equations determine the matrices in (4.4). After some algebra we find

$$\begin{aligned}
&E = \frac{(e^{4\pi i/9})^{\kappa}}{i\sqrt{3}} \begin{pmatrix} 1 & \omega^2 & \omega^2 \\ 1 & 1 & \omega \\ 1 & \omega & 1 \end{pmatrix}, \quad A_1 = \frac{i}{\sqrt{3}} \begin{pmatrix} 1 & \omega & \omega \\ \omega^2 & \omega^2 & \omega \\ \omega^2 & \omega & \omega^2 \end{pmatrix},\\
&A_2 = \frac{1}{i\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \quad T_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},\\
&G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad T_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega^2 & 0 \\ 0 & 0 & \omega \end{pmatrix}.
\end{aligned} \tag{4.5}$$

The three-dimensional representation is then given by

$$T^{(1)}(g_{\binom{\tau_1}{\tau_2}k}^{\epsilon\binom{\alpha_1}{\alpha_2}\gamma}) = \omega^{k-\tau_1\tau_2} T_1^{\tau_1} T_2^{\tau_2} G^{\gamma} A_1^{\alpha_1} A_2^{\alpha_2} E^{\epsilon}, \tag{4.6}$$

while the $k = 2$ representation is just the complex conjugated one. As mentioned, all representations of the extended Hessian group are tensor products of the representations of

T^* and (4.6) and its conjugate one, along with the representations (3.9)–(3.12) of $\Sigma(216)$ itself. Formula (4.6) contains, in an obvious manner, the representations of the extensions of $\Sigma(72)$ and $\Sigma(36)$.

From (4.5) we notice that all matrices have determinant $= 1$, except E; in fact, $\det E = 1$ if $\kappa = 1$, and $\det E = \omega$, if $\kappa = 0$. Thus, of the two extensions defined by (4.3) only the case $\kappa = 1$ leads to a subgroup of SU(3).

ACKNOWLEDGMENTS

The authors are indebted to Mihail Marcu for many helpful discussions. One of us (A.B.) is especially grateful to Uwe Abresch for interesting discussions on mathematical questions.

APPENDIX

In this appendix we give the formulas for the characters of the groups $\Sigma(216)$, $\Sigma(72)$, and $\Sigma(36)$ as derived from the expressions for the representations in Sec. III. They coincide with the tables given by Fairbairn *et al.*[5]

1. $\Sigma(216)$

$$\begin{aligned}\chi^{[(\frac{1}{1}),k]}(g_{\boldsymbol{\tau}}^{\epsilon,\alpha,\gamma}) &= \delta_{\epsilon,0}\delta_{\alpha,0}\delta_{\beta,0}\delta_{\gamma,0}(8\delta_{\boldsymbol{\tau},\mathbf{0}} - \bar{\delta}_{\boldsymbol{\tau},\mathbf{0}}) \\ &+ \delta_{\alpha,0}\delta_{\beta,0}\delta_{\gamma,0}(\delta_{\epsilon,1}\omega^k f(\tau_1) + \delta_{\epsilon,2}\omega^{2k} f(\tau_2)) \\ &+ \delta_{\alpha,0}\delta_{\beta,1}\delta_{\gamma,1}(\delta_{\epsilon,1}\omega^k f(\tau_1 - \tau_2) + \delta_{\epsilon,2}\omega^{2k} f(\tau_2 - \tau_1)) \\ &+ \delta_{\alpha,1}\delta_{\beta,0}\delta_{\gamma,1}(\delta_{\epsilon,1}\omega^k f(\tau_1 + \tau_2) + \delta_{\epsilon,2}\omega^{2k} f(\tau_1)) \\ &+ \delta_{\alpha,1}\delta_{\beta,1}\delta_{\gamma,1}(\delta_{\epsilon,1}\omega^k f(\tau_2) + \delta_{\epsilon,2}\omega^{2k} f(\tau_1 + \tau_2)),\end{aligned} \tag{A1}$$

where

$$f(\tau) = 2\delta_{\tau,0} - \bar{\delta}_{\tau,0}$$

and

$$\bar{\delta}_{\tau,0} = \begin{cases} 0, & \text{if } \tau = 0, \\ 1, & \text{if } \tau \neq 0. \end{cases}$$

$$\begin{aligned}\chi^{[0,\mathbf{m},i,\kappa]}(g_{\boldsymbol{\tau}}^{\epsilon,\boldsymbol{\alpha},\gamma}) &= \sum_{a,\rho} (-1)^{\kappa\gamma}\omega^{i\epsilon}(-1)^{\bar{M}^a\mathbf{m}\cdot M^{\gamma}\boldsymbol{\alpha}}[\delta_{\mathbf{m},\mathbf{0}} + \bar{\delta}_{\mathbf{m},\mathbf{0}}\delta_{\gamma,a}] \\ &\times\left[\delta_{\kappa,1}\left[\begin{pmatrix}0 & i\\ i & 0\end{pmatrix}^{\beta}\begin{pmatrix}-i & 0\\ 0 & i\end{pmatrix}^{\alpha}\Gamma^{\epsilon}\right]_{\rho,\rho} + \delta_0^{\kappa}\right].\end{aligned} \tag{A2}$$

Specialized, this gives

$$\chi^{[0,\mathbf{m},0,0]}(g_{\boldsymbol{\tau}}^{\epsilon,\boldsymbol{\alpha},\gamma}) = \delta_{\epsilon,0}(3\delta_{\boldsymbol{\alpha},\mathbf{0}} - \bar{\delta}_{\boldsymbol{\alpha},\mathbf{0}}), \tag{A3}$$

$$\chi^{[0,\mathbf{0},i,0]}(g_{\boldsymbol{\tau}}^{\epsilon,\boldsymbol{\alpha},\gamma}) = \omega^{i\epsilon}, \tag{A4}$$

$$\chi^{[0,0,i,l]}(g_{\boldsymbol{\tau}}^{\epsilon,\boldsymbol{\alpha},\gamma}) = (-1)^{\gamma}\omega^{i\epsilon}(2\delta_{\epsilon,0}\delta_{\alpha,0}\delta_{\beta,0} - \delta_{\epsilon,1} + \delta_{\epsilon,2}). \tag{A5}$$

2. $\Sigma(72)$

$$\chi^{[(\frac{1}{1})]}(g_{\boldsymbol{\tau}}^{\boldsymbol{\alpha},\gamma}) = \delta_{\alpha,0}\delta_{\beta,0}\delta_{\gamma,0}(8\delta_{\boldsymbol{\tau},\mathbf{0}} - \bar{\delta}_{\boldsymbol{\tau},0}), \tag{A6}$$

$$\chi^{[\mathbf{0},i,j,k]}(g_{\boldsymbol{\tau}}^{\boldsymbol{\alpha},\gamma}) = \frac{2}{\bar{k}}\delta_{k\alpha,0}\delta_{k\beta,0}(-1)^{i\alpha + i\beta + k\gamma}, \tag{A7}$$

where

$$\bar{k} = \begin{cases} 1, & \text{for } k = 1, \\ 2, & \text{for } k = 0. \end{cases}$$

3. $\Sigma(36)$

$$\chi^{[(\mathbf{t})]}(g_{\boldsymbol{\tau}}^{\beta,\gamma}) = \delta_{\beta,0}\delta_{\gamma,0}(6\delta_{\boldsymbol{\tau},0} + \delta_{\boldsymbol{\tau}\in\{\mathbf{t}\}} - 2\delta_{\boldsymbol{\tau}\notin\{\mathbf{t}\}}), \tag{A8}$$

$$\chi^{[\mathbf{0},k,l]}(g_{\boldsymbol{\tau}}^{\beta,\gamma}) = i^{2l\beta + k(2\gamma + \beta^2)}. \tag{A9}$$

[1] U. Abresch, A. Bovier, O. Lechtenfeld, M. Lüling, V. Rittenberg, and G. Weymans, J. Math. Phys. **21**, 2481 (1980).

[2] A. Bovier, M. Lüling, and D. Wyler, J. Math. Phys. **22**, 1536 (1981).

[3] A. Bovier, M. Lüling, and D. Wyler, J. Math. Phys. **22**, 1543 (1981).

[4] A. Bovier, G. Grenet, M. Kibler, M. Lüling, M. Marcu, and D. Wyler, to be published.

[5] W. M. Fairbairn, T. Fulton, and W. H. Klink, J. Math. Phys. **5**, 1038 (1964).

[6] H. S. Coxeter and W. O. S. Moser, *Generators and Relations for Discrete Groups* (Springer-Verlag, Berlin, 1957).

[7] G. A. Miller, H. E. Dickson, and C. E. Blichfeldt, *Theory and Applications of Finite Groups* (Steckart, New York, 1938).

[8] C. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Interscience, New York, 1962); G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).

[9] See, for example, G. Ljubarskii, *The Theory of Groups and Its Applications to Physics* (Pergamon, Oxford, 1960).

[10] For a qualitative discussion, see Ref. 2; for examples, Refs. 1–3.

[11] See the appendix of Ref. 2. The fact that $\Sigma(72)$ possesses a two-dimensional representation besides the eight-dimensional and the one-dimensional ones seems to contradict the assumption made there that such a "minimal" group should possess only n-dimensional and one-dimensional representations. However, this assumption was made only to find the minimal number of elements such a group should have, and is sufficient but not necessary. A two-dimensional real representation occurring twice in the product $8 \otimes 8$ absorbs the same number of states and makes the same number of group elements necessary as four one-dimensional representations.

[12] This group rule differs from the one given in Ref. 1, and yields however, a group isomorphic to $Q(3)$ (note that there are only two nonabelian groups of order 27). We use this form because P, Q, and M are unitary with respect to the quadratic form $\langle \boldsymbol{\tau},\boldsymbol{\tau}'\rangle = \tau_1\tau_2' - \tau_2\tau_1'$.

The Racah algebra for groups with time reversal symmetry. II

J. D. Newmarch [a)] and R. M. Golding

Department of Physical Chemistry, University of New South Wales, P.O. Box 1, Kensington, New South Wales 2033, Australia

(Received 8 March 1980; accepted for publication 20 August 1980)

In a previous paper[J. Math. Phys. **22**, 233 (1981)]it was shown that a Racah algebra could be developed for groups containing the antilinear operator of time reversal. Here the $1jm$ and $2jm$ symbols are constructed explicitly, and it is shown that the $3jm$ symbols may be found in terms of those of the linear subgroup. Thus the Racah algebra of these groups is known once the Racah algebra of the linear subgroup is known.

PACS numbers: 02.02.Hj, 03.65.Fd

1. INTRODUCTION

In a previous paper[1] henceforth denoted as I, we gave the elements of the Racah algebra for the compact groups with time reversal symmetry using the theory of corepresentations.[2] Although the algebra of such a group differs in many important respects from that of a linear group, the broad sweep remains the same: coupling coefficients and $3jm$ symbols may be found, and both the Wigner–Eckart theorem and Racah's lemma hold.

A compact group G of this type containing both linear and antilinear operators has a subgroup H consisting of the linear operators only. On restriction to this linear subgroup each irreducible corepresentation (ICR) of G becomes a possibly reducible representation of H. It was noted in I that according to the three types of irreducible representations of H under the Frobenius–Schur classification, there are exactly six types of ICR. These ICR's are built up in a very definite manner from the IRs of H.[2,3]

In view of this fact that the ICR's of G can be easily found once the IR's of H are known, we investigate in this paper how the n-jm symbols of G can be found from a knowledge of the n-jm symbols of H. Thus we here give prescriptions for producing the n-jm symbols and their symmetries in G from given symbols and symmetries in H. To achieve this in a definite manner, we have found it necessary to take specific matrix forms for the time reversal operator θ. This in turn implies that the IR's of H must be taken in certain standard forms. There is no loss of generality in this as a unitary transformation may always be applied to bring these into forms which may be commonly used for specific groups.

In the next section the symmetry properties of the $2jm$ and $3jm$ symbols of G are given in terms of those of H. An interesting result is that in certain cases a $3jm$ symbol in G must possess [21] symmetry even if no such symmetry exists in H. Sections 3 and 4 give complete constructions for the $1jm$ and $2jm$ symbols. In Sec. 5 we deal with the $3jm$ symbol. Due to the large number of cases it is not possible here to give exhaustive constructions, so we give a number of examples showing the methods involved. A complete list is available from the authors.

[a)]Present address: Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia.

2. SYMMETRIZED POWERS OF ICR'S

In I it was shown that the symmetries of the n-jm symbols for a grey group may be found in exactly the same way as for a linear group, namely by reducing the representations of S_n induced by the permutations of the n-jm symbols. Now the ICRs of G may be constructed from the IRs of the linear subgroup H in a definite manner (Sec. 3 of I), and we may use these constructions to relate the symmetries in G to those in H. We shall illustrate the method for a few of the more interesting cases.

The symmetry of the $2jm$ symbol in representation theory is intimately related to the Frobenius–Schur invariant which classifies the IRs into the first, second and third kinds. Let

$$C_k = \frac{1}{H}\int_H \chi_k(u^2)\,du.$$

For an IR of the first kind, $C_k = 1$, 0 is contained in the symmetric product $k\otimes[2]$ (or $[k^2]$), and there is no change in the value of the $2jm$ symbol $(k\ k\ 0)^r{}_{m_1 0 m_2}$ under transpositions. For an IR of the second kind, $C_k = -1$,0 and the symmetry of the $2jm$ symbol is undefined. Butler and King[4] have discussed various possibilities for this case of $k \not\equiv k^*$.

In the grey groups the situation differs in two respects: firstly, the multiplicity of 0 in $j\otimes j$ may be higher than one, and secondly, *every* ICR is equivalent to its complex conjugate. Defining the Frobenius–Schur invariant for grey groups by

$$C_j = \frac{2}{|G|}\int_H \chi_j(u^2)\,du,$$

it is not hard to show that C_j completely characterizes the symmetry. Consider for example an ICR of type (E) which is

TABLE I. Symmetry types of 0 in $j\otimes j$.

Type of ICR	Multiplicity of 0 in $j\otimes[2]=[j^2]$	Multiplicity of 0 in $j\otimes[1^2]=\{j^2\}$	Frobenius–Schur invariant
A	1	0	1
B	1	3	−2
C	1	1	0
D	0	1	−1
E	3	1	2
F	1	1	0

 0022-2488/81/102113-07$01.00

formed from two IR's of the first kind:

$$j(u)=\begin{pmatrix} k(u) & 0 \\ 0 & k(u)\end{pmatrix}.$$

Then

$$\chi_j(u)=2\chi_k(u) \text{ and } \chi_j(u^2)=2\chi_k(u^2).$$

The multiplicity $m_{[2]}$ of 0 in $j\otimes[2]$ is

$$m_{[2]}=\frac{1}{|G|}\int_H [\chi_j(u)]^2+\chi_j(u^2)\,du,$$

which, since k is of the first kind, gives

$$m_{[2]}=3.$$

Similarly, the multiplicity $m_{[1^2]}$ of 0 in $j\otimes[1^2]$ is

$$m_{[1^2]}=\frac{1}{|G|}\int_H [\chi_j(u)]^2-\chi_j(u^2)\,du$$
$$=1.$$

This rather curious result may be illustrated with grey C_3^* treated as a subgroup of grey SU(2). The ICR E' of type (E) has couplings to A of

$$(3/2)E'\otimes(3/2)E'\to 0A,1A,3aA,3bA,$$

and taking the symmetry from the sum of the j values there are three symmetric and one antisymmetric terms.

The symmetries of the $2jm$ symbol for all types of ICRs are given in Table I.

Turning now to the $3jm$ symbol, let us look at the case when all three j's are equivalent. Permutations of the $3jm$ symbol generate a representation Γ_j of S_3 which may be reduced to

$$\Gamma_j=m^j_{[1^3]}[1^3]\oplus m^j_{[3]}[3]\oplus m^j_{[21]}[21],$$

where the Young diagrams have been used to label the IR's of S_3 and the multiplicities are given by

$$m^j_{[3]}=\frac{1}{3|G|}\int_H\{[\chi_j(u)]^3+3\chi_j(u^2)\chi_j(u)+2\chi_j(u^3)\}\,du,$$

$$m^j_{[1^3]}=\frac{1}{3|G|}\int_H\{[\chi_j(u)]^3-3\chi_j(u^2)\chi_j(u)+2\chi_j(u^3)\}\,du,$$

and

$$m^j_{[21]}=\frac{2}{3|G|}\int_H\{[\chi_j(u)]^3-\chi_j(u^3)\}\,du.$$

If the ICR is of type (a), then $j(u)=k(u)$. [The Frobenius–Schur classification is irrelevant for the triple product and so it is not necessary to separate this into types (A) or (D).] the representation Γ_j is clearly equal to Γ_k, the representation of S_3 generated by the $3jm$ symbol of the linear subgroup, and so there is no change in these multiplicities. For a type (b) ICR, letting

$$m^k_{[3]}=\frac{1}{6|H|}\int_H\{[\chi_k(u)]^3+3\chi_k(u^2)\chi_k(u)+2\chi_k(u^3)\}\,du,$$

etc., and $\chi_j(u^n)=2\chi_k(u^n)$, gives the results

$$m^j_{[1^3]}=4m^k_{[1^3]}+2m^k_{[21]},$$
$$m^j_{[3]}=4m^k_{[3]}+2m^k_{[21]},$$

and

$$m^j_{[21]}=2m^k_{[1^3]}+2m^k_{[3]}+6m^k_{[21]}$$

or

$$\Gamma_j=(4[3]\oplus 2[21])\otimes\Gamma_k .$$

This shows that there must be at least two [21] symmetry terms in the permutation matrix of the $3jm$ symbols.

The last type of ICR, type (c) has

$$j(u)=\begin{pmatrix} k(u) & 0 \\ 0 & k(u)^*\end{pmatrix},$$

with $k(u)\not\equiv k(u)^*$. The triple product $j\otimes j\otimes j$ will give the products $k\otimes k\otimes k$, $k^*\otimes k^*\otimes k^*$, $k\otimes k\otimes k^*$, and $k\otimes k^*\otimes k^*$. The $3jm$ symbols for $k\otimes k\otimes k$ and $k^*\otimes k^*\otimes k^*$ yield equivalent representations of S_3 (Ref. 5) and as before we may set $m^k_{[1^3]}$ equal to the multiplicity of $[1^3]$ in the representation generated by $k\otimes k\otimes k$, etc.. Setting $m^k_{[1^2]}$ to be the multiplicity of k in $k\otimes[1^2]$ and $m^k_{[2]}$ to be the multiplicity of k in $k\otimes[2]$ gives

$$m^j_{[1^3]}=2m^k_{[1^3]}+2m^k_{[1^2]},$$
$$m^j_{[3]}=2m^k_{[3]}+2m^k_{[2]},$$

and

$$m^j_{[21]}=2m^k_{[21]}+2m^k_{[1^2]}+2m^k_{[2]},$$

showing that if k is contained in $k\otimes k$ then again we must have the mixed symmetry term [21].

In a similar manner the symmetries when $j_1\equiv j_2\not\equiv j_3$ may be found in terms of the symmetries of the linear subgroup. These are given in Table II. When none of j_1, j_2, and j_3 are equivalent the symmetry may be chosen arbitrarily.

3. THE $1jm$ SYMBOL

The $1jm$ symbol (Ref. 5) is defined to be the matrix which transforms an ICR to its complex conjugate. It was shown in Ref. 1 Sec. 6 that although this matrix is not necessarily unique, the time reversal operator may always be used:

$$j(u)=j(\theta)\,j(u)^*j(\theta)^{-1} \text{ and } j(a)=j(\theta)\,j(a)^*j(\theta)^{-1*}.$$

Thus the problem reduces to finding a matrix form of $j(\theta)$. Some discussion in Ref. 1 was given to block diagonalizing $j(\theta)$ but it turns out that this is not a particularly useful form and that it is better to find the $1jm$ symbol in H and extend it to G. This may be done explicitly for IR's of the first and second kinds by a suitable standardization of the IR

TABLE II. Symmetry structure $\Gamma_{j_1j_1j_2}$ of the $3jm$ tensor $(j_1j_1j_2)$ in terms of the symmetry structure $\Gamma_{k_1k_1k_2}$ of the tensor $(k_1k_1k_2)$ of the linear subgroup. $\otimes$ is the inner direct product, $\odot$ the outer direct product.

$j_1\ j_1\ j_2$	$\Gamma_{j_1j_1j_2}$
(a) (a) (a)	$\Gamma_{k_1k_1k_2}$
(a) (a) (b)	$2\Gamma_{k_1k_1k_2}$
(a) (a) (c)	$2\Gamma_{k_1k_1k_2}$
(b) (b) (a)	$(3[2]\oplus[1^2])\otimes\Gamma_{k_1k_1k_2}$
(b) (b) (b)	$(6[2]\oplus 2[1^2])\otimes\Gamma_{k_1k_1k_2}$
(b) (b) (c)	$(6[2]\oplus 2[1^2])\otimes\Gamma_{k_1k_1k_2}$
(c) (c) (a)	$2[2]\otimes\Gamma_{k_1k_1k_2}\oplus 2[1]\odot\Gamma_{k_1k_1^*k_2}$
(c) (c) (b)	$4[2]\otimes\Gamma_{k_1k_1k_2}\oplus 4[1]\odot\Gamma_{k_1k_1^*k_2}$
(c) (c) (c)	$2[2]\otimes\Gamma_{k_1k_1k_2}\oplus 2[2]\otimes\Gamma_{k_1k_1}k_2^*\oplus 4[1]\odot\Gamma_{k_1k_1^*k_2}$

matrices.

An IR of the first kind is equivalent to its conjugate and also to a real representation. Choose this real form:

$$k(u) = k(u)^*.$$

An irreducible representation of the second kind is equivalent to its conjugate but not to a real representation. A suitable form is

$$k(u) = \begin{pmatrix} k_1(u) & k_2(u) \\ -k_2(u)^* & k_1(u)^* \end{pmatrix},$$

with

$$k(u) = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} k(u)^* \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

with each block square. This may be obtained from the form given by Wybourne[6] by a permutation of the basis vectors.

For an IR of the third kind we make the obvious choice: if k is represented by $k(u)$ then the complex conjugate k^* is represented by $k(u)^*$.

These standardizations may be combined with the results of Bradley and Davies,[3] Sec. 2.4, to fix $j(\theta)$ to within phase. Since, however, under a change of basis $j'(\theta) = Pj(\theta)P^{*-1}$, this phase may be chosen arbitrarily and we choose the following as the various kinds of ICR:

type (A) $j(u) = k(u)$ and $j(\theta) = I$,

type (B) $j(u) = \begin{pmatrix} k(u) & 0 \\ 0 & k(u) \end{pmatrix}$

and

$$j(\theta) = \begin{pmatrix} 0 & 0 & 0 & I \\ 0 & 0 & -I & 0 \\ 0 & -I & 0 & 0 \\ I & 0 & 0 & 0 \end{pmatrix},$$

type (C) $j(u) = \begin{pmatrix} k(u) & 0 \\ 0 & k(u)^* \end{pmatrix}$ and $j(\theta) = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}$,

type (D) $j(u) = k(u)$ and $j(\theta) = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$,

type (E) $j(u) = \begin{pmatrix} k(u) & 0 \\ 0 & k(u) \end{pmatrix}$ and $j(\theta) = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$,

type (F) $j(u) = \begin{pmatrix} k(u) & 0 \\ 0 & k(u)^* \end{pmatrix}$ and $j(\theta) = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$,

in which the $1jm$ symbol is given explicitly. The $1j$ symmetry ϕ_j upon interchange of the m values is trivially found from this.

4. THE $2jm$ SYMBOL

Using the conventions of the previous section the $2jm$ symbol[7] may easily be constructed by use of Eq. (7.9) of I

$$(j\,0\,j)^r{}_{m_1 0 m_2} = [j]^{-1/2}\delta_{m_4 \dot m_3}\begin{bmatrix} \dot m_3 & \\ & m_2 \end{bmatrix}\langle j0|j\rangle^{rm_4}{}_{m_1 0},$$

where $[\,{}^{\dot m_3}\;{}_{m_2}\,] = j(\theta)^{-1}$ and the coupling coefficients are given in Sec. 7 of I. Whereas in representation theory there is essentially only one coupling coefficient, in corepresentation theory there may be up to four. Each row of the $2jm$ symbol may be found separately and after checking orthogonality between rows the complete symbol may be found. It turns out to be already in symmetrized form.

type (A): $\langle j0|j\rangle = I$ and $j(\theta)^{-1} = I$.

Now $\langle j0|j\rangle$ is a $[j]$ by $[j]$ matrix, as is $j(\theta)^{-1}$, whereas $(j\;0\;j)$ is one by $[j]^2$. The δ term "smears" the product on the right as we illustrate by the case $[j] = 2$. We have

$$(j\;0\;j)^1{}_{101} = 2^{-1/2}\langle j0|j\rangle^{m_4}{}_{10}\delta_{m_4\dot m_3}\begin{bmatrix} \dot m_3 & \\ & 1 \end{bmatrix}.$$

the only nonvanishing term occurs when $\dot m_3 = m_4 = 1$ so

$$(j\;0\;j)^1{}_{101} = 2^{-1/2}.$$

Similarly,

$$(j\;0j)^1{}_{201} = (j\;0j)^1{}_{102} = 0$$

and

$$(j\;0j)^1{}_{202} = 2^{-1/2}.$$

Thus

$$(j\;0j)^1{}_{m_1 0 m_2} = 2^{-1/2}(1\;0\;0\;1)^1{}_{m_1 0 m_2}.$$

The general case when $[j] = n$ follows similarly as

$$(j\;0j)^1{}_{m_1 0 m_2} = [j]^{-1/2}(E_1 E_2 \cdots E_n)^1{}_{m_1 0 m_2},$$

where E_i is an elementary row vector with n columns and a one in the ith position and zeros elsewhere.

type (B):

$$\langle j0|j\rangle = \begin{pmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{pmatrix}, \begin{pmatrix} -iI & 0 & 0 & 0 \\ 0 & -iI & 0 & 0 \\ 0 & 0 & iI & 0 \\ 0 & 0 & 0 & iI \end{pmatrix}, \begin{pmatrix} 0 & 0 & -iI & 0 \\ 0 & 0 & 0 & -iI \\ -iI & 0 & 0 & 0 \\ 0 & -iI & 0 & 0 \end{pmatrix}, \text{ or } \begin{pmatrix} 0 & 0 & -I & 0 \\ 0 & 0 & 0 & -I \\ I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \end{pmatrix}$$

and

$$j(\theta)^{-1} = \begin{pmatrix} 0 & 0 & 0 & I \\ 0 & 0 & -I & 0 \\ 0 & -I & 0 & 0 \\ I & 0 & 0 & 0 \end{pmatrix},$$

where the coupling coefficients drawn from I Sec. 7 have had their block size adjusted to conform to the block size of $j(\theta)^{-1}$. A construction similar to the last applied to each of the four coupling coefficients gives the four rows of the $2jm$ symbol, which may be combined as

$$(j\,0\,j)^r{}_{m_10m_2} =$$

$$[j]^{-1/2}\begin{array}{c} \begin{array}{cccccccccccccccc} 101 & 102 & 103 & 104 & 201 & 202 & 203 & 204 & 301 & 302 & 303 & 304 & 401 & 402 & 403 & 404 \end{array} \\ \left\{\begin{array}{cccccccccccccccc} 0 & -E & 0 & 0 & E & 0 & 0 & 0 & 0 & 0 & 0 & -E & 0 & 0 & E & 0 \\ 0 & iE & 0 & 0 & -iE & 0 & 0 & 0 & 0 & 0 & 0 & -iE & 0 & 0 & iE & 0 \\ \\ 0 & 0 & 0 & -iE & 0 & 0 & iE & 0 & 0 & -iE & 0 & 0 & iE & 0 & 0 & 0 \\ 0 & 0 & 0 & E & 0 & 0 & -E & 0 & 0 & -E & 0 & 0 & E & 0 & 0 & 0 \end{array}\right\}^r_{m_10m_2} \end{array}$$

with E as the row vector $(E_1,E_2,\ldots,E_n)$. The block structure of the m values is shown along the top from which it is found that the first three rows possess $[1^2]$ symmetry and the last $[2]$ symmetry as required.

$$type\ \text{(C)}:\ \langle j0|j\rangle = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \text{ or } \begin{pmatrix} -iI & 0 \\ 0 & iI \end{pmatrix}$$

and

$$j(\theta)^{-1} = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix},$$

to give

$$(j0j)^r{}_{m_10m_2} = [j]^{-1/2} \begin{array}{c} \begin{array}{cccc} 101 & 102 & 201 & 202 \end{array} \\ \begin{pmatrix} 0 & -iE & iE & 0 \\ 0 & E & E & 0 \end{pmatrix} \end{array},$$

with the first row antisymmetric and the second symmetric.

$$type\ \text{(D)}: \langle j0|j\rangle = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \text{ and } j(\theta)^{-1} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix},$$

giving one antisymmetric row

$$(j\ 0\ j)^1{}_{m_10m_2} = [j]^{-1/2} \begin{array}{c} \begin{array}{cccc} 101 & 102 & 201 & 202 \end{array} \\ (\begin{array}{cccc} 0 & -E & E & 0 \end{array})^1_{m_10m_2} \end{array}.$$

$$type\ \text{(E)}:\ \langle j0|j\rangle = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix},$$

$$\begin{pmatrix} -iI & 0 \\ 0 & iI \end{pmatrix}, \begin{pmatrix} 0 & -iI \\ -iI & 0 \end{pmatrix}, \text{ or } \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$$

and $j(\theta)^{-1} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix}$,

yielding

$$(j\ 0\ j)^r_{m_10m_2} = [j]^{-1/2} \begin{array}{c} \begin{array}{cccc} 101 & 102 & 201 & 202 \end{array} \\ \begin{pmatrix} 0 & -E & E & 0 \\ 0 & iE & iE & 0 \\ -iE & 0 & 0 & iE \\ E & 0 & 0 & E \end{pmatrix}^r_{m_10m_2} \end{array}.$$

The first row is antisymmetric and the other three symmetric.

$$type\ \text{(F)}:\ \langle j0|j\rangle = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \text{ or } \begin{pmatrix} -iI & 0 \\ 0 & iI \end{pmatrix}$$

from which,

$$(j\ 0\ j)^r{}_{m_10m_2} = [j]^{-1/2} \begin{array}{c} \begin{array}{cccc} 101 & 102 & 201 & 202 \end{array} \\ \begin{pmatrix} 0 & -iE & -iE & 0 \\ 0 & E & -E & 0 \end{pmatrix}_{m_10m_2} \end{array}$$

where the first row is symmetric and the second antisymmetric.

5. THE 3*jm* SYMBOL

We have seen in the last two sections how the 1*jm* and 2*jm* symbols may be constructed explicitly. In considering the 3*jm* symbol, however, there are such a large number of cases that it is not possible to deal with them all here. We give a number of examples illustrating the methods and the problems which arise. A complete list of constructions is available.

Example 1: j_1, j_2, and j_3 all of type (A).

This is the simplest case, for $j(\theta) = I$ and on descent to the linear subgroup H there is no branching. Suppose P reduces the triple product in G:

$$j_1 \otimes j_2 \otimes j_3(u) = P^{*T}P \oplus \cdots$$

and

$$j_1 \otimes j_2 \otimes j_3(a) = P^{*T}P^* \oplus \cdots,$$

or more conveniently,

$$P(j_1 \otimes j_2 \otimes j_3)(u) = P \tag{5.1}$$

and

$$P(j_1 \otimes j_2 \otimes j_3)(a) = P^*, \tag{5.2}$$

where the number of rows of P equals the multiplicity of 0 in $j_1 \otimes j_2 \otimes j_3$ and the number of columns is $[j_1]\times[j_2]\times[j_3]$. From Eq. (5.1), since $j_1(u) = k_1(u)$ etc., P must also reduce $k_1 \otimes k_2 \otimes k_3$ and hence is a 3*jm* symbol of H. By setting $a = \theta$ in the second equation we find $P = P^*$. Thus any *real* 3*jm* symbol of H is also a 3*jm* symbol of G.

Example 2: j_1 of type (A), j_2 and j_3 of type (C).

Here

$$\begin{aligned} j_1 \otimes j_2 \otimes j_3(u) &= k_1(u) \otimes \begin{pmatrix} k_2(u) & 0 \\ 0 & k_2(u)^* \end{pmatrix} \otimes \begin{pmatrix} k_3(u) & \\ 0 & k_3(u)^* \end{pmatrix} \\ &= \begin{pmatrix} k_1 \otimes k_2 \otimes k_3(u) & 0 & 0 & 0 \\ 0 & k_1 \otimes k_2 \otimes k_3^*(u) & 0 & 0 \\ 0 & 0 & k_1 \otimes k_2^* \otimes k_3(u) & 0 \\ 0 & 0 & 0 & k_1 \otimes k_2^* \otimes k_3^*(u) \end{pmatrix} \end{aligned} \tag{5.3}$$

and

$$j_1 \otimes j_2 \otimes j_3(\theta) = \begin{pmatrix} 0 & 0 & 0 & I \\ 0 & 0 & I & 0 \\ 0 & I & 0 & 0 \\ I & 0 & 0 & 0 \end{pmatrix}. \tag{5.4}$$

There are two relevant $3jm$ symbols of H: P_1 which reduces $k_1 \otimes k_2 \otimes k_3$ [with P_1^* reducing $k_1 \times k_2^* \times k_3^*$] and P_2 which reduces $k_1 \times k_2 \times k_3^*$ [with P_2^* reducing $k_1 \otimes k_2^* \otimes k_3$]. From the linear equation a trial value for P would be

$$P = \begin{pmatrix} P_1 & & & 0 \\ & P_2 & & \\ & & P_2^* & \\ 0 & & & P_1^* \end{pmatrix}$$

but Eq. (5.4) mixes the first and last rows and also the second and third rows, and a little experimentation shows that P may be taken as

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} P_1 & 0 & 0 & P_1^* \\ 0 & P_2 & P_2^* & 0 \\ 0 & -iP_2 & iP_2^* & 0 \\ -iP_1 & 0 & 0 & iP_1^* \end{pmatrix}. \tag{5.5}$$

Note that there are no restrictions on the choice of P_1 and P_2 so this is one of the easiest cases to construct.

Example 3: j_1 and j_2 of type (A), j_3 of type (B). We have

$$j_1 \otimes j_2 \otimes j_3(u) = \begin{pmatrix} k_1 \otimes k_2 \otimes k_3(u) & 0 \\ 0 & k_1 \otimes k_2 \otimes k_3(u) \end{pmatrix} \tag{5.6}$$

and

$$j_1 \otimes j_2 \otimes j_3(\theta) = \begin{pmatrix} 0 & J \\ -J & 0 \end{pmatrix}, \tag{5.7}$$

where

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}. \tag{5.8}$$

Now the multiplicity of 0 in $k_1 \otimes k_2 \otimes k_3$ is even[4] and $[k_3]$ is even, so the matrix P_1 which reduces $k_1 \otimes k_2 \otimes k_3$ may be written as

$$P_1 = \begin{pmatrix} P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} p_1 & p_2 \\ p_3 & p_4 \end{pmatrix}$$

and with the standardization of k_3 we also have

$$k_1 \otimes k_2 \otimes k_3(u) = \begin{pmatrix} k_4(u) & k_5(u) \\ -k_5(u)^* & k_4(u)^* \end{pmatrix}.$$

It follows that P_1 may be taken as

$$P_1 = \begin{pmatrix} p_1 & p_2 \\ -p_2^* & p_1^* \end{pmatrix} \tag{5.9}$$

or that

$$P_2 J = P_3^* \text{ and } P_3 J = -P_2^*. \tag{5.10}$$

The reason for this particular choice is as follows: from Table I of I, the multiplicity of j_3 in $j_1 \otimes j_2$ is exactly one-half the multiplicity of k_3 in $k_1 \otimes k_2$ and the multiplicity of 0 in $j_3 \otimes j_3$ is four. Equation (7.13) of I which gives the $3jm$ symbol in terms of the two sets of coupling coefficients implies that that we should look for a 4×2 block structure of the $3jm$ symbol rather than a 2×2 block structure as suggested by Eq. (5.6) and that each block row should be made up of P_2 and P_3. As the $3jm$ symbol must also reduce Eq. (5.7) a relation between P_2, P_3, and J is required, of the form of Eq. (5.10). After these preliminaries it is straightforward to verify that P may be taken as

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} P_2 & P_3 \\ P_3 & -P_2 \\ iP_2 & -iP_3 \\ iP_3 & iP_2 \end{pmatrix}. \tag{5.11}$$

This comparative complexity seems to be shared by all couplings involving an IR of the second kind due to the form of the matrix which gives equivalence to the conjugate. We conclude this section with probably the most extreme case:

Example 4: $j_1 = j_2 = j_3$ of type (B). Here

$$j \otimes j \otimes j(u) = \begin{pmatrix} k \otimes k \otimes k\,(u) & & 0 \\ & k \otimes k \otimes k\,(u) & \\ 0 & & \ddots \; k \otimes k \otimes k\,(u) \end{pmatrix} \tag{5.12}$$

is an 8×8 block matrix, and

$j \otimes j \otimes j(\theta)$

$$= \begin{pmatrix} & & & & & & & J\otimes J\otimes J \\ & & & & & & -J\otimes J\otimes J & \\ & & 0 & & & -J\otimes J\otimes J & & \\ & & & & J\otimes J\otimes J & & & \\ & & & -J\otimes J\otimes J & & & & \\ & & J\otimes J\otimes J & & & & & \\ & J\otimes J\otimes J & & & & & & \\ -J\otimes J\otimes J & & & & & 0 & & \end{pmatrix}. \tag{5.13}$$

Again the multiplicity of 0 in $k \otimes k \otimes k$ is even and now $[k] \times [k] \times [k]$ is divisible by eight. Thus P_1 which reduces $k \otimes k \otimes k$ may be written as

$$P_1 = \begin{pmatrix} P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \end{pmatrix}.$$

It is straightforward but tedious to verify that, similar to the last example, we may choose P_3 to satisfy

$$P_2(J \otimes J \otimes J) = P_3^* \text{ and } P_3(J \otimes J \otimes J) = -P_2^*. \tag{5.14}$$

If each block row is to be constructed from P_2 and P_3 we shall need sixteen such rows, with eight block columns. There is an additional complication, for from Sec. 2, $\Gamma_j = (4[3] + 2[21]) \otimes \Gamma_k$ where Γ_j and Γ_k are the S_3 symmetries of the $3jm$ symbols in each group. Taking these into account and using the permutation matrices for [21] of Butler[5] or Hamermesh[8] gives, with the block labelling at the top and the symmetry structure down the side

$$P = \begin{array}{c} \begin{array}{cccccccc} 111 & 112 & 121 & 122 & 211 & 212 & 221 & 222 \end{array} \\ \begin{pmatrix} P_2/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & P_3/\sqrt{2} \\ P_3/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & -P_2/\sqrt{2} \\ iP_2/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & -iP_3/\sqrt{2} \\ iP_3/\sqrt{2} & 0 & 0 & 0 & 0 & 0 & 0 & iP_2/\sqrt{2} \\ 0 & P_2/\sqrt{6} & P_2/\sqrt{6} & -P_3/\sqrt{6} & P_2/\sqrt{6} & -P_3/\sqrt{6} & -P_3/\sqrt{6} & 0 \\ 0 & P_3/\sqrt{6} & P_3/\sqrt{6} & P_2/\sqrt{6} & P_3/\sqrt{6} & P_2/\sqrt{6} & P_2/\sqrt{6} & 0 \\ 0 & iP_2/\sqrt{6} & iP_2/\sqrt{6} & iP_3/\sqrt{6} & iP_2/\sqrt{6} & iP_3/\sqrt{6} & iP_3/\sqrt{6} & 0 \\ 0 & iP_3/\sqrt{6} & iP_3/\sqrt{6} & -iP_2/\sqrt{6} & iP_3/\sqrt{6} & -iP_2/\sqrt{6} & -iP_2/\sqrt{6} & 0 \\ 0 & P_2/\sqrt{3} & -P_2/2\sqrt{3} & P_3/2\sqrt{3} & -P_2/2\sqrt{3} & P_3/2\sqrt{3} & -P_3/\sqrt{3} & 0 \\ 0 & P_3/\sqrt{3} & -P_3/2\sqrt{3} & -P_2/2\sqrt{3} & -P_3/2\sqrt{3} & -P_2/2\sqrt{3} & P_2/\sqrt{3} & 0 \\ 0 & 0 & P_2/2 & P_3/2 & -P_2/2 & -P_3/2 & 0 & 0 \\ 0 & 0 & P_3/2 & -P_2/2 & -P_3/2 & -P_2/2 & 0 & 0 \\ 0 & iP_2/\sqrt{3} & -iP_2/2\sqrt{3} & -iP_3/2\sqrt{3} & -iP_2/2\sqrt{3} & -iP_3/2\sqrt{3} & iP_3/\sqrt{3} & 0 \\ 0 & iP_3/\sqrt{3} & -iP_3/2\sqrt{3} & iP_2/2\sqrt{3} & -iP_3/2\sqrt{3} & iP_2/2\sqrt{3} & -iP_2/\sqrt{3} & 0 \\ 0 & 0 & iP_2/2 & -iP_3/2 & -iP_2/2 & iP_3/2 & 0 & 0 \\ 0 & 0 & iP_3/2 & iP_2/2 & -iP_3/2 & -iP_2/2 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \left.\begin{array}{c} \\ \\ \end{array}\right\} [3] \otimes \Gamma_k \\ \left.\begin{array}{c} \\ \\ \end{array}\right\} [3] \otimes \Gamma_k \\ \left.\begin{array}{c} \\ \\ \end{array}\right\} [3] \otimes \Gamma_k \\ \left.\begin{array}{c} \\ \\ \end{array}\right\} [3] \otimes \Gamma_k \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} [21] \otimes \Gamma_k \\ \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} [21] \otimes \Gamma_k \end{array} \tag{5.15}$$

It hoped that these examples are sufficient to show the problems which arise and their method of solution.

6. CONCLUSION

The examples of the last section show how we can express the n-jm symbols of a grey group in terms of those of the linear subgroup. This means that we do not have to start from scratch if time reversal is added to a particular linear group, which should give a great saving in computations. This is of course not the only way of finding these symbols—tables are available from the authors of $3jm$ symbols and some $6j$ and isoscalar symbols for the grey double point groups which were calculated using descent in symmetry from grey SU(2).

The coupling coefficient has not been dealt with explicitly, but it can be constructed in much the same way as the $3jm$ symbol.

[1]J. D. Newmarch and R. M. Golding, J. Maths. Phys., **22**, 233 (1981).

[2]E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).

[3]C. J. Bradley and B. L. Davies, Rev. Mod. Phys. **40**, 359 (1968).

[4]R. C. King, Canad. J. Math. **26**, 328 (1974).

[5]P. H. Butler, Philos. Trans. R. Soc. London Ser. A **277**, 545 (1975).

[6]B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).

[7]A distinction is being made from the $1jm$ symbol here. The $1jm$ symbol relates j to j^* and is always taken as $j(\theta)$—a square matrix. The $2jm$ symbol reduces the product $j \otimes j$ (or $j \otimes 0 \otimes j$) to 0 and is rectangular.

[8]M. Hamermesh, *Group Theory and its Application to Physical Problems* (Addison–Wesley, Reading, Mass., 1962).

Generalized Bessel series and multiplicity problem in complex semisimple Lie algebra theory

C. Bretin and J. P. Gazeau
Laboratoire de Chimie Physique[a] *de l'Université Pierre et Marie Curie, 11, rue Pierre et Marie Curie, 75 231 Paris Cedex 05, France*

(Received 25 February 1980; accepted for publication 23 July 1980)

The connection between Bessel series and SU(2) is reviewed from the standpoint of the outer multiplicity problem for this group. Its extension to any complex semisimple Lie algebras allows one to introduce new objects called "generalized Bessel series." Some applications concerning the special function theory (addition theorems) are given.

PACS numbers: 02.20.Sv, 02.20.Qs, 02.30.Gp, 02.30.Lt

In this work we generalize the connection that exists between Bessel functions and the representations of certain Lie groups.

In this first place one notices that the Bessel series coefficients[1] are, to within a factorial, the multiplicities of the unitary irreducible representations (UIR) of SU(2) in the Clebsch–Gordan (CG) decomposition of the successive tensor powers of the fundamental representation of the group. This is brought out by means of the generating function $\exp(i\mathbf{k}\cdot\mathbf{r})$. Indeed, $i\mathbf{k}\cdot\mathbf{r} = ikr\cos\theta = 2ikr\chi(\theta)$, where χ is the character of the SU(2) fundamental representation.

The generalization then becomes apparent.[2,3] We introduce for all semisimple compact Lie group G a generating function $\exp z\chi_\lambda(g)$ where $g\in G$ and χ_λ is the character of a UIR λ of this group, and we define a generalized Bessel function associated with G as the Fourier coefficient of the expansion of this exponential on the χ_μ basis. In fact, one can do without the Weyl "unitarian trick"[4] by examining the problem in $\mathbb{Z}[P]$ (and its extension $\mathbb{Q}[P]$) which is the algebra of the weight group P of a semisimple Lie algebra $\mathfrak{g}$.[3,4]

The physicist can exploit the formula thus established in two ways since:

(1) It is of interest to have a general formula for certain outer multiplicities in the reduction of certain tensor products (whether it be in problems concerned with unitary symmetries in particle physics or in problems originating in couplings in nuclear, atomic, or molecular physics). As an example, we give here the multiplicities of the UIR's of SU(n) in the CG decomposition of any tensor power of a certain fundamental representation of this group [in the SU(3) case, that called "quark" or its conjugate "antiquark"]. Another example concerning Spin (5) is given.

(2) Besides, we generalize certain addition theorems met with in special function theory, especially those relating to solid harmonics with decentered argument and used for instance in many-body problems in which appear potentials with spherical symmetry. This generalization consists in expanding matrix elements $\mathscr{D}^{[m]_n}(x'+x'')_{m^\cdot m^\cdot}$ of finite analytic irreducible linear representations of GL $(n,\mathbb{C})$, where x', x'', $x'+x''\in\mathrm{GL}(n,\mathbb{C})$.[5]

Actually it can be shown that a broad category of addition theorems can be expressed in this way (provided some analytic continuation on the indices are made in many cases).

I. INTRODUCTION

Let us consider the following formula, drawn from the special function machinery[1]:

$$\exp ab = (a/2)^{-\lambda}\Gamma(\lambda)\sum_{n=0}^{+\infty}(\lambda+n)I_{\lambda+n}(a)C_n^\lambda(b), \tag{1}$$

where C_n^λ is a Gegenbauer polynomial and I_ν a modified Bessel function.

Let x be a nonsingular 2×2 complex matrix. We can verify that

$$C_n^1(\mathrm{Tr}\,x/2(\det x)^{1/2}) = (\det x)^{-n/2}\chi_n(x), \tag{2}$$

where χ_n is the character of the totally symmetrical, $(n+1)$-dimensional analytic irreducible representation of GL(2,$\mathbb{C}$), that is labeled by $(n,0)$ in the Gel'fand notation.[5]

We put $\lambda = 1$, $a = 2(\det x)^{1/2}$, and $b = (\mathrm{Tr}\,x)/2(\det x)^{1/2}$ in Eq. (1).

This leads to the following expansion, taking into account that of the I_{n+1} function in powers of a

$$\exp(\mathrm{Tr}\,x)\equiv\exp\chi_1(x) = \sum_{n,q\geqslant 0}\frac{(n+1)(\det x)^q}{q!(q+n+1)!}\chi_n(x), \tag{3}$$

formula to compare with the usual development

$$\exp\chi_1(x) = \sum_{r=0}^{+\infty}(\chi_1(x))^r/r!. \tag{4}$$

By restricting the above expansions to the elements x of GL(2,$\mathbb{C}$) which can be written under the form

$$x = (\det x)^{1/2}u,\quad u\in\mathrm{SU}(2), \tag{5}$$

and by using the evident homogeneity properties

$$\chi_n((\det x)^{1/2}u) = (\det x)^{n/2}\chi_n(u),$$

$$(\chi_1(x))^r = (\det x)^{r/2}(\chi_1(u))^r,$$

the development of the rth power of $\chi_1(u)$ in terms of the orthogonal set of the characters $\chi_n(u)$, or equivalently, the Clebsch–Gordan decomposition of the rth tensor power of the fundamental representation of SU(2), can be considered

$$(\chi_1(u))^r = \sum_n m^r_{1;n}\chi_n(u). \tag{6}$$

[a] Laboratoire "Matière et Rayonnement" associé au C. N. R. S.

 0022-2488/81/102120-07$01.00

By replacing (6) in Eq. (4) and comparing the latter with Eq. (3), we obtain the interesting result[6]

$$m^r_{1;n} = \delta_{n,r-2q} \frac{(n+1)r!}{q!(q+n+1)!}. \tag{7}$$

As a consequence of all this, the following series

$$I_{1;n}(z) \equiv \sum_{r=0}^{+\infty} m^r_{1;n} \frac{z^r}{r!} = (n+1)z^n \sum_{q=0}^{+\infty} \frac{z^{2q}}{q!(n+q+1)!}, \tag{8}$$

is closely related to the modified Bessel function

$$I_{1;n}(z) = [(n+1)/2z] I_{n+1}(2z). \tag{9}$$

That is the reason why we shall name the former "Bessel series" associated to the Lie algebra A_1 of the group SL(2,$\mathbb{C}$), the real compact form of which being SU(2).

The following section makes the most of this idea by extending it to any complex semisimple Lie algebra.

II. GENERALIZED BESSEL SERIES ASSOCIATED WITH COMPLEX SEMISIMPLE LIE ALGEBRAS

In the following, the notations used are related to Ref. 3.

Let $\mathfrak{g}$ be a complex semisimple Lie algebra of rank l. Let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$ (dim $\mathfrak{h} = l$). Let R be the set of roots of $\mathfrak{g}$ with respect to $\mathfrak{h}$. P_{++} designates the set of dominant weights with respect to a basis of the root system R. For $\lambda \in P_{++}$, let $V(\lambda)$ be the corresponding finite-dimensional simple $\mathfrak{g}$-module with highest weight λ. For all $\mu \in P_{++}$, let $m^r_{\lambda;\mu}$ be the multiplicity of $V(\mu)$ in the Clebsch–Gordan decomposition of $\otimes^r V(\lambda)$, the rth tensor power of $V(\lambda)$:

$$\otimes^r V(\lambda) = \bigoplus_{\mu} \Big(\underbrace{\oplus \cdots \oplus}_{m^r_{\lambda;\mu} \text{ times}} \Big) V(\mu). \tag{10}$$

The following series

$$I_{\lambda;\mu}(z) = \sum_{r \geqslant 0} m^r_{\lambda;\mu} z^r / r! \tag{11}$$

will be named "($\lambda;\mu$) generalized Bessel series associated with $\mathfrak{g}$".

For instance, in the A_1 case, and for $\lambda = \bar{\omega}$, unique fundamental weight, we recover the definition (8) by identifying $I_{\bar{\omega};n\bar{\omega}}$ with $I_{1;n}$.

An equivalent definition lies in the fundamental expansion formula below, involving such series, and using the "Weyl unitarian trick[11] of $\mathfrak{g}$. Let G be the simply connected compact Lie group having the compact real form of $\mathfrak{g}$ as Lie algebra, χ_λ the character of the unitary irreducible representation D^λ of G. Any invariant function on the group G is completely defined by its restriction to T, the Cartan subgroup of G, and it is apparent that the functions χ_λ are complete with respect to any continuous invariant function on G.

The properties in analyticity of the invariant function $\exp(z\chi_\lambda(g))$, $g \in G$, entail the pointwise convergence of its development in terms of χ_μ

$$\exp(z\chi_\lambda(u)) = \sum_{\mu} I_{\lambda;\mu}(z) \chi_\mu(u), \quad u \in T. \tag{12}$$

III. GENERATING FUNCTIONS FOR THE MULTIPLICITY $m^r_{\lambda;\mu}$, AND ALGEBRAIC FORMULAS

We denote by P the group of weights of $\mathfrak{g}$, by $\mathbb{Z}[P]$ the algebra of the group P, and by $\mathbb{Z}[P]^W$ the subalgebra of $\mathbb{Z}[P]$ whose elements are fixed with respect to the action of the Weyl group W of R.[3,4] Let $\{e^\eta\}_{\eta \in P}$ be a basis of $\mathbb{Z}[P]$.

For each $\varphi \in \mathbb{Z}[P]^W$, we have

$$\varphi = \sum_{\mu \in P_{++}} \varphi_\mu \mathrm{ch}_\mu,$$

where ch_μ is the character of $V(\mu)$; i.e., the characters ch_μ generate the set of the W-fixed elements.

Now, if $\tilde{\varphi}$ is defined by $\tilde{\varphi} = de^{-\rho}\varphi$, where

$$\rho = \tfrac{1}{2} \sum_{\alpha > 0} \alpha, \quad d = \sum_{w \in W} \epsilon(w) e^{w(\rho)},$$

[$\epsilon(w)$: signature of the transformation w], we have $\tilde{\varphi} \in \mathbb{Z}[P]$, and accordingly

$$\tilde{\varphi} = \sum_{\eta \in P} \tilde{\varphi}_\eta e^\eta.$$

The Weyl formula for the characters is written as follows

$$\mathrm{ch}_\mu = d^{-1} \Big(\sum_{w \in W} \epsilon(w) e^{w(\mu+\rho)} \Big). \tag{13}$$

We use this expression in Eq. (12) and we observe that, for all $\mu \in P_{++}$, $w(\rho+\mu) - \mu \in P_{++}$, only if $w = 1$, the identity in the Weyl group, so that it can be concluded easily:

Proposition: Let $\varphi \in \mathbb{Z}[P]^W$ satisfy $\varphi_\mu = 0$ except for a finite number of indices μ. Then, if $\eta \in P_{++}$, we have $\tilde{\varphi}_\eta = \varphi_\eta$.

We turn now to the case in which $\varphi = (\mathrm{ch}_\lambda)^r \equiv \varphi^{(r)}, \lambda \in P_{++}$ and $\tilde{\varphi}^{(r)} = de^{-\rho}\varphi^{(r)}$. We have, therefore,

$$\varphi^{(r)} = \sum_{\mu \in P_{++}} m^r_{\lambda;\mu} \mathrm{ch}_\mu, \tag{14}$$

and

$$\tilde{\varphi}^{(r)} = \sum_{\eta \in P} \tilde{m}^r_{\lambda;\eta} e^\eta. \tag{14'}$$

The following relation is a consequence of the above proposition

$$\tilde{m}^r_{\lambda;\mu} = m^r_{\lambda;\mu} \quad \text{if} \quad \mu \in P_{++}. \tag{15}$$

The possible extension of the proposition to infinite series in $\mathbb{Q}[P]$ can be tackled in the frame of the Weyl unitarian trick of $\mathfrak{g}$ and the analyticity properties of the objects considered. Thus, we can introduce the extended Bessel series $I_{\lambda;\mu}$ and two analytic generating functions for the extended multiplicities $\tilde{m}^r_{\lambda;\mu}$

$$de^{-\rho}(1 - z\mathrm{ch}_\lambda)^{-1} = \sum_{r \geqslant 0} z^r \tilde{\varphi}^{(r)} = \sum_{r \geqslant 0, \mu \in P} \tilde{m}^r_{\lambda;\mu} z^r e^\mu, \tag{16}$$

$$\tilde{I}_{\lambda;\mu}(z) = \sum_{r \geqslant 0} \tilde{m}^r_{\lambda;\mu} z^r / r! \tag{17}$$

$$de^{-\rho}\exp(z\mathrm{ch}_\lambda) = \sum_{\mu \in P} \tilde{I}_{\lambda;\mu}(z) e^\mu. \tag{18}$$

A general expression for the multiplicities involving inner multiplicities, can be obtained from the character formula. We denote by $\Pi(\lambda)$ the weight system of $V(\lambda)$ and by $m(\lambda;\theta)$ the multiplicity of $\theta\in P$ in $\Pi(\lambda)$.

We have by definition

$$\mathrm{ch}_\lambda = \sum_{\theta\in\Pi(\lambda)} m(\lambda;\theta)e^{\theta}. \tag{19}$$

We develop the expression $de^{-\rho}(\mathrm{ch}_\lambda)^r$ in order to apply the relation (15)

$$de^{-\rho}(\mathrm{ch}_\lambda)^r = \sum_{w\in W}\epsilon(w)\sum_{j_\theta}\delta_{r,\sum_{\theta\in\Pi(\lambda)} j_\theta} \times\binom{r}{\ldots j_\theta\ldots}\prod_{\theta\in\Pi(\lambda)}(m(\lambda;\theta))^{j_\theta} \times e^{\sum_\theta j_\theta\theta + w(\rho)-\rho}$$

$\binom{r}{\ldots j_\theta\ldots}$ denotes the multinomial coefficient $r!/\Pi_{\theta\in\Pi(\lambda)}j_\theta!$. It follows that:

$$m^r_{\lambda;\mu} = \sum_{w\in W}\sum_{j_\theta}\delta_{r,\sum_{\theta\in\Pi(\lambda)} j_\theta} \times\delta_{\mu+\rho,\sum_{\theta\in\Pi(\lambda)} j_\theta\theta + w(\rho)} \times\binom{r}{\ldots j_\theta\ldots} \times\prod_{\theta\in\Pi(\lambda)}(m(\lambda;\theta))^{j_\theta}. \tag{20}$$

Now we notice the following relations verified by the extended multiplicities $\tilde m^r_{\lambda;\mu}$. They are a direct consequence of Eq. (14′):

$$\sum_{\mu\in P}\tilde m^r_{\lambda;\mu}e^{\mu} = \Big(\sum_{\mu\in P}\tilde m^{r-1}_{\lambda;\mu}e^{\mu}\Big)\mathrm{ch}_\lambda,$$

$$\sum_{\mu\in P}\tilde m^0_{\lambda;\mu}e^{\mu} = de^{-\rho} = \sum_{w\in W}\epsilon(w)e^{w(\rho)-\rho}.$$

Therefore the extended multiplicities $m^r_{\lambda;\mu}$ satisfy the finite difference equations

$$\tilde m^r_{\lambda;\mu} = \sum_{\theta\in\Pi(\lambda)} m(\lambda;\theta)\tilde m^{r-1}_{\lambda;\mu-\theta}, \tag{21}$$

with the initial conditions

$$\tilde m^0_{\lambda;\mu} = \sum_{w\in W}\epsilon(w)\delta_{\mu,w(\rho)-\rho}.$$

Thus we can obtain numerical expressions for $\tilde m^r_{\lambda;\mu}$ by using standard routines for solving linear finite difference equations.

IV. PARTICULAR CASES OF INTEREST

Explicit expressions for a few multiplicities $m^r_{\lambda;\mu}$ can be obtained by considering the following special cases.

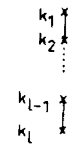


FIG. 1. Graph for $(k_1,k_2,\ldots,k_l)$. Recall that $(k_1,k_2,\ldots,k_l)$ is taken to be zero if any of k_i is strictly negative.

For the classical algebras A_l, B_l, C_l, D_l[3–5] we denoted by $\bar\omega_1,\ldots,\bar\omega_l$ the fundamental weights. In each case, we denote by **1** the chosen $\mathfrak{g}$ module, by $\Pi(\mathbf{1})$ its weight system. We have calculated the weight system by means of the Dynkin algorithm.[4] Then any weight $\eta\in\Pi(\mathbf{1})$ can be written

$$\eta = \sum_{i=1}^{l} k_i\bar\omega_i,$$

where k_i equals zero or $+1$ or -1. Furthermore, any $\eta\in\Pi(\mathbf{1})$ is in the W-orbit of **1**, the inner multiplicity $m(\mathbf{1};\eta)$ is one and by denoting $m(\nu,\nu';\mu)$ the multiplicity of $V(\mu)$ in $V(\nu)\otimes V(\nu')$ ("outer multiplicity") we have

$$m(\mathbf{1},\nu;\nu) = 0 \quad \text{for all } \nu\in P_{++}.$$

The diverse choice of fundamental representations are shown in Table I.

The figures also help to visualize the effect of the tensor product of **1**, with $V(\mu)$, where $\mu = \Sigma_{i=1}^{l}k_i\bar\omega_i$, is an arbitrary dominant weight, on the components k_i of μ. For instance, in the B_2 case, **1**, is chosen to be the four-dimensional fundamental representation $V(\bar\omega_2)$ (spinor representation[3]), and for any weight $\mu = k_1\bar\omega_1 + k_2\bar\omega_2$, $k_i\geqslant 0$, we denote $V(\mu)$ as well as μ by (k_1,k_2) and we have the CG reduction

$$(0,1)\otimes(k_1,k_2) = (k_1,k_2+1)\oplus(k_1,k_2-1) \oplus(k_1-1,k_2+1)\oplus(k_1+1,k_2-1). \tag{22}$$

It goes without saying that we define (k_1,k_2) to be zero if k_1 or k_2 is strictly negative. It will now be attractive to read Eq. (22) through the following apparent graphic illustration

Figures 2–5 are concerned with similar processes for the algebras A_l, B_l, C_l, D_l, respectively.

According to Eq. (22) or Eq. (22′), we can compute the multiplicity $m^r_{1;\mu}$ in the B_2 case. The result is

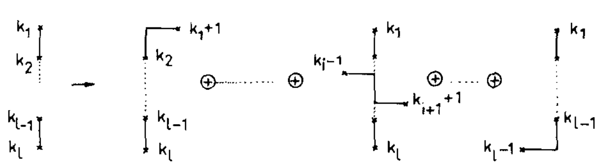


FIG. 2. A_l case; $\mathbf{1} = (1,0,\ldots,0)$; $\dim(\mathbf{1}) = l+1$.

TABLE I. Diverse choices of fundamental representation.

Simple Lie algebra	A_l	B_l	C_l	D_l
Dynkin diagram	$\alpha_1\ \alpha_2\ \alpha_{l-1}\ \alpha_l$	$\alpha_1\ \alpha_2\ \alpha_{l-2}\ \alpha_{l-1}\ \alpha_l$	$\alpha_1\ \alpha_2\ \alpha_{l-2}\ \alpha_{l-1}\ \alpha_l$	$\alpha_1\ \alpha_2\ \alpha_{l-3}\ \alpha_{l-2}\ \alpha_{l-1}\ \alpha_l$
Highest weight of chosen representation	$\bar\omega_1$	$\bar\omega_l$	$\bar\omega_1$	$\bar\omega_1$

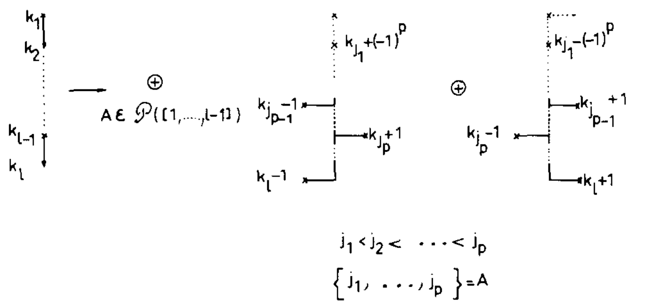
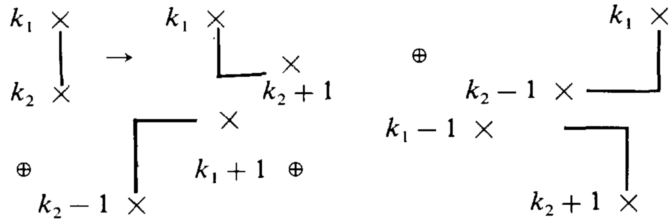


FIG. 3. B_l case; $\mathbf{1} = (0,0,\ldots,1)$; $\dim(\mathbf{1}) = 2^l$. We denote by $\mathscr{P}(X)$ the class of all subsets of X ("power set of X") and by $[1,\ldots,n]$ the set of all integers i such that $1 \leqslant i \leqslant n$.



(22′)

$$m^r_{\mathbf{1};\mu} = \delta_{r,2k_1+k_2+2q} \times \frac{r!(r+2)!(k_1+1)(k_2+1)(k_1+k_2+2)(2k_1+k_2+3)}{q!(q+k_1+1)!(q+k_1+k_2+2)!(q+2k_1+k_2+3)!}. \tag{23}$$

The presence in the above expression of the Kronecker symbol, i.e., $r - 2k_1 - k_2$ must be even, means a "2-ality" phenomenon exactly like the SU(2) case illustrated by Eq. (7).

Formula (23) can be rewritten by using the bilinear form $\langle\ \rangle$ on $\mathfrak{h}^* \times \mathfrak{h}$, deduced from the Killing form restricted to $\mathfrak{h} \times \mathfrak{h}$ where $\mathfrak{h}$ is the Cartan subalgebra of B_2 corresponding to the fundamental weights $\bar{\omega}_1$, and $\bar{\omega}_2$. We recall that $H_\alpha \in \mathfrak{h}$ is defined by: $\langle \alpha, H_\alpha \rangle = 2$ for some $\alpha \in R$ and $\langle \bar{\omega}_i, H_{\alpha_j} \rangle = \delta_{ij}$, where α_j, $j = 1$ or 2, is a basis element of R.[3] Equation (23) now reads

$$m^r_{\mathbf{1};\mu} = \delta_{r,2k_1+k_2+2q} \frac{r!(r+2)!}{q!} \times \frac{\Pi_{\alpha>0} \langle \mu+\rho, H_\alpha \rangle}{\Pi_{\alpha>0, \alpha\neq\alpha_2} (q + \langle \mu+\rho, H_\alpha \rangle)!}. \tag{24}$$

Similar expressions can be given for A_l, for all l, and the chosen representation $\mathbf{1}$ is here $\mathbf{1} = (1,0,\ldots,0)$. For $\mu = \Sigma^l_{i=1} k_i \bar{\omega}_i$, we have

$$m^r_{\mathbf{1};\mu} = \delta_{r,k_1+2k_2+\cdots+lk_l+(l+1)q} \times \frac{r!\Pi_{i<j}(k_i + k_{i+1} + \cdots + k_{j-1} + j - i)}{q!\Pi_{1\leqslant i\leqslant l}(q + k_i + k_{i+1} + \cdots + k_l + l + 1 - i)!}. \tag{25}$$

This equation is consistent with the A_1 case [Eq (7)]. It can be also rewritten by using the "root-weight" outfit

$$m^r_{\mathbf{1};\mu} = \delta_{r,k_1+2k_2+\cdots+lk_l+(l+1)q} \times \frac{r!\Pi_{\alpha>0}\langle \rho+\mu, H_\alpha \rangle}{q!\Pi_{1\leqslant i\leqslant l}(q + \langle \rho+\mu, H_{\alpha_i+\cdots+\alpha_l} \rangle)!}. \tag{26}$$

The factor $\delta_{r,k_1+2k_2+\cdots+lk_l+(l+1)q}$ is a consequence of the concept of "$l+1$-ality", well familiar in the representation theory of SU($l+1$), the Lie algebra of which is the real

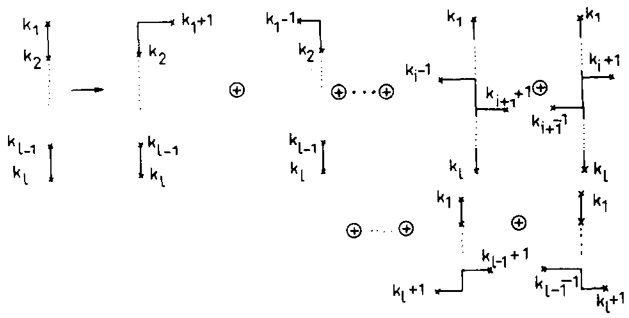


FIG. 4. C_l case; $\mathbf{1} = (1,0,\ldots,0)$; $\dim(\mathbf{1}) = 2l$.

compact form A_l. In this connection, it is useful to employ the Gel'fand–Zeitlin[5] notations, namely,

$$\begin{aligned} &k_1 + k_2 + \cdots + k_l = m_1, \\ &k_2 + \cdots + k_l = m_2, \\ &\vdots \\ &k_l = m_l. \end{aligned} \tag{27}$$

We put $|\mu| = \Sigma^l_{i=1} m_i$. Then, we have

$$m^r_{\mathbf{1};\mu} = \delta_{r,|\mu|+(l+1)q} \frac{r!\Pi_{i<j}(m_i - m_j + j - i)}{q!\Pi_{1\leqslant i\leqslant l}(q + m_i + l + 1 - i)!}. \tag{28}$$

We can note in the numerators of the expressions (24) and (26) the appearance of the numerator of the Weyl formula for the dimension of a $\mathfrak{g}$-module $V(\mu)$[3]:

$$d_\mu \equiv \dim V(\mu) = \prod_{\alpha>0} \frac{\langle \mu+\rho, H_\alpha \rangle}{\langle \rho, H_\alpha \rangle}. \tag{29}$$

This oddness is certainly not accidental and calls for a deeper investigation which could allow us to perform the sum in Eq. (20) for more general cases. However, we think that the above results, Eq. (23) and Eq. (25), are new.

V. ADDITION THEOREMS

The multiplicities $m^r_{\lambda;\mu}$ satisfy a recurrence formula, easily established by exploiting $(\mathrm{ch}_\lambda)^{r+r'} = (\mathrm{ch}_\lambda)^r(\mathrm{ch}_\lambda)^{r'}$,

$$m^{r+r'}_{\lambda;\mu} = \sum_{\nu,\nu' \in P_+} m(\nu,\nu';\mu) m^r_{\lambda;\nu} m^{r'}_{\lambda;\nu'}. \tag{30}$$

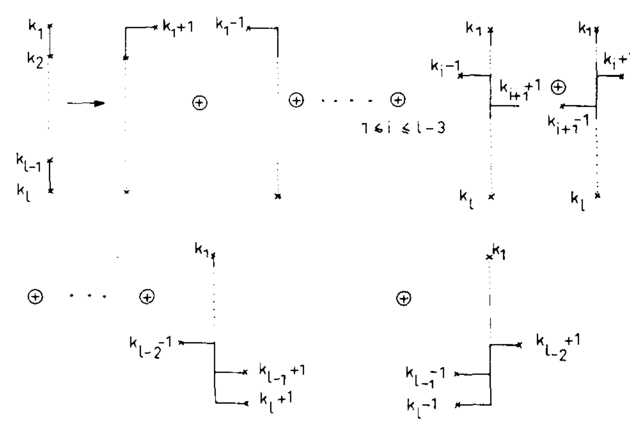


FIG. 5. D_l case; $\mathbf{1} = (1,0,\ldots,0)$; $\dim(\mathbf{1}) = 2l$.

Equation (30) can be particularized

$$m^{r+1}_{\lambda;\mu} = \sum_{\nu\in P_{++}} m(\lambda,\nu;\mu) m^{r}_{\lambda;\nu}. \tag{31}$$

These formulas are of interest for computing both $m^r_{\lambda;\mu}$ and $m(\lambda,\nu;\mu)$.

The $(\lambda;\mu)$ generalized Bessel series exhibit an interesting addition theorem deduced from Eq. (12) by combining $\exp(z+z')\chi_\lambda(u) = \exp z\chi_\lambda(u) \exp z'\chi_\lambda(u)$ with a CG reduction,

$$I_{\lambda;\mu}(z+z') = \sum_{\nu,\nu'} m(\nu,\nu';\mu) I_{\lambda;\nu}(z) I_{\lambda;\nu'}(z'). \tag{32}$$

We turn now to an important addition theorem generalizing to A_l or $\mathrm{GL}(l+1,\mathbb{C})$ which is encountered in the special function theory, particularly in the three Euclidean space group approach to spherical harmonics and Bessel functions.[7]

We shall use henceforward the Gel'fand–Zeitlin parametrization with a light modification with respect to that introduced in Eq. (27) for labeling the unitary irreducible representations D^μ of $\mathrm{SU}(n)$,

$$\mu = (m_{1n}, m_{2n}, \ldots, m_{n-1n}),$$
$$m_{1n} \geqslant m_{2n} \geqslant \cdots \geqslant m_{n-1n} \geqslant 0.$$

Similarly[5] the n-tuple

$$[m]_n = (\mu, m_{nn}) = (m_{1n}, \ldots, m_{n-1n}, m_{nn}) \tag{33}$$

with

$$0 \leqslant m_{nn} \leqslant m_{n-1n} \quad \text{and} \quad |m| = \sum_{i=1}^{n} m_{in}$$

indexes a finite analytic irreducible representation $\mathscr{D}^{[m]_n}$ of $\mathrm{GL}(n,\mathbb{C})$.

Consider $x\in \mathrm{GL}(n,\mathbb{C})$ such that

$$x = |x|^{1/n} u, \quad |x| \equiv \det x, \quad u \in \mathrm{SU}(n). \tag{34}$$

The link between respective representation matrix elements of $\mathrm{GL}(n,\mathbb{C})$ and $\mathrm{SU}(n)$ with respect to Gel'fand basis[5] is given by

$$\mathscr{D}^{[m]_n}(x)_{m^< m^>}$$
$$= |x|^{m_{nn}} \mathscr{D}^{[\underline{m}]_n}(x)_{\underline{m}^< \underline{m}^>} \tag{35}$$
$$= |x|^{|\underline{m}|/n} \mathscr{D}^{\mu}(u)_{\underline{m}^< \underline{m}^>}, \tag{35'}$$

where

$$\begin{pmatrix} m^> \\ [m]_n \\ m^< \end{pmatrix}$$

denotes a double Gel'fand pattern[5]

$$\begin{pmatrix} m^> \\ [m]_n \\ m^< \end{pmatrix} \equiv \begin{array}{ccccccc} & & & m^>_{11} & & & \\ & & & \vdots & & & \\ & m^>_{1n-1} & m^>_{2n-1} & \cdots & m^>_{n-1n-1} & & \\ m_{1n} & & m_{2n} & \cdots & m_{n-1n} & & m_{nn} \\ & m^<_{1n-1} & m^<_{2n-1} & \cdots & m^<_{n-1n-1} & & \\ & & & \vdots & & & \\ & & & m^<_{11} & & & \end{array} \tag{36}$$

and

$$\begin{pmatrix} \underline{m}^> \\ [\underline{m}]_n \\ \underline{m}^< \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \underline{m}^> \\ \mu \\ \underline{m}^< \end{pmatrix}$$

are deduced from the former by subtracting m_{nn} from each of its indices,

$$\underline{m}_{ij} \equiv m_{ij} - m_{nn}.$$

Let $\chi_{[m]_n}(x)$ [resp. $\chi_\mu(u)$] be the character of the representation $[m]_n$ of $\mathrm{GL}(n,\mathbb{C})$ [resp. μ of $\mathrm{SU}(n)$],

$$\chi_{[m]_n}(x) = \sum_{m^< = m^>} \mathscr{D}^{[m]_n}(x)_{m^< m^>}.$$

We introduced in Sec. 4 the particular A_{n-1}-module $\mathbf{1} = \underbrace{(1,0,\ldots,0)}_{n-1}$. For $x = |x|^{1/n}u$ and $[m]_n = (\mathbf{1},0)$, i.e., for the n-dimensional matrix representation of GL $(n,\mathbb{C})$, we have

$$\chi_{[m]_n}(x) = \operatorname{Tr} x = |x|^{1/n} \operatorname{Tr} u = |x|^{1/n}\chi_{\mathbf{1}}(u),$$

and the use of Eq. (12) leads to

$$\exp(\operatorname{Tr} x)$$
$$= \sum_\mu I_{\mathbf{1};\mu}(|x|^{1/n})\chi_\mu(u)$$

$$= \sum_{[m]_n=(\mu,0)} |x|^{-|m|/n} I_{\mathbf{1};\mu}(|x|^{1/n})\chi_{[m]_n}(x). \tag{37}$$

It is apparent that expansion (37) can be extended analytically to any matrix x in $\mathrm{GL}(n,\mathbb{C})$.

We consider now two elements x' and x'' of $\mathrm{GL}(n,\mathbb{C})$ such that $x' + x''$ is in $\mathrm{GL}(n,\mathbb{C})$. Let u be an element of $\mathrm{SU}(n)$ and z a complex variable. By expanding according to (37) on $\exp\operatorname{Tr} z(x'+x'')u = (\exp\operatorname{Tr} zx'u)(\exp\operatorname{Tr} zx''u)$ and by using the main matrix representation group property

$$\chi_{[m]_n}(xu)$$
$$= \sum_{m^<} \mathscr{D}^{[m]_n}(xu)_{m^< m^<}$$
$$= \sum_{m^<, m^>} \mathscr{D}^{[m]_n}(x)_{m^< m^>} \mathscr{D}^{\mu}(u)_{\underline{m}^> \underline{m}^<},$$

we obtain

$$\sum_{\substack{[m]_n=(\mu,0) \\ m^<, m^>}} |x'+x''|^{-|m|/n} I_{\mathbf{1};\mu}(z|x'+x''|^{1/n})$$
$$\times \mathscr{D}^{[m]_n}(x'+x'')_{m^< m^>} \mathscr{D}^{\mu}(u)_{\underline{m}^> \underline{m}^<}$$

$$
= \sum_{\substack{[m']_n=(\mu',0)\\ [m'']_n=(\mu'',0)}} \sum_{\substack{m'^<,m'^>\\ m''^<,m''^>}} |x'|^{-|m'|/n}|x''|^{-|m''|/n}
\times I_{1;\mu'}(z|x'|^{1/n}) I_{1;\mu''}(z|x''|^{1/n})
\times \mathscr{D}^{[m']_n}(x')_{m'^< m'^>} \mathscr{D}^{[m'']_n}(x'')_{m''^< m''^>}
\times \mathscr{D}^{\mu'}(u)_{m'^> m'^<} \mathscr{D}^{\mu''}(u)_{m''^> m''^<} . \tag{38}
$$

The product of two representation matrix elements of SU(n) is transformed in a sum of matrix elements by means of a CG reduction procedure

$$
\mathscr{D}^{\mu'}(u)_{m'^> m'^<} \mathscr{D}^{\mu''}(u)_{m''^> m''^<}
= \sum_{b,\mu,m^>,m^<} \langle \mu'\mu'' m'^> m''^> | \mu b m^> \rangle
\times \langle \mu'\mu'' m'^< m''^< | \mu b m^< \rangle^* \mathscr{D}^{\mu}(u)_{m^> m^<} . \tag{39}
$$

The index b indicates here a summation about the equivalent representations.

A first GL(n,ℂ) addition theorem is established by identifying the coefficients of $\mathscr{D}^{\mu}(u)_{m^> m^<}$ on both sides of Eq. (38),

$$
|x'+x''|^{-|\mu|/n} I_{1;\mu}(z|x'+x''|^{1/n}) \mathscr{D}^{(\mu,0)}(x'+x'')_{m^< m^>}
= \sum_{\mu',\mu''} \sum_{\substack{m'^<,m'^>\\ m''^<,m''^>}} \sum_b \langle \mu'\mu'' m'^> m''^> | \mu b m^> \rangle
\times \langle \mu'\mu'' m'^< m''^< | \mu b m^< \rangle^* |x'|^{-|\mu'|/n}|x''|^{-|\mu''|/n}
\times I_{1;\mu'}(z|x'|^{1/n}) I_{1;\mu''}(z|x''|^{1/n})
\times \mathscr{D}^{(\mu',0)}(x')_{m'^< m'^>} \mathscr{D}^{(\mu'',0)}(x'')_{m''^< m''^>} . \tag{40}
$$

A second GL(n,ℂ) addition theorem is then deduced from the preceding one by expanding the generalized Bessel series and identifying the coefficients of identical powers of z. For this purpose, we make use of the homogeneity property (35) and we put

$$
[m]_n = (m_{1n}, m_{2n}, \ldots, m_{nn}), \quad |m| = \sum_{i=1}^{n} m_{in},
$$

$$
\underline{\mu} = (m_{1n} - m_{nn}, m_{2n} - m_{nn}, \ldots, m_{n-1n} - m_{nn}), \text{ etc. } \ldots
$$

We have then

$$
\mathscr{D}^{[m]_n}(x'+x'')_{m^< m^>}
= \sum_{[m']_n,[m'']_n} \delta_{|m|,|m'|+|m''|} \frac{|m|!}{|m'|!|m''|!}!
\times \frac{m_{1;\underline{\mu}'}^{|m'|} m_{1;\underline{\mu}''}^{|m''|}}{m_{1;\underline{\mu}}^{|m|}} \sum_{\substack{m'^<,m'^>\\ m''^<,m''^>}} \sum_b \langle \underline{\mu}'\underline{\mu}'' \underline{m}'^> \underline{m}''^> | \underline{\mu} b \underline{m}^> \rangle
\times \langle \underline{\mu}'\underline{\mu}'' \underline{m}'^< \underline{m}''^< | \underline{\mu} b \underline{m}^< \rangle^*
\times \mathscr{D}^{[m']_n}(x')_{m'^< m'^>} \mathscr{D}^{[m'']_n}(x'')_{m''^< m''^>} . \tag{41}
$$

These new and striking addition theorems appear as a straight generalization of those obtained by Talman (Formulas 12.40, 12.41, 12.42a, 12.54, and 12.55a of Ref. 7) for ordinary Bessel and spherical harmonics.

They are also related with certain formulas, established by Hua (Formulas 5.3.1 and 5.3.2 of Ref. 8) and by Gazeau *et al.* [Formulas (61) of Ref. 9)].

As an illustration the above results, we return to the GL(2,ℂ)–SU(2) case for which the formulas can seem more familiar to reader. The notations of Talman[7] will be used.

Let $x(x_{ij})$ be a complex nonsingular 2×2 matrix.

Let us introduce the homogeneous polynomials $\mathscr{D}^j(x)_{m_1m_2}$, j integer or half integer and $0 \leqslant |m_1|, |m_2| \leqslant j$, which constitute an analytic continuation to GL(2,ℂ) of the matrix elements $\mathscr{D}^j(u)_{m_1m_2}$, $u \in$SU(2), of the unitary irreducible representations of SU(2). The connection with the precedent notations is the following.

Let q be an integer. We associate with the set of indices $\{j, m_1, m_2, q\}$ the double Gel'fand pattern

$$
\begin{pmatrix} m^> \\ [m]_2 \\ m^< \end{pmatrix} \equiv \begin{pmatrix} & m_{11}^> & \\ m_{12} & & m_{22} \\ & m_{11}^< & \end{pmatrix},
$$

by putting

$$
m_{22} = q, \quad m_{12} = 2j + q,
$$

$$
m_{11}^< = j - m_1 + q, \quad m_{11}^> = j - m_2 + q,
$$

and we define

$$
\mathscr{D}^{[m]_2}(x)_{m^< m^>} \equiv |x|^q \mathscr{D}^j(x)_{m_1m_2} . \tag{42}
$$

It is apparent that this polynomial is a finite irreducible analytic representation matrix element of GL(2,ℂ). Its expression is given by

$$
|x|^q \mathscr{D}^j(x)_{m_1m_2} = (\sigma^j_{m_1}\sigma^j_{m_2})^{-1}|x|^q
\times (-1)^{m_1-m_2} \sum_t \frac{(x_{11})^{j-m_2-t}}{(j-m_2-t)!} \frac{(x_{22})^{j+m_1-t}}{(j+m_1-t)!}
\times \frac{(x_{21})^t}{t!} \frac{(x_{12})^{t+m_2-m_1}}{(t+m_2-m_1)!}, \tag{43}
$$

for

$$
x = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \in \mathrm{GL}(2,\mathbb{C}),
$$

$$
\sigma^j_m = [(j+m)!(j-m)!]^{-1/2} .
$$

We shall express the SU(2) CG reduction [Eq. (39)] by using preferably the Wigner 3-j symbols. Expressions (40) and (41) are now written

$$
[\det(x'+x'')]^{-(j+1/2)} I_{2j+1}\,[2z(\det(x'+x''))^{1/2}]
\times \mathscr{D}^j(x'+x'')_{m_1m_2} = (-1)^{m_1-m_2}/2z
\times \sum_{j'j''} \sum_{\substack{m_1',m_2'\\ m_1'',m_2''}} [j'][j''] \begin{pmatrix} j' & j'' & j \\ m_1' & m_1'' & -m_1 \end{pmatrix}
\times \begin{pmatrix} j' & j'' & j \\ m_2' & m_2'' & -m_2 \end{pmatrix}
\times (\det x')^{-(j'+1/2)} (\det x'')^{-(j''+1/2)}
\times I_{2j'+1}\,(2z(\det x')^{1/2}) I_{2j''+1}\,(2z(\det x'')^{1/2})
\times \mathscr{D}^{j'}(x')_{m_1'm_2'} \mathscr{D}^{j''}(x'')_{m_1''m_2''} , \tag{44}
$$

where $[j] \equiv 2j+1$ and $I_\nu(z)$ is a modified Bessel function [Eq. (9)],

$$
[\det(x'+x'')]^q \mathscr{D}^j(x'+x'')_{m_1m_2}
$$

$$
\begin{aligned}
&= (-1)^{m_1-m_2} \sum_{\substack{j'j'' \\ q',q''}} \delta_{j+q,j'+j''+q'+q''} \\
&\times \frac{q!(q+2j+1)!}{q'!q''!(q'+2j'+1)!(q''+2j''+1)!} \\
&\times \sum_{\substack{m_1'm_2' \\ m_1''m_2''}} [j'][j''] \begin{pmatrix} j' & j'' & j \\ m_1' & m_1'' & -m_1 \end{pmatrix} \\
&\times \begin{pmatrix} j' & j'' & j \\ m_2' & m_2'' & -m_2 \end{pmatrix} \\
&\times (\det x')^{q'} (\det x'')^{q''} \\
&\times \mathscr{D}^{j}(x')_{m_1'm_2'} \mathscr{D}^{j''}(x'')_{m_1''m_2''}.
\end{aligned}
$$

Taking $q = 0$ yields a surprisingly elegant formula, already used by one of us in another work [Ref. 10, Appendix A, formula (A5)],

$$
\begin{aligned}
&\sigma^{j}_{m_1m_2} \mathscr{D}^{j}(x'+x'')_{m_1m_2} \\
&\quad = \sum_{j'',m_1'',m_2''} \sigma^{j-j''}_{m_1-m_1''m_2-m_2''} \\
&\quad \times \mathscr{D}^{j-j''}(x')_{m_1-m_1''m_2-m_2''} \sigma^{j''}_{m_1''m_2''} \\
&\quad \times \mathscr{D}^{j''}(x'')_{m_1''m_2''}\,. \qquad (46)
\end{aligned}
$$

Note added in proof: $\sigma^{j}_{m_1m_2} = \sigma^{j}_{m_1} \sigma^{j}_{m_2}$.

VI. CONCLUDING REMARKS

We have outlined a generalization of the ordinary Bessel functions group theoretically. The latter exhibit remarkable properties in the frame of the special function theory. Reciprocally, it would be interesting to investigate the link between the $(\lambda;\mu)$ generalized Bessel series and the special functions, particularly from the standpoint of the differential equations which they probably satisfy.

For instance in a special case already studied for A_l, we have

$$
\begin{aligned}
I_{1;\mu}(z) &= z^{\langle \mu, H_{(l+1)\bar{\omega}_l} \rangle} \\
&\times \prod_{\alpha>0} \langle \mu+\rho, H_\alpha \rangle \times \prod_{i=1}^{l} \Gamma(\langle \mu+\rho, H_{\alpha_i+\cdots+\alpha_l} \rangle + 1) \\
&\times {}_0F_l(;\ldots,\langle \mu+\rho, H_{\alpha_i+\cdots+\alpha_l} \rangle + 1,\ldots;z), \qquad (47)
\end{aligned}
$$

where ${}_pF_q$ is a generalized hypergeometric function.[1]

[1] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer–Verlag, Berlin, 1966).

[2] C. Bretin and J. P. Gazeau, in *Proceedings of the Sixth International Colloquium in Group Theoretical Methods in Physics*, Tübingen, 1977; *Lecture Notes in Physics* **79** (Springer-Verlag, Berlin, 1978).

[3] N. Bourbaki, *Groupes et Algèbres de Lie* (Hermann, Paris, 1975), Chaps. 4–8.

[4] E. B. Dynkin, Tr. Mosk. Mat. Ova., **1**, 39 (1952) [Am. Math. Soc. Transl. (2), **6**, 245 (1957)].

[5] J. D. Louck, Am. J. Phys. **38**, 3 (1970); A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific Publishers, Warsaw, 1977).

[6] L. Landau and E. Lifshitz, *Mécanique Quantique, Théorie Non Relativiste* (Editions Mir, Moscow, 1967), p. 264.

[7] J. D. Talman, *Special Functions. A Group Theoretic Approach* (Benjamin, New York, 1968) (based on lectures by Eugene P. Wigner).

[8] L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains*; Transl. Math. Mon. **6** (Am. Math. Soc., Providence, Rhode Island, 1963).

[9] J. P. Gazeau, M. Cl Dumont Lepage, and A. Ronveaux, J. Math. Phys. **19**, 734, (1978). See also: J. P. Gazeau, in *Proceedings of the Fifth International Colloquium in Group Theoretical Methods in Physics, Montreal*, 1976 (Academic, New York, 1977).

[10] J. P. Gazeau, J. Math. Phys. **19**, 1041 (1978).

Canonical realizations of Lie superalgebras: Ladder representations of the Lie superalgebra $A(m,n)$

Tchavdar D. Palev
Institute für Theoretische Physik der Technischen, Universität Clausthal, Clausthal-Zellerfeld, West Germany

(Received 15 October 1980; accepted for publication 27 February 1981)

A simple formula for realizations of Lie superalgebras in terms of Bose and Fermi creation and annihilation operators is given. The essential new feature is that Bose and Fermi operators mutually anticommute. The Fock representation of these operators is used in order to construct a class of irreducible finite-dimensional representations of the simple Lie superalgebra $A(m,n)$. The matrix elements of the generators are written down. For $m > 0$ all representations turn out to be nontypical.

PACS numbers: 02.20.Sv, 03.65.Fd

1. INTRODUCTION

The creation and annihilation operators (CAO's) of Bose and Fermi type were initially introduced for the purposes of the quantum physics and continue at present to play an important role in it. For an illustration we mention only that the dynamical variables in any second quantized (elementary particles, nuclear, or solid state) theory are expressed in terms of these operators. At the same time the creation and annihilation operators turned out to be useful in several other branches of physics and mathematics.[1,2] In particular, they were often used for construction of representations of Lie algebras via canonical realizations.

We recall that a Bose (resp., Fermi) canonical realization of a given Lie algebra $\mathscr{A}$ is isomorphic to $\mathscr{A}$ Lie-subalgebra $\hat{\mathscr{A}}$ of the algebra of all polynomials of a certain number n of Bose (resp., Fermi) operators, which we denote as $E(n,0)$ [resp., $E(0,n)$]. In the case of Bose operators this definition is generalized in several ways: $\hat{\mathscr{A}}$ can be taken to be a subalgebra from the quotient algebra $D(n,0)$ of $E(n,0)$[3] or from the completion $\overline{E(n,0)}$ or $\overline{D(n,0)}$ if, for instance, $E(n,0)$ or $D(n,0)$ are metric spaces. The elements of $\overline{D(n,0)}$ are rational functions of CAO's, whereas those of $\overline{E(n,0)}$ are analytical with respect to the operators under consideration. It can be shown that every (finite-dimensional) Lie algebra has canonical realizations.[4,5]

If the canonical realization $\hat{\mathscr{A}}$ is in $E(n,0)$ or $\overline{E(n,0)}$, then one can construct a representation of $\mathscr{A}$ in the way usual for quantum mechanics; namely, considering the Fock representation of the CAO's (which for finite number of operators is also unique). This was the way a lot of physically important representations, usually called ladder representations,[6] were constructed,[7] and among them some new representations of the classical noncompact Lie algebras were found.[8] The representations of $\mathscr{A}$ obtained in this way depend essentially on the choice of the canonical realization $\hat{\mathscr{A}}$. It is important to point out, however, that once a canonical realization $\hat{\mathscr{A}}$ is chosen, the further steps for building ladder representations do not make any use of the underlying algebraical structure of $\mathscr{A}$. Therefore, the same method can be applied for constructing representations of any other (nonassociative) algebra. The nontrivial part is to embed this algebra isomorphically into $E(m,0)$, $E(0,n)$, or, more generally, into a polynomial algebra $E(m,n)$ of both Bose and Fermi operators.

In the present paper we choose the algebraical structure to be a Lie superalgebra (LS). First, we show how one can build canonical realizations of an arbitrary LS in terms of Bose and Fermi operators. The method we use is similar to the one repeatedly used in the literature[9] for constructing canonical realizations of Lie algebras. It is based on the circumstance that if $(A_{ij}), i,j = 1,\dots,n$ is an exact $n\times n$ matrix representation of a Lie algebra $\mathscr{A}$ and $a_1^\xi,\dots,a_n^\xi$ are only Bose or only Fermi creation ($\xi = +$) or annihilation ($\xi = -$) operators, then one particular realization can be obtained from what is sometimes referred to as a trace formula:

$$\hat{\mathscr{A}} = \sum_{i,j=1}^{n} a_i^+ A_{ij} a_j^- . \qquad (1)$$

In case of a Lie superalgebra the trace realization (1) remains formally unaltered. There are, however, two essential differences to be added: (a) for a proper LS one has to use in (1) both Bose and Fermi CAO's, and moreover (b) to assume that Bose and Fermi operators mutually anticommute.

In the second part of the paper we apply the trace formula to the basic classical Lie subalgebra $A(m-1,n-1)$ for any m and n.[10,11] We find and analyze a class of representations of this LS, which we call ladder representations. The trace realization of $A(m-1,n-1)$ is expressed in terms of m pairs of Bose and n pairs of Fermi operators. The corresponding Fock space $W(m,n)$ is infinite-dimensional and decomposes into an infinite direct sum of irreducible finite-dimensional $A(m-1,n-1)$ modules $W_N(m,n), N = 1,2,\cdots$. In the cases $1 < m \neq n$ and $1 = m \neq n, n > N$, every $W_N(m,n)$ carries a nontypical[12] representation of $A(m-1,n-1)$. $W(m,n)$ is a Hilbert space with a metric defined in the way usual for the Fock space. The ladder representations are star representations[13] if the star-operation is Hermitian conjugation.

We notice that the Fock space $W(m,n)$ can be viewed as an infinite-dimensional irreducible $B(m,n)$ module. The latter follows from the observation that any m pairs of Bose operators, considered as odd elements, and n pairs of Fermi operators, considered as even elements, generate the basic classical Lie superalgebra $B(m,n)$.[14] Therefore, the trace re-

 0022-2488/81/102127-05$01.00

presentations of $A(m-1,n-1)$ are realized in the irreducible $A(m-1,n-1)$ submodules of the $B(m,n)$ module $W(m,n)$.

2. CANONICAL REALIZATIONS OF LIE SUPERALGEBRAS[15]

Let α be a function on the set

$$I=(1,2,\dots,m,m+1,\dots,m+n=r), \tag{2}$$

with values α_i, $i\in I$, in Z_2,

$$\alpha_i=\begin{cases}0 & \text{for } i>m\\ 1 & \text{for } i\leqslant m\end{cases}. \tag{3}$$

By $E(m,n)$ we denote the associative algebra of all formal polynomials of the indeterminates ($r=m+n$)

$$a_1^\pm,a_2^\pm,\dots,a_m^\pm,\dots,a_r^\pm, \tag{4}$$

with additional relations

$$a_i^- a_j^+ + (-)^{\alpha_i\alpha_j} a_j^+ a_i^- = \delta_{ij},$$
$$a_i^\xi a_j^\xi + (-)^{\alpha_i\alpha_j} a_j^\xi a_i^\xi = 0,\quad \xi=\pm. \tag{5}$$

To turn $E(m,n)$ into a Z_2-graded algebra, we postulate that the indeterminates $a_1^\pm,\dots,a_m^\pm$ are odd, whereas $a_{m+1}^\pm,\dots,a_r^\pm$ are even elements in $E(m,n)$, i.e.,

$$\deg a_i^\pm=\alpha_i,\quad i\in I. \tag{6}$$

We consider $E(m,n)$ as a Lie subalgebra with a product $\{\ ,\ \}$ defined in a natural way.

$$\{a,b\}=ab-(-)^{\deg a\cdot\deg b}ba,\quad a,b\in E(m,n). \tag{7}$$

The relations (5) show that $a_1^\pm,\dots,a_m^\pm$ are Bose operators, whereas $a_{m+1}^\pm,\dots,a_r^\pm$ are Fermi operators. The essential new feature is that the Bose operators anticommute with the Fermi operators. We call $E(m,n)$ a Bose–Fermi Lie superalgebra.

We now proceed to show that every finite-dimensional LS can be considered as a subalgebra of a certain Bose-Fermi LS. To this end we essentially use

Ado's theorem[10]: Every finite-dimensional Lie superalgebra has a finite-dimensional faithful matrix representation.

Let $\mathscr{A}$ be a LS which we identify with one of its finite-dimensional faithful matrix representations. Then $\mathscr{A}=\mathscr{A}_0\oplus\mathscr{A}_1$ is an algebra[16] of endomorphisms of a certain Z_2-grades space $V=V_0+V_1$, with a product defined by Eq. (7) for any homogeneous elements $a,b\in\mathscr{A}$. By definition $\mathscr{A}$ preserves the grading in V,

$$a_\alpha(V_\beta)\subset V_{\alpha+\beta},\quad a_\alpha\in\mathscr{A}_\alpha,\quad \alpha,\beta\in Z_2. \tag{8}$$

Choose an arbitrary homogeneous basis ($r=m+n$)

$$e_1,e_2,\dots,e_m,\dots,e_r \tag{9}$$

in V, where

$e_1,\dots e_m$ is a basis in V_1 and

$e_{m+1},\dots,e_r$ is a basis in V_0.

In this basis every element $M\in\mathscr{A}$ is represented by an $r\times r$ matrix of the form

$$M=\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}, \tag{10}$$

with α,β,γ, and δ being $m\times m$, $m\times n$, $n\times m$, and $n\times n$ matrices, respectively. The element $M\in\mathscr{A}$ is even (resp., odd) if $\beta=\gamma=0$ (resp., $\alpha=\delta=0$). The algebra $l(m,n)$ of all endomorphisms of V will be of particular interest for us. In the matrix form (10) this is the algebra of all $r\times r$ matrices with a grading as defined above.

Denote

$$A^\pm=(a_1^\pm,\dots,a_m^\pm,\dots,a_r^\pm), \tag{11}$$

and for every $M\in\mathscr{A}$ (with matrix elements M_{ij}) define a mapping $\theta:\mathscr{A}\rightarrow E(m,n)$ as follows (T = transposition)

$$\theta M=A^+M\overset{\mathrm{T}}{A^-}\equiv\sum_{i,j=1}^{r}a_i^+M_{ij}a_j^-. \tag{12}$$

Theorem: The mapping θ is an isomorphism of $\mathscr{A}$ into $E(m,n)$.

Proof: Since $\mathscr{A}$ is a subalgebra of $l(m,n)$, it suffices to prove that θ is an isomorphism of $l(m,n)$ in $E(m,n)$. Let e_{ij} be an $r\times r$ matrix with one on the intersection of the ith row and jth column and zero elsewhere. As a homogeneous basis in $l(m,n)$ we take the matrices e_{ij}, $ij=1,\dots,r$. Since $\deg e_{ij}=\alpha_i+\alpha_j$,

$$\{e_{ij},e_{kl}\}=e_{ij}e_{kl}-(-)^{(\alpha_i+\alpha_j)(\alpha_k+\alpha_l)}e_{kl}e_{ij}. \tag{13}$$

Applying θ to the basis, one obtains

$$\theta e_{ij}=a_i^+a_j^-. \tag{14}$$

Clearly, θ is a linear operator. It preserves the grading since $\deg e_{ij}=\deg a_i^+a_j^-=\alpha_i+\alpha_j$. Moreover, the elements e_{ij} (resp., $a_i^+a_j^-$), $i,j\in I$ are linearly independent in V [resp., in $E(m,n)$]. Hence, θ is injective. To prove that θ preserves the LS product, compute

$$\begin{aligned}
\{\theta e_{ij},\theta e_{kl}\}&=\{a_i^+a_j^-,a_k^+a_l^-\}=a_i^+a_j^-a_k^+a_l^-\\
&\quad-(-)^{(\alpha_i+\alpha_j)(\alpha_k+\alpha_l)}a_k^+a_l^-a_i^+a_j^-\\
&=a_i^+a_j^-a_k^+a_l^-+(-)^{\alpha_j\alpha_k}a_i^+a_k^+a_j^-a_l^-\\
&\quad-(-)^{(\alpha_i+\alpha_j)(\alpha_k+\alpha_l)}a_k^+a_l^-a_i^+a_j^-\\
&\quad-(-)^{(\alpha_i+\alpha_j)(\alpha_k+\alpha_l)+\alpha_i\alpha_l}a_k^+a_i^+a_l^-a_j^-\\
&=a_i^+[a_j^-a_k^+ +(-)^{\alpha_j\alpha_k}a_k^+a_j^-]a_l^-\\
&\quad-(-)^{(\alpha_i+\alpha_j)(\alpha_k+\alpha_l)}a_k^+\\
&\quad\times[a_l^-a_i^+ +(-)^{\alpha_i\alpha_l}a_i^+a_l^-]a_j^-\\
&=\delta_{jk}a_i^+a_l^- -(-)^{(\alpha_i+\alpha_j)(\alpha_k+\alpha_l)}\delta_{il}a_k^+a_j^-.
\end{aligned}$$

Hence,

$$\{\theta e_{ij},\theta e_{kl}\}=\theta\{e_{ij},e_{kl}\}.\blacksquare \tag{15}$$

If $\mathscr{A}$ is a LS isomorphic to a subalgebra $\hat{\mathscr{A}}\subset E(m,n)$, we shall refer to $\hat{\mathscr{A}}$ as to a canonical realization (realization in terms of CAO's) of $\mathscr{A}$ in $E(m,n)$. We call (12) a trace formula for realizations of LS's and the corresponding realization, a trace realization. In the case $m=0$ (resp., $n=0$), (12) reduces to the trace formula (1) for realizations of Lie algebras with Fermi (resp., Bose) creation and annihilation operators.

Identifying $\mathscr{A}$ with its image $\hat{\mathscr{A}}=\theta\mathscr{A}$, we have as an immediate consequence of the theorem.

Corollary 1: Every finite-dimensional Lie superalgebra is a subalgebra of a certain Bose–Fermi algebra, i.e., has a canonical realization.

Applying the trace formula to $l(m,n)$, we conclude

Corollary 2: The linear envelope of all bilinear combinations of creation and annihilation operators is a subalgebra of $E(m,n)$ isomorphic to $l(m,n)$, i.e.,

$$\text{lin. env.}\{a_i^+ a_j^- | i,j \in I\} = l(m,n). \tag{16}$$

The algebra $E(m,n)$ contains also other LS's. A more detailed analysis shows, for instance, that[14]

$$\text{lin. env.}\{\{a_i^\xi, a_j^\eta\} | i,j \in I; \xi,\eta = \pm\} = D(m,n), \tag{17}$$

whereas

$$\text{lin. env.}\{a_i, \{a_i, a_j\} | i,j \in I; \xi,\eta = \pm\} = B(m,n). \tag{18}$$

Here $B(m,n)$ and $D(m,n)$ are the classical simple orthogonal-symplectic LS's [See Refs. 10 and 11)]. It turns out that any third order polynomial (with respect to the LS operations) of the CAO's can be expressed as a linear combination of CAO's. This observation together with (18) gives:

Corollary 3: Any m pairs of Bose operators and n pairs of Fermi operators considered as elements of a Bose–Fermi Lie superalgebra generate the algebra $B(m,n)$.

3. LADDER REPRESENTATIONS OF *A(m,n)*

In this section we construct and study a class of finite-dimensional representations of the LS $A(m,n)$, which we call ladder representations. For this purpose we use a trace realization obtained from the defining matrix representation of $A(m,n)$.

A. Preliminaries and notation

Here we list some of the properties of the LS's $l(m,n)$ and $A(m-1,n-1)$ which can be found in Refs. 10–12, or are a consequence of the results contained therein. Whenever possible, we also follow the notation of these papers.

We consider the algebra $l(m,n)$ always in its matrix realization (10). Let

$$g_{ij} = -(-)^{\alpha_j}\delta_{ij}, \quad i,j \in I. \tag{19}$$

The algebra $l(m,n)$ contains a one-codimensional ideal

$$sl(m,n) = \text{lin. env.}\{g_{ii}e_{ii} - g_{jj}e_{jj}, e_{kl} | k \neq l;\ i,j,k,l \in I\}. \tag{20}$$

If $m \neq n$, $sl(m,n)$ is a simple LS. The algebra $sl(n,n)$ contains a one-dimensional ideal $\lambda\, 1_{2n}$ spanned on the unit matrix 1_{2n}. The basic simple LS $A(m-1,n-1)$ is defined as

$$A(m-1,n-1) = sl(m,n), \quad m \neq n,\ m,n > 0, \tag{21}$$

$$A(n-1,n-1) = sl(n,n)/1_{2n}, \quad n > 0. \tag{22}$$

The grading on $A(m-1,n-1)$ is the one induced from $l(m,n)$. In this paper we consider for simplicity only the algebras (21). As a Cartan subalgebra of $A(m-1,n-1)$ we choose

$$H = \text{lin. env.}\{g_{ii}e_{ii} - g_{jj}e_{jj} | i,j \in I\}. \tag{23}$$

Then

$$A(m-1,n-1) = N^- \oplus H \oplus N^+, \tag{24}$$

where

$$N^+ = \text{lin. env.}\{e_{ij} | i < j \in I\}, \tag{25}$$

$$N^- = \text{lin. env.}\{e_{ij} | i > j \in I\}, \tag{26}$$

are subalgebras spanned on all positive root vectors $e_{ij}, i < j \in I$ and on all negative root vectors $e_{ij}, i > j \in I$, respectively. The Killing form $(\ ,\ )'$ of $l(m,n)$, restricted to its Cartan subalgebra

$$H' = \text{lin. env.}\{e_{ii} | i \in I\}, \tag{27}$$

reads as

$$(e_{ii}, e_{jj})' = 2(m-n)g_{ij} - 2(-)^{\alpha_i + \alpha_j}. \tag{28}$$

On H the form $(\ ,\ )'$ coincides with the Killing form $(\ ,\ )$ of $A(m-1,n-1)$. Since the second term in (28) vanishes on H, we neglect it and write

$$(e_{ii}, e_{jj}) = 2(m-n)g_{ij}. \tag{29}$$

This equality defines a nondegenerate $(n \neq m)$ bilinear form on H', which restricted to H gives the Killing form of $A(m-1,n-1)$.

Choose as a basis in H' the vectors

$$\epsilon_i = e_{ii}, \quad i = 1,\ldots,r. \tag{30}$$

As an ordered basis in $\overset{\bullet}{H}{}'$ we take the dual to the (30) basis of linear functionals

$$\epsilon^1, \epsilon^2, \ldots, \epsilon^r. \tag{31}$$

The bilinear form on $\overset{\bullet}{H}{}'$ induced from (29) reads

$$(\epsilon^i, \epsilon^j) = g_{ij}/2(m-n). \tag{32}$$

The correspondence between the root vectors and their roots is

$$e_{ij} \longleftrightarrow \epsilon^i - \epsilon^j. \tag{33}$$

Therefore, a root is positive if its first nonzero coordinate in the basis (31) is positive. For the half-sum of the even positive roots minus the half-sum of the odd positive roots one obtains

$$\rho = \frac{1}{2}\sum_{i=1}^{r}[3m + n - 2\alpha_i(m+n) - 2i + 1]\epsilon^i. \tag{34}$$

The operators $(i = 1,2,\ldots,r-1)$

$$e_i = e_{i,i+1}, \quad f_i = e_{i+1,i},$$
$$h_i = e_{ii} - (-)^{\alpha_i + \alpha_{i+1}} e_{i+1,i+1} \tag{35}$$

generate the algebra $A(m-1,n-1)$ and satisfy the relations $([x,y] = xy - yx)$

$$\{e_i, f_j\} = \delta_{ij}h_i, \quad [h_i, h_j] = 0,$$
$$[h_i, e_j] = a_{ij}e_j, \quad [h_i, f_j] = -a_{ij}f_j, \ i,j = 1,\ldots,r-1, \tag{36}$$

where the Cartan matrix $A = (a_{ij})$ reads as

$$a_{ij} = \delta_{ij} + (-)^{\alpha_i + \alpha_{i+1}}\delta_{ij} - \delta_{i,j+1} - (-)^{\alpha_i + \alpha_{i+1}}\delta_{i+1,j}. \tag{37}$$

In particular,

$$a_{mm} = 0, \quad a_{m,m+1} = 1, \ \text{and} \ a_{ii} = 2, \ i \neq m. \tag{38}$$

Moreover,

$$(\rho, \epsilon^i - \epsilon^{i+1}) = \tfrac{1}{2}(\epsilon^i - \epsilon^{i+1}, \epsilon^i - \epsilon^{i+1}) = (g_{ii} + g_{i+1,i+1})/2(m-n). \tag{39}$$

Consider an irreducible $A(m-1,n-1)$ module $V(\Lambda)$ with a highest weight vector $v_\Lambda, \Lambda \in \overset{*}{H}$,

$$hv_\Lambda = \Lambda(h)v_\Lambda,\quad h\in H,\quad N^+v_\Lambda = 0,\quad \deg v_\Lambda = 0. \tag{40}$$

The corresponding representation is finite-dimensional if and only if[12]

$$a_i = \Lambda(h_i) \in Z_+ \quad \text{for any } i\neq m = 1,\ldots,r-1, \tag{41}$$

i.e., for nonnegative integers $a_i, i\neq m$. The number $a_m = \Lambda(h_m)$ can be an arbitrary complex number. A finite-dimensional $A(m-1,n-1)$ module is said to be typical[12] if

$$(\Lambda+\rho,\alpha^{ij})\neq 0 \quad \text{for any } \alpha^{ij} = \epsilon^i - \epsilon^j \in \Delta_1, \tag{42}$$

where

$$\Delta_1 = \{\epsilon^i - \epsilon^j | i\leqslant m, j>m\} \tag{43}$$

is the set of all odd positive roots.

B. Matrix elements and analysis of the ladder representations

Since every finite-dimensional irreducible $l(m,n)$ module is simultaneously $A(m-1,n-1)$-irreducible, we consider in the beginning the representation spaces as $l(m,n)$ modules. The canonical realization we use is the trace realization (16). As a representation space $W(m,n)$ we take the ordinary quantum-mechanical Fock space, namely the Hilbert space generated out of a vacuum vector $|0\rangle$ by means of creation and annihilation operators. By definition the vacuum is a vector which vanishes under the action of the annihilation operators,

$$a_i^-|0\rangle = 0,\quad i\in I. \tag{44}$$

The metric in $W(m,n)$ is uniquely defined from the requirements that the norm of the vacuum be one and the Hermitian conjugate of a_i^+ be equal to a_i^-, $i\in I$. To turn $W(m,n)$ into a Z_2-graded space we assume in addition, that the vacuum is an even element, $\deg|0\rangle = 0$.

As an orthonormal basis in $W(m,n)$ we choose the vectors

$$|p_1,p_2,\ldots,p_r\rangle = \frac{(a_1^+)^{p_1}(a_2^+)^{p_2}\cdots(a_r^+)^{p_r}}{(p_1!\,p_2!\cdots p_r!)^{1/2}}, \tag{45}$$

where

$$p_i \in Z_+ \quad \text{for } i\leqslant m\in I, \qquad p_i = 1,2 \quad \text{for } i>m\in I. \tag{46}$$

For $m>0$ the representation space is infinite-dimensional. A straightforward computation gives

$$a_i^+|\ldots,p_i,\ldots\rangle = (g_{ii})^{p_1+\cdots+p_{i-1}}(1+g_{ii}p_i)^{1/2}|\ldots,p_i+1,\ldots\rangle,$$
$$a_i^-|\ldots,p_i,\ldots\rangle = (g_{ii})^{p_1+\cdots+p_{i-1}}\sqrt{p_i}|\ldots,p_i-1,\ldots\rangle. \tag{47}$$

It is understood that the unwritten indices on the left- and on the right-hand side in (47) are the same. From these equalities one easily calculates the transformation properties of the basis (45) under the action of the $l(m,n)$ generators:

Cartan generators ($\epsilon_i = e_{ii} = a_i^+a_i^-, i\in I$)

$$\epsilon_i|\ldots,p_i,\ldots\rangle = p_i|\ldots,p_i,\ldots\rangle; \tag{48}$$

Positive root vectors ($i<j\in I$)

$$e_{ij}|\ldots,p_i,\ldots,p_j,\ldots\rangle = (g_{ii})^{p_1+\cdots+p_{i-1}}(g_{jj})^{p_1+\cdots+p_{j-1}} \times((1+g_{ii}p_i)p_j)^{1/2}|\ldots,p_i+1,\ldots,p_j-1,\ldots\rangle; \tag{49}$$

Negative root vectors ($i>j\in I$)

$$e_{ij}|\ldots,p_j,\ldots,p_i,\ldots\rangle = (g_{ii})^{p_1+\cdots+p_{i-1}-1}(g_{jj})^{p_1+\cdots+p_{j-1}} \times((1+g_{ii}p_i)p_j)^{1/2}|\ldots,p_j-1,\ldots,p_i+1,\ldots\rangle. \tag{50}$$

Introducing a θ function

$$\theta(k) = \begin{matrix} 0 & k\leqslant 0, \\ 1 & k>1 \end{matrix} \tag{51}$$

one can write (48–50) in an unified form ($i,j\in I$),

$$e_{ij}|\ldots,p_i,\ldots,p_j,\ldots\rangle = (g_{ii})^{p_1+\cdots+p_{i-1}-\theta(i-j)}(g_{jj})^{p_1+\cdots+p_{j-1}} \times[p_j(1+g_{ii}p_i-g_{ij})]^{1/2}|\ldots,p_i+1,\ldots,p_j-1,\ldots\rangle. \tag{52}$$

Consider the finite-dimensional subspace

$$W_N(m,n) = \text{lin. env.}\{|p_1,\ldots,p_r\rangle | p_1+\cdots+p_r = N\}. \tag{53}$$

Clearly, $W_N(m,n)$ is an invariant $l(m,n)$ submodule. Since, moreover, any vector $|p_1,\ldots,p_r\rangle\in W_N(m,n)$ can be transformed by means of the generators of $l(m,n)$ onto any other vector $|p_1',\ldots,p_r'\rangle\in W_N(m,n)$, the subspace $W_N(m,n)$ is irreducible. Hence, $W(m,n)$ resolves into an infinite direct sum of finite-dimensional irreducible $l(m,n)$ modules $W_N(m,n)$,

$$W(m,n) = \bigoplus_{N=0}^{\infty} W_N(m,n). \tag{54}$$

Let $h=\Sigma_{i=1}^r \xi^i\epsilon_i$ be an arbitrary element from the Cartan subalgebra H'. From (48), (49), and (25) one has

$$h|N,0,\ldots,0\rangle = N\xi^1|N,0,\ldots,0\rangle, \qquad N^+|N,0,\ldots,0\rangle = 0. \tag{55}$$

Hence, $|N,0,\ldots,0\rangle$ is the highest weight vector in $W_N(m,n)$. The corresponding highest weight $\Lambda_N\in \overset{*}{H}'$ is

$$\Lambda_N = N\epsilon^1. \tag{56}$$

Thus, we have constructed a countable set of irreducible representations of the LS $l(m,n)$, which we call ladder representations. Every $W_N(m,n)$ is also an irreducible $sl(m,n)$ module. If $m\neq n$, $sl(m,n) = A(m-1,n-1)$ and, therefore, every $W_N(m,n)$ carries an irreducible representation of $A(m-1,n-1)$. Since

$$a_i = \Lambda_N(h_i) = N\delta_{1i},\quad i=1,\ldots,r-1, \tag{57}$$

the signature of the representation $(a_1,a_2,\ldots,a_{r-1})$ with highest weight Λ_N is $(N,0,\ldots,0)$.

Assertion: The representation spaces $W_N(m,n), N\in Z_+$, are nontypical if and only if $m>1$ or $m=1$ and $N<n$.

Proof: Inserting (34) and (56) in (42) and taking α^{jk} to be an odd positive root, i.e., $\alpha^{jk} = \epsilon^j - \epsilon^k, j \leqslant m, k > m$, one obtains

$$(\Lambda_N + \rho, \epsilon^j - \epsilon^k) = (N\delta_{1j} + 2m + 1 - j - k)/2(m - n). \tag{58}$$

For $m > 1$ and $j = k - 1 = m$,

$$(\Lambda_N + \rho, \epsilon^m - \epsilon^{m+1}) = 0, \tag{59}$$

and, therefore, for $m > 1$ all $A(m-1, n-1)$ modules $W_N(m,n)$ are nontypical. If $m = j = 1$,

$$(\Lambda_N + \rho, \epsilon^1 - \epsilon^k) = (N + 2 - k)/2(1 - n),$$
$$k = 2, \ldots, n + 1, \ n > 1. \tag{60}$$

This expression vanishes for $k = N + 2$. Therefore, only those representations are nontypical for which k can take the value $N + 2$, i.e., in the cases $N < n$.■

Thus, all ladder representations of the LS $A(m,n), n > m > 0$ are nontypical, whereas for $A(0,n)$ only a finite set of them has the same property.

We have mentioned already that the CAO's $a_1^{\pm}, \ldots, a_r^{\pm} \in E(m,n)$ generate the canonical realization (18) of the simple LS $B(m,n)$, $m + n = r$. The space $W(m,n)$ can be viewed as an irreducible $B(m,n)$ module with a highest weight vector $|0\rangle$. Therefore, the ladder representations of $A(m-1, n-1)$ appear when this particular representation of $B(m,n)$ is restricted to a representation of its subalgebra $A(m-1, n-1)$, and then decomposed into irreducible representations.

4. CONCLUDING REMARKS

The trace formula for realizations of Lie algebras with creation and annihilation operators was generalized in the present paper to the case of Lie superalgebras. A crucial role for this generalization is played by the assumption that the Bose and Fermi operators mutually anticommute.

Applying the trace formula to the defining matrix representation of the LS, $A(m,n)$, we have constructed a class of irreducible nontypical representations of the algebra. The representations obtained in this way depend essentially on the canonical realization. In general, different realizations will lead to nonequivalent representations. In this respect it is interesting to study various realizations for a given LS. Likewise for the Lie algebras (see Ref. 5), one can try to construct Schur realizations, which lead immediately to irreducible representations. On the other hand, it is of physical interest to find minimal realizations, i.e., for a given algebra $\mathscr{A}$ to determine the minimum number $m + n = r$ of pairs of creation and annihilation operators, for which $\mathscr{A}$ is a subalgebra of $E(m,n)$. The automorphisms of the Bose–Fermi algebra provide a convenient way for generating new realizations out of a given one and, thus, for constructing new representations. In all cases the main difficulty one has to overcome in constructing ladder representations of a given Lie superalgebra $\mathscr{A}$ from a canonical realization $\hat{\mathscr{A}} \subset E(m,n)$ is to decompose the irreducible $B(m,n)$ module $W(m,n)$ into irreducible $\mathscr{A}$ modules.

[1]See, for instance, P. Cordero and G. C. Chirardi, Fortschr. Phys. **20**, 105 (1972) and A. O. Barut, "Some Unusual Applications of Lie-Algebra Representations in Quantum Theory," Preprint ICTP, IC/72/79(1972), where further references can be found.
[2]B. R. Judd in *Group Theory and its Appliations*, edited by E. M. Loebel (Academic, New York, 1968, 1971, 1975); W. Miller, *On Lie Theory and Some Special Functions in Mathematical Physics* (Amer. Math. Soc., Providence, R. I., 1964).
[3]I. M. Gel'fand and A. A. Kirillov, "On the Division Rings Connected with Enveloping Algebras" (in Russian), Preprint of Steclov Inst. Math. (1965); DAN USSR **167**, 503 (1966) [French transl. Publ. Math. France **31** (1966)]; Y. Chow, J. Math. Phys. **10**, 975 (1969).
[4]More generally, every Lie algebra can be realized in terms of paraoperators, H. D. Doebner and T. D. Palev, Acta Phys. Austriaca, Suppl. **VII**, 597 (1970); T. D. Palev, C. R. Acad. Bulg. Sci. **24**, 565 (1971).
[5]For a systematic study of canonical realizations of the classical Lie algebras see the review article P. Exner, M. Havlíček, and W. Lassner, Czech. J. Phys. B **26**, 1213 (1976) and the references therein.
[6]Y. Dothan, M. Gell-Mann, and Y. Ne'eman, Phys. Lett. **17**, 148 (1965).
[7]I. T. Todorov, "Discrete Series of Hermitian Representations of the Algebra $U(p,q)$," Lecture notes, Preprint ICTP IC/66/71 (1966); G. Mack and I. T. Todorov, J. Math. Phys. **10**, 2078 (1969).
[8]For instance, the ladder representations of $U(p,q)$ are not among the representations previously classified by I. M. Gel'fand and M. I. Graev, Izv. Akad. Nauk SSSR, Ser. Math. **29**, 1329 (1965); for more details see Ref. 7.
[9]See, for example, H. Lipkin, *Lie Groups for Pedestrians* (North-Holland, Amsterdam, 1965); T. D. Palev, Preprint ICTP IC/68/23 (1968).
[10]V. G. Kac, Adv. Math. **26**, 8 (1977). We follow the notation of this paper.
[11]V. G. Kac, Comm. Math. Phys. **53**, 31 (1977).
[12]V. G. Kac, Lecture notes in Math. **676**, 597 (1978).
[13]M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146, 155 (1977).
[14]T. D. Palev, "Para-Bose and Para-Fermi Operators as Generators of Orthosymlpectic Lie Superalgebras," Preprint ICTP IC/79/167 (1979) (to appear in J. Math. Phys.)
[15]See also T. D. Palev, C. R. Acad. Bulg. Sci. (to appear).
[16]Unless otherwise stated an algebra means a Lie superalgebra.

Fredholm determinants associated with Wiener integrals

Bruce Nelson [a)]
Institute for Theoretical Physics, [b)] *University of California, Santa Barbara, California 93106*
Benny Sheeks
Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Massachusetts 02173

(Received 10 June 1980; accepted for publication 20 August 1980)

We present a method for evaluating a large class of Fredholm determinants that are associated with the evaluation of certain Wiener integrals. The infinite-dimensional determinant is shown to be equal to a single finite-dimensional determinant.

PACS numbers: 02.30.Cj

INTRODUCTION

A common integral arising in the theory of Wiener integrals is that of the exponential of a quadratic combination of the integration variables. The evaluation of such integrals is reducible to the evaluation of an infinite-dimensional Fredholm determinant, often of the Volterra type. In this paper we show that these infinite-dimensional determinants are equal to simple finite-dimensional determinants (not a limit of such). Our approach makes use of several important mathematical telchniques. First, we use the linear transformations introduced by Cameron and Martin[1] to relate the Fredholm determinant to the determinant of a linear transformation. Second, we use the variation techniques of Schwinger[2] to derive a differential equation for the determinant. Finally, we take advantage of the Gaussian nature of the Wiener measure, as presented in terms of its Fourier transform by Cécile DeWitt-Morette,[3–6] to integrate the differential equation.

WIENER MEASURES AND THEIR FOURIER TRANSFORMS (REF. 6)

We are interested in the space of all paths in R^n that begin at the origin and are parameterized by the interval $T=[0,1]$. Let X be this space and x an element of it. On X there is the Wiener measure $d\gamma(x)$, defined to be the Guassian measure with covariance inf:

$$\int_X d\gamma(x)\, x^\alpha(t)\, x^\beta(s) = \inf(t,s)\, \delta^{\alpha\beta}\,. \tag{1}$$

The fact that γ is Gaussian is enough to allow one to compute all of the finite-dimensional distributions, as was Wiener's original definition. We prefer to define γ in terms of its Fourier transform, however. The Fourier transform is defined on the space X^* dual to X. X^* is essentially the space of bounded measures: For $\mu\in X^*$, the duality is given by

$$\langle \mu,x\rangle \equiv \int_T d\mu_\alpha(t)\, x^\alpha(t)\,, \tag{2}$$

γ is then defined by its Fourier transform:

$$(\mathscr{F}\gamma)(\mu) = \int_X d\gamma(x) \exp(i\langle\mu,x\rangle)$$
$$\equiv \exp\Big(-\frac{1}{2}\int_T d\mu_\alpha(t)\int d\mu_\beta(s)\, \delta^{\alpha\beta} \inf(t,s)\Big). \tag{3}$$

The important fact about the covariance $G^{\alpha\beta}(t,s) \equiv \delta^{\alpha\beta}\inf(t,s)$ of γ is that it is a Green's function of the simple differential operator $-d^2/dt^2$ $(\equiv \mathscr{J}_0(t))$:

$$-\frac{d^2}{dt^2} G^{\alpha\beta}(t,s) = \delta^{\alpha\beta}\,\delta(t-s)\,, \tag{4a}$$

with the boundary conditions

$$G^{\alpha\beta}(t=0,s)=0\,, \tag{4b}$$

$$\frac{dG^{\alpha\beta}}{dt}(t=1,s)=0\,. \tag{4c}$$

These boundary conditions guarantee that G is symmetric:

$$G^{\alpha\beta}(t,s) = G^{\beta\alpha}(s,t)\,. \tag{4d}$$

Equation (4) can be verified by writing

$$\inf(t,s) = t\theta(s-t) + s\theta(t-s)\,, \tag{4e}$$

where θ is the step function. The importance of $-d^2/dt^2$ and the boundary conditions on G is that they in some sense (see Ref. 6 pp. 271–82 for details) make G the inverse of $\int \dot{x}^2\, dt$. This agrees with the property that the Fourier transform of an infinite-dimensional Gaussian measure defined by a matrix A is just the exponential of the inverse matrix A^{-1}. Further, it is by viewing γ in terms of its Fourier transform, or its covariance G, that allows one to define conviently a large class of generalized Wiener measures. In particular, consider the second-order operator $\mathscr{J}$ given by

$$\mathscr{J}_{\alpha\beta} = -\delta_{\alpha\beta}\frac{d}{dt^2} + F_{\alpha\beta}(t)\frac{d}{dt} + R_{\alpha\beta}(t)\,. \tag{5}$$

(We are interested only in the case where F is skew, in particular where $F_{\alpha\beta} = A_{\alpha\beta} - A_{\beta\alpha}$ for some A. Note in the development below that the effects of a d/dt term are trivial for F symmetric and hence trivial in one dimension. This observation is intimately connected with the fact that in mechanics, a linear velocity dependent term in a one-dimensional Lagrangian is equivalent to a total time derivative and hence irrelevant. Further, we want $\mathscr{J}$ to be self-adjoint; thus $R_{\alpha\beta} - R_{\beta\alpha} = dA_{\beta\alpha}/dt$.) We define the Green's function G associated with $\mathscr{J}$ by

$$\mathscr{J}_{\alpha\beta}(t)\, G^{\alpha\gamma}(t,s) = \delta^\gamma_\alpha\, \delta(t-s)\,, \tag{6a}$$

[a)] Present address: Department of Physics, University of Utah, Salt Lake City, Utah 84112.
[b)] Research supported by the National Science Foundation, Grant No. PHY77-27084.

with the boundary conditions

$$G^{\beta\gamma}(t=0,s)=0\,, \tag{6b}$$

$$\frac{d}{dt}G^{\beta\gamma}(t=1,s)=A^{\beta}_{\alpha}(t=1)\cdot G^{\alpha\gamma}(t=1,s)\,. \tag{6c}$$

These boundary conditions are those that define the symmetric Green's function

$$G^{\alpha\beta}(t,s)=G^{\beta\alpha}(s,t)\,. \tag{6d}$$

This new Green's function can be used as a covariance to define a new Gaussian measure $\bar{\gamma}$ on X:

$$(\mathscr{F}\bar{\gamma})(\mu)=\exp\left(-\frac{1}{2}\int_T d\mu_\alpha(t)\int_T d\mu_\beta(s)\,G^{\alpha\beta}(t,s)\right). \tag{7}$$

In the same way that the original Wiener measure corresponded in some way to $\exp(-\frac{1}{2}\int_T \dot{x}^2(t)\,dt)$, this new measure corresponds to

$$\exp\left(-\frac{1}{2}\int_T(\dot{x}^2+x\cdot F\cdot\dot{x}+x\cdot R\cdot x)\,dt\right).$$

The Green's function G can be expressed conveniently in terms of θ functions and homogeneous solutions of $\mathscr{J}$. These expressions are generalizations of Eq. (4e) for the inf. Thus, consider the two functions J and K defined by

$$\mathscr{J}_{\alpha\beta}(t)\,J^{\beta\alpha}(t,s)=0\,,$$
$$J^{\beta\gamma}(t=s,s)=0\,,$$
$$\frac{d}{dt}J^{\beta\gamma}(t=s,s)=\delta^{\beta\gamma}\,, \tag{8}$$

and

$$\mathscr{J}_{\alpha\beta}(t)\,K^{\beta\gamma}(t,s)=0\,,$$
$$K^{\beta\gamma}(t=s,s)=\delta^{\beta\gamma}\,,$$
$$\frac{d}{dt}K^{\beta\gamma}(t=s,s)=-\tilde{A}^{\gamma\beta}(s)\,. \tag{9}$$

Let N represent the inverse of K:

$$N_{\alpha\beta}(t,s)\,K^{\beta\gamma}(s,t)=\delta^{\gamma}_{\alpha}\,. \tag{10}$$

Then G can be written as

$$\begin{aligned}G(t,s)=\;&\theta(s-t)\,J(t,0)\,\tilde{N}(0,1)\,\tilde{K}(1,s)\\&-\theta(t-s)\,K(t,1)\,N(1,0)\,J(0,s)\,.\end{aligned} \tag{11}$$

[Here, indices have been suppressed, juxtaposition of matrices implies matrix multiplication, and $\tilde{K}^{\alpha\beta}(t,s)\equiv K^{\beta\alpha}(s,t)$.] We point out here that the determinant of the matrix $K^{\alpha\beta}(0,1)$ will be equal to the Fredholm determinant we wish to calculate. Thus we will have reduced the calculation to solving the second order, linear equation (9). Cameron and Martin[7] first showed that a Fredholm determinant (for one-dimensional paths) could be written in terms of a Sturm–Liouville equation. The extension to cases of paths in more than one dimension was given by Cecile DeWitt-Morette.[5] Our discussion is the first that we know of that successfully treats cases with a first order term in $\mathscr{J}$.

THE GAUSSIAN INTEGRAL

We wish now to evaluate the integral

$$I\equiv\int_X d\gamma(x)\exp(-\tfrac{1}{2}H(x,x))\,, \tag{12}$$

where γ is the usual Wiener measure and H is a quadratic combination of the x's:

$$\begin{aligned}H(x,x)\equiv&\int_T(x(t)\cdot F(t)\,dx(t)+x(t)\cdot R(t)\cdot x(t)\,dt)\\&+x(1)\cdot\tilde{A}(1)\cdot x(1)\,.\end{aligned} \tag{13}$$

We choose to include explicitly the boundary term. If we wanted to calculate an integral without it, we would include them as above, and then compensate by integrating (with the new measure, discussed below) the simple function $\exp(x(1)\cdot\tilde{A}(1)\cdot x(1))$. Note that the boundary term is absent if $F=0$ (remember $F=A-\tilde{A}$).

Because H is quadratic, it is possible to define a new measure $\bar{\gamma}$ that incorporates the $\exp(-(1/2)H)$ term into $d\gamma$

$$d\bar{\gamma}(x)\sim d\gamma(x)\exp(-\tfrac{1}{2}H(x,x))\,. \tag{14}$$

It proves convenient to define the Fourier transforms of both γ and $\bar{\gamma}$ as simply the exponential of a quadratic form. As such, both γ and $\bar{\gamma}$ are normalized to 1. In general, the integral I will not be 1, and the correspondence (14) is only a proportionality. Thus

$$D\,d\bar{\gamma}(x)=d\gamma(x)\exp(-\tfrac{1}{2}H(x,x))\,. \tag{15}$$

D will be a Fredholm determinant. However, it will be independent of x. This fact is best argued by showing that $\bar{\gamma}$ can be obtained from γ by a *linear* change of variables, and hence the Jacobian is constant. (See Ref. 6 for the details of this argument.)

Viewed in this way as the Jacobian of a linear transformation, there are many ways to compute D. One direct way is to approximate the integral in H by a sum. Each term in the sum is a cylindrical function, so the Wiener integral can be evaluated. Then the limit can be taken in passing from a sum to an integral. In some cases, the limit can actually be evaluated. This is Feynman's approach to path integration.[8] (More precisely, this is his *definition* of the integral.). Next, one can find the appropriate change of variables, approximate it by discretized versions, evaluate the Jacobians of the discretized version, and pass to the limit. This is the approach taken originally by Cameron and Martin.[1] Once one has an explicit form of the transformation M that takes γ to $\bar{\gamma}$, one can generalize the finite-dimensional result that

$$\det M=\exp(\mathrm{tr}\ln M) \tag{16a}$$

to the infinite-dimensional case. M will in fact be of the Volterra type and $\ln M$ can be defined. (See Ref. 6, p. 280, for a discussion of this.) The trouble with this approach is that the usual definition of the trace of a continuous matrix neglects important contributions due to boundary terms and hence gives the wrong answer. Finally, one can generalize the relationship

$$\left(\frac{\det G}{\det G_0}\right)^{1/2}=\det M \tag{16b}$$

from the case of finite-dimensional Gaussian measures to the infinite-dimensional case. One has thus reduced the problem of defining $\det M$ to that defining $(\det G/\det G_0)$.

This ratio can be defined and evaluated in two ways. The first is to discretize both G_0 and G, use their representations in terms of the θ functions and the homogeneous solutions J and K, evaluate the ratio of the finite-dimensional determinants, and then pass to the appropriate limit in the discretization.[5] The second is to use an eigenfunction expansion of G_0 and G and use the product of the eigenvalues to define the determinants.[9] (Part of our calculation is similar to that in Ref. 9, despite the difference in approaches.)

We propose to use the variational principle of Schwinger to derive a differential equation for I. We will then use the expression for G in terms of J and K to integrate this equation. We will show that

$$I = D = \left| \frac{\det K_0(0,1)}{\det K(0,1)} \right|^{1/2}, \tag{17}$$

where K_0 and K are the homogeneous solutions of $\mathscr{J}_0$ and $\mathscr{J}$ with the boundary conditions by (9). The det here is the simple finite-dimensional determinant.

For simplicity of discussion, consider first the case when $A = F = 0$. The modifications needed for $F \neq 0$ are presented in the Appendix. Then

$$H(x,x) = \int_T x(t)\cdot R(t)\cdot x(t)\, dt. \tag{18}$$

We are then interested in the integral

$$I = \int_X d\gamma(x) \exp(-\tfrac{1}{2} H(x,x)). \tag{19}$$

As argued in Ref. 6, p. 271, this can be written in terms of a new measure $\bar{\gamma}$ as

$$I = \int_X D\, d\bar{\gamma}(x), \tag{20}$$

where the covariance G of $\bar{\gamma}$ is the Green's function of

$$\mathscr{J}(t) = -\frac{d^2}{dt^2} + R(t). \tag{21}$$

Since $\bar{\gamma}$ is normalized, we know that

$$I = D \int d\bar{\gamma} = D; \tag{22}$$

thus we want to evaluate D. To do this, we consider a whole family of integrals, parameterized by $\lambda \in [0,1]$:

$$D(\lambda) = \int_X D(\lambda)\, d\bar{\gamma}_\lambda(x) \tag{23}$$

$$\equiv \int_X d\gamma(x) \exp\left(-\frac{\lambda}{2} \int_T x\cdot R\cdot x\, dt \right). \tag{24}$$

Clearly

$$D(0) = 1 \quad \text{and} \quad D(\lambda = 1) = I. \tag{25}$$

We can now derive a differential equation for D:

$$\frac{dD(\lambda)}{d\lambda} = \frac{d}{d\lambda} \int_X d\gamma(x) \exp\left(-\frac{\lambda}{2} \int_T x\cdot R\cdot x\, dt \right) \tag{26}$$

$$= \int_X d\gamma(x) \left(-\frac{1}{2} \int_T x(t)\cdot R(t)\cdot x(t)\, dt \right) \times \exp\left(-\tfrac{1}{2}\lambda \int_T x\cdot R\cdot x\, dt \right). \tag{27}$$

This expression can be written in terms of $d\gamma_\lambda$, the generalized Wiener measure defined by the Green's function G_λ of the operator

$$\mathscr{J}_\lambda = \mathscr{J}_0 + (F(d/dt) + R):$$

$$\frac{dD(\lambda)}{d\lambda} = \int_T dt\, R_{\alpha\beta}(t) \int_X D(\lambda)\, d\bar{\gamma}_\lambda(x)\, x^\beta(t)\, x^\alpha(t)$$

$$= D(\lambda) \int_T dt\, R_{\alpha\beta}(t)\, G_\lambda(t,t), \tag{28}$$

or

$$\frac{1}{D(\lambda)} \frac{dD(\lambda)}{d\lambda} = \int_T dt\, \mathrm{tr}(R(t)\, G_\lambda(t,t)). \tag{29}$$

Thus we have the desired equation for D:

$$\frac{d \ln D(\lambda)}{d\lambda} \equiv \mathrm{Tr}(RG_\lambda). \tag{30}$$

(Here, "tr AB" denotes multiplying A and B together (or letting A operate on B if A is an operator) then taking the trace of the resulting matrix. The large trace Tr is a continuous trace

$$\mathrm{Tr}G \equiv \int_T \mathrm{tr}\, G(t,s)|_{t=s}\, dt. \tag{31}$$

We would now like to write G in terms of θ, J, N, and K to suggest a method for evaluating D. In Eq. (11) if we set $t = s$ and use the fact that G is continuous at $t = s$, we have that

$$G_\lambda(t,t) = -K_\lambda(t,1)\, N_\lambda(1,0)\, J_\lambda(0,t). \tag{32}$$

Then

$$\frac{d \ln D}{d\lambda} = -\tfrac{1}{2}\mathrm{tr} \int_0^1 dt\, (-R(t)\, K_\lambda(t,1)\, N_\lambda(1,0)\, J_\lambda(0,t))$$

$$= \tfrac{1}{2}\, \mathrm{tr} \int_0^1 J_\lambda(0,t) R(t) K_\lambda(t,1) N_\lambda(1,0). \tag{33}$$

We have derived an interesting expression for the Fredholm determinant. We would like now to show that this expression is equal to the ln of the ratio of two determinants of two matrices, one assocaited with the operator $\mathscr{J}_{\lambda=1}$ and the other with $\mathscr{J}_{\lambda=0}$. We consider $\ln(C)$ where C is defined to be

$$C(\lambda) \equiv \frac{\det K_\lambda(0,1)}{\det K_0(0,1)}. \tag{34}$$

We now want to show that

$$\frac{d \ln C(\lambda)}{d\lambda} = -\,\mathrm{tr} \int_0^1 J_\lambda(0,t)\, R(t)\, K_\lambda(t,1)\, N_\lambda(1,0) \tag{35}$$

$$= -2 \frac{d}{d\lambda} \ln D, \tag{36}$$

with $C(0) = 1$, so that

$$\ln C = -2 \ln D \tag{37}$$

or

$$D = C^{-1/2}, \tag{38}$$

i.e.,

$$I = D = \left(\frac{\det K_1(0,1)}{\det K_0(0,1)} \right)^{-1/2} \tag{39}$$

as desired. We will thus have evaluated the Wiener integral I, or equivalently the Fredholm determinant D, in terms of determinants of finite-dimensional matrices.

First, then

$$\frac{d\ln C}{d\lambda}=\frac{1}{|K_\lambda|}\frac{d|K_\lambda|}{d\lambda} \tag{40}$$

$$=\mathrm{tr}(N_\lambda\cdot K'_\lambda) \tag{41}$$

$$(\text{where } K'\equiv dK/d\lambda \quad \text{and} \quad |K|\equiv \det K). \tag{42}$$

To evaluate this, we need an expression for K'. For this, take $d/d\lambda$ of the equation $\mathscr{J}_\lambda K_\lambda = 0$ for K_λ. We find that

$$\mathscr{J}_\lambda(t)\,K'_\lambda(t,0) = -R(t)\cdot K_\lambda(t,0)\,. \tag{43}$$

Thus K'_λ is an inhomogeneous solution of the operator $\mathscr{J}_\lambda$. As such, it can be solved for by the method of Green's functions. Thus

$$K'_\lambda(t,1) = -\int_0^1 G_\lambda^{\mathrm{adv}}(t,s)\,R(s)\,K_\lambda(s,1)\,ds, \tag{44}$$

where G is a Green's function of $\mathscr{J}_\lambda$ with boundary conditions appropriate to K' at $t=1$. An explicit representation of G_λ^{adv} in terms of θ functions and homogeneous solutions of $\mathscr{J}$ is

$$G_\lambda^{\mathrm{adv}}(t,s) = \theta(s-t)\,J_\lambda(t,s)\,. \tag{45}$$

Thus

$$K'_\lambda(0,1) = -\int_0^1 J_\lambda(0,s)\,R(s)\,K_\lambda(s,1)\,ds\,. \tag{46}$$

Putting this expression for K'_λ into Eq. (44) we find that

$$\frac{d\ln C}{d\lambda}$$

$$=\mathrm{tr}\Big(N_\lambda(0,1)\int_0^1(-1)\,J_\lambda(0,s)R(s)K_\lambda(s,1)\Big)\,ds \tag{47}$$

$$=-\,\mathrm{tr}\int_0^1 J_\lambda(0,s)\,R(s)\,K_\lambda(s,1)\,N_\lambda(1,0)\,ds\,. \tag{48}$$

Comparison with (33) gives the desired result:

$$\frac{d\ln C(\lambda)}{d\lambda} = -2\frac{d}{d\lambda}\ln D(\lambda)\,. \tag{49}$$

The values of $\ln C$ and $\ln D$ at $\lambda=0$ are both zero. Thus

$$D(\lambda)=(C(\lambda))^{-1/2} \tag{50}$$

or

$$I=D(1)=\left|\frac{\det K_1(0,1)}{\det K_0(0,1)}\right|^{-1/2}. \tag{51}$$

As noted earlier, we relegate to the Appendix a discussion of the modification of the above proof to include the case when $F\neq 0$.

THE CASE OF PATH INTEGRALS (REF. 10)

The techniques developed here are directly applicable to the theory of path integrals and their imaginary Gaussian prodistributions. (See Ref. 6, pp. 259–82, for a discussion of prodistributions.) The Gaussian integrands arise when one does a semiclassical expansion of path integral, say for the propagator or for a wave function with a given initial value. (It is when doing the case of the wave function that the boundary terms in the integrand become important. They are usually neglected because one is interested in the propagator.) Note that by defining the new measure $\bar{\gamma}$ one is able not only to calculate the semiclassical approximation easily, but also to have the "semiclassical measure" to compute the higher order terms in the semiclassical expansion. For the path integral, $\mathscr{J}$ will be the Jacobi operator and the fields J and K are Jacobi fields. The determinant of K can be interpreted geometrically in terms of volume expansions and contractions of a congruence of classical flows defined in terms of a given initial (canonical) momentum. (See Ref. 11 for details.)

CONCLUSION

We have given a method for evaluating a large class of Fredholm determinants that arise when evaluating Wiener integrals of Gaussian integrands. Our method offers the simplicity of not resorting to any kind of discretization or expansion procedures to define the determinant; and it is manifestly covariant, making it directly applicable to cases when the paths are more than one dimensional. Finally our method is able to handle cases when there is a stochastic term of the form $\int x\cdot F\cdot dx$ in the integrand. These cases are cumbersome, and usually not treated, in other approaches to the Wiener integral. The power of our method derives from a combination of the linear change of variables introduced by Cameron and Martin (and extended by DeWitt-Morette), the powerful insight gained by viewing the Wiener measures in terms of their Fourier transforms and the associated Green's functions, and the idea of Schwinger's of looking not at one integrand but at a parametric family of integrands.

APPENDIX

We here show what modifications of the proof relating the Fredholm determinant D to the finite-dimensional determinant $|K|$ are needed for the case when $A\neq 0$. There are now two added complications, the boundary term $x\cdot\tilde{A}\cdot x$ and the stochastic term $\int x\cdot F\cdot\dot{x}$. [The two are not unrelated; in fact, the operator

$$\mathscr{J}(t)\equiv -\frac{d^2}{dt^2}+F\frac{d}{dt}+R \tag{A1}$$

is not self-adjoint on the space of paths that vanish only at $t=0$, unless the proper boundary terms are included. Equivalently, the quadratic form Q defined by

$$Q[x,y]\equiv x(1)\cdot\dot{y}(1)+x(1)\cdot\tilde{A}(1)\cdot y(1) + \int_T x(t)\cdot\mathscr{J}(t)\cdot y(t)\,dt \tag{A2}$$

is not symmetric without the boundary terms. (The first term is effectively contained in the usual Wiener measure γ that we begin with.)]

So, we want to prove that

$$D(\lambda)\equiv\int_X d\gamma(x)\,\exp(-\tfrac{1}{2}\lambda H(x,x)) = |\det K_\lambda(1,0)|^{-1/2}, \tag{A3}$$

where now H is given by

$$H(x,x) = x(1)\cdot A(1)\cdot x(1) + \int_0^1 x(t)\cdot\left(F(t)\frac{d}{dt} + R(t)\right)\cdot x(t)\,dt, \qquad \text{(A4)}$$

and K is defined to be the solution to

$$\mathscr{J}_\lambda(t)K_\lambda(t,0) = 0 \qquad \text{(A5a)}$$

[where $\mathscr{J}_\lambda \equiv -(d/dt^2) + \lambda(F(d/dt) + R)$] with the boundary condition (A6)

$$K_\lambda(0,0) = g^{-1} \qquad \text{(A5b)}$$

(where g is the metric tensor, just δ in Cartesian coordinates) and

$$\frac{dK_\lambda}{dt}(t=0,0) = -\lambda\tilde{A}(t). \qquad \text{(A5c)}$$

Again we are considering a parametric family of integrals. Note that λ enters into the definition of $\mathscr{J}_\lambda$ and into the boundary conditions of K_λ. Following the same analysis that led to Eq. (28) it is obvious that the differential equation for D is now

$$\frac{d}{d\lambda}\ln D(\lambda) = -\tfrac{1}{2}\mathrm{tr}\tilde{A}(1)\,G_\lambda(1,1) - \frac{1}{2}\int_0^1 dt\,\mathrm{tr}\left(F(t)\frac{d}{dt}G_\lambda(t,s)|_{s=t} + R(t)\,G_\lambda(t,t)\right), \qquad \text{(A6)}$$

where we have used the fact that

$$\int dt \int_X d\gamma(x)\,x(t)\,\dot{x}(t) = \int dt\frac{d}{dt}G(t,s)|_{s=t}. \qquad \text{(A7)}$$

We now again replace $G(t,t)$ by the $-K\cdot N\cdot J$ term. We also replace

$$\left.\frac{dG(t,s)}{dt}\right|_{t=s}$$

by

$$-\left(\frac{dK(t,1)}{dt}\right)N(1,0)\,J(0,t) - \tfrac{1}{2}g^{-1}, \qquad \text{(A8)}$$

and note that because F is skew, the g^{-1} does not contribute to $\mathrm{tr}(F\dot{G})$. Thus the desired equation for $D(\lambda)$ is

$$\frac{d}{d\lambda}(\ln D(\lambda)) = \tfrac{1}{2}\mathrm{tr}\tilde{A}(1)\,K_\lambda(1,1)\,N_\lambda(1,0)\,J_\lambda(0,1) + \frac{1}{2}\int_0^1 dt\,\mathrm{tr}(F(t)\,\dot{K}_\lambda(t,1) + R(t)\,K_\lambda(t,1)) \times N_\lambda(1,0)\,J_\lambda(0,t). \qquad \text{(A9)}$$

We next consider $C(\lambda)$ as defined by (34). Again

$$\frac{d\ln C(\lambda)}{d\lambda} = \mathrm{tr}(N_\lambda(1,0)\cdot K'_\lambda(0,1)). \qquad \text{(A10)}$$

The expression for K' is now

$$K'_\lambda = -\int_0^1 \theta(s-t)\,J_\lambda(t,s)\left(F(s)\frac{d}{ds} + R(s)\right)K_\lambda(s,1)\,ds - J_\lambda(t,1)\,\tilde{A}(1). \qquad \text{(A11)}$$

[The boundary term is needed because the boundary value of $(d/dt)\,K'_\lambda(t,1)$ at $t=1$ is no longer zero.]

Putting this expression for $K'_\lambda(0,1)$ into the equation for $C(\lambda)$ (A10) gives that

$$\frac{d\ln C(\lambda)}{d\lambda} = -\int_0^1 dt\,\mathrm{tr}\left[\left(F(t)\frac{d}{dt} + R(t)\right)\times K_\lambda(t,1)\,N_\lambda(1,0)\,J_\lambda(0,t)\right] - \mathrm{tr}(A(1)\,K_\lambda(1,1)\,N_\lambda(1,0)\,J_\lambda(0,1)) = -2\frac{d}{d\lambda}\ln D(\lambda), \qquad \text{(A12)}$$

where use has been made of the cyclic property of the trace and the fact that N is the inverse of K. Thus, once again we have the relationship between the Fredholm determinant D and that of K:

$$D(\lambda) = |\det K_\lambda(0,1)|^{-1/2}. \qquad \text{(A13)}$$

(We have used the fact that in Cartesian coordinates

$$K_0^{\alpha\beta}(t,1) = \delta^{\alpha\beta}.$$

So $\det K_0 = 1$.)

[1] R. H. Cameron and W. T. Martin, "Transformation of Wiener Integrals Under a General Class of Transformations," Trans. Am. Math. Soc. **58**, 184–219 (1945).

[2] J. Schwinger, for example "On Gauge Invariance and Vacuum Polarization," Phys. Rev. **82**, 664–79 (1951).

[3] C. DeWitt-Morette, "Feynman Path Integrals, Definition Without Limiting Procedure," Commun. Math. Phys. **28**, 47–67 (1972).

[4] C. DeWitt-Morette, "Feynman Path Integrals: I. Linear and Affine Techniqes; II. The Feynman–Green Function," Commun. Math. Phys. **37**, 63–81 (1974).

[5] C. DeWitt-Morette, "The Semi-classical Expansion," Ann. Phys. **97**, 367–99 (1976).

[6] C. DeWitt, A Maheshwari, and B. Nelson, "Path Integration in Non-relativistic Quantum Mechanics," Phys. Rep. **50**, 255–372 (1979).

[7] R. H. Cameron and W. T. Martin, "Evaluation of Certain Wiener Integrals by use of Certain Sturm–Liouville Differential Equations," Bull. Am. Math. Soc. **51**, 73–90 (1945).

[8] R. P. Feynman, "Spacetime Approach to Non-relativistic Quantum Mechanics," Rev. Mod. Phys. **20**, 367–87 (1948).

[9] S. Levit and V. Smilansky, "A Theorem on Infinite Products of Eigenvalues of Sturm–Liouville Type Operators" (1976), preprint from Weizmann Institute of Science, Rehovot, Israel.

[10] B. Sheeks, "Some Applications of Path Integration Using the Prodistribution Method," May 1980, Ph.D. Dissertation, University of Texas, Austin.

[11] B. Nelson and B. Sheeks, "Path Integration for Velocity-dependent Potentials," Center of Relativity Preprint, 1979, University of Texas, Austin.

Modified singular perturbation method for a stiff system of linear evolution equations

Tomasz Bleński
Computing Center CYFRONET, Institute of Nuclear Research, Swierk, 05-400 Otwock, Poland

(Received 10 October 1979; accepted for publication 15 May 1980)

A stiff system of linear evolution equations in Banach space is investigated. It is shown that the standard singular perturbation algorithm can be considerably simplified. In this modified algorithm the asymptotic solutions can be obtained in each order independently. The initial conditions are given explicitly so there is no need to solve the so-called "inner" equations. Two examples of application are considered.

PACS numbers: 02.30.Jr

A small parameter multiplying the time derivative term may appear in a system of evolution equations when the time constants characterizing the system differ from each other by orders of magnitude. Such a system of evolution equations may be called stiff by analogy to the case of ordinary differential equations.[1] For stiff systems the standard perturbation approach breaks down for t close to the origin (the initial moment). Thus the standard expansion should be for small t supplemented by an inner solution. That procedure leads to the singular perturbation method.[2-5]

The aim of this paper is to present a new algorithm based on the singular perturbation method developed by Mika[5] for the following system of linear equations:

$$\epsilon \frac{d}{dt} x(t) = A(t)x(t) + P(t)y(t) + q(t), \tag{1a}$$

$$\frac{d}{dt} y(t) = Q(t)x(t) + B(t)y(t) + r(t), \tag{1b}$$

with the initial conditions

$$x(0) = \mu, \quad y(0) = \eta.$$

As in Ref. 5 the above equation will be considered as evolution equations in a Banach space $\mathscr{X}$ with the norm $\| \ \|$:

For $t \in [0,t_0]$, $q(t), r(t) \in \mathscr{X}$, $P(t)$, $Q(t)$, and $B(t)$ are bounded linear operators in $\mathscr{X}$. The linear operator $A(t)$ with the domain D_A dense in $\mathscr{X}$ is of the form

$$z \in D_A, \quad A(t)z = A_0 z + A_1(t)z,$$

where $A_1(t)$ is bounded and the closed operator A_0 is a generator of a strongly continuous semigroup of bounded evolution operators $G(t)$[6] such that

$$\|G(t)\| \leqslant \exp(\gamma t), \quad \text{for some real } \gamma,$$

where $\| \ \|$ denotes the operator norm in $\mathscr{X}$; moreover for each $\tau \in [0,t_0]$ the resolvent operator $R(\lambda, A(\tau)) = (\lambda I - A(\tau))^{-1}$ satisfies the inequality

$$\|R(\lambda, A(\tau))\| \leqslant (\lambda - \gamma_\tau)^{-1}, \gamma_\tau < \lambda, \tag{2}$$

and $\sup_{\tau \in [0,t_0]} \gamma_\tau = \bar{\gamma} < 0$, $\mu \in D_A$, $\eta \in \mathscr{X}$. The positive number ϵ plays the role of the small parameter.

1. THE SINGULAR PERTURBATION METHOD

We sketch briefly the results of Ref. 5, where the singular perturbation method is applied to the system (1a), (1b).

It is assumed that for $t \in [0,t_0]$ the operator functions $A_1(t)$, $P(t)$, $Q(t)$, and $B(t)$ are $n+1$ times continuously differentiable in the operator norm and the functions $q(t)$ and $r(t)$ are $(n+1)$ times strongly continuously differentiable.

Then the unique, strongly differentiable solutions $x(t)$, $y(t)$ to the system (1a), (1b) have asymptotic expansions $x^{(n)}(t)$, $y^{(n)}(t)$ such that for each $t \in [0,t_0]$

$$\|x(t) - x^{(n)}(t)\| = O(\epsilon^{n+1}),$$
$$\|y(t) - y^{(n)}(t)\| = O(\epsilon^{n+1}). \tag{3}$$

The asymptotic solutions of the nth order $x^n(t)$, $y^n(t)$ are given by the sums

$$x^{(n)}(t) = \tilde{x}^{(n)}(t/\epsilon) + \bar{x}^{(n)}(t),$$
$$y^{(n)}(t) = \tilde{y}^{(n)}(t/\epsilon) + \bar{y}^{(n)}(t), \tag{4}$$

of the inner asymptotic solutions $\tilde{x}^{(n)}(t/\epsilon)$, $\tilde{y}^{(n)}(t/\epsilon)$ and the outer asymptotic solutions $\bar{x}^{(n)}(t)$, $\bar{y}^{(n)}(t)$.

The outer solution is obtained from the standard perturbation approach applied to (1)

$$\bar{x}^{(n)}(t) = \sum_{k=0}^{n} \epsilon^k \bar{x}_k(t),$$
$$\bar{y}^{(n)}(t) = \sum_{k=0}^{n} \epsilon^k \bar{y}_k(t), \tag{5}$$

where $\bar{x}_k(t)$, $\bar{y}_k(t)$, $k = 0,1,\dots,n$, are solutions to the outer equations

$$A(t)\bar{x}_k(t) + P(t)\bar{y}_k(t) + q_k(t) = 0, \tag{6a}$$

$$\frac{d}{dt}\bar{y}_k(t) = Q(t)\bar{x}_k(t) + B(t)\bar{y}_k(t) + \delta_{0k} r(t), \tag{6b}$$

with

$$q_k(t) = \begin{cases} q(t), & k = 0, \\ -\dfrac{d}{dt}\bar{x}_{k-1}(t), & k \neq 0. \end{cases}$$

The outer asymptotic solutions approximate the exact solutions to (1) for t large in comparison with ϵ.

The inner asymptotic solutions $\tilde{x}^{(n)}(\tau)$ and $\tilde{y}^{(n)}(\tau)$ are given by the series of

$$\tilde{x}^{(n)}(\tau) = \sum_{k=0}^{n} \epsilon^k \tilde{x}_k(\tau), \tag{7}$$

$$\bar{y}^{(n)}(\tau) = \sum_{k=0}^{n} \epsilon^k \tilde{y}_k(\tau)$$

which satisfy the equations

$$\frac{d}{d\tau}\tilde{x}_k(\tau) = \sum_{m=0}^{k} \frac{\tau^{k-m}}{(k-m)!} \times (A^{(k-m)}\tilde{x}_m(\tau) + P^{(k-m)}\tilde{y}_m(\tau)), \quad (8a)$$

$$\frac{d}{d\tau}\tilde{y}_k(\tau) = \sum_{m=0}^{k-1} \frac{\tau^{k-m-1}}{(k-m-1)!} \times (Q^{(k-m-1)}\tilde{x}_m(\tau) + B^{(k-m-1)}\tilde{y}_m(\tau)), \quad (8b)$$

where the notation

$$V^{(m)} \equiv \frac{d^m}{d\tau^m} V(t)|_{\tau=0}, m = 0,1,...,n$$

is used. Since the reason for introducing them is to match the exact solutions in the region where t is of the order of ϵ it is demanded that

$$\lim_{\tau\to\infty} \tilde{y}_k(\tau) = 0; \quad k = 0,1,...,n. \quad (9)$$

The above requirement and the structure of (6) and (8) are sufficient to specify the initial conditions for the inner and outer solutions. Let us assume $\mu = O(1)$, $\eta = O(1)$. Taking into account (1) and (4) we have

$$\begin{aligned} &\bar{x}_0(0) + \tilde{x}_0(0) = \mu, \\ &\bar{y}_0(0) + \tilde{y}_0(0) = \eta, \\ &\bar{x}_k(0) + \tilde{x}_k(0) = 0, \quad k = 1,...,n, \\ &\bar{y}_k(0) + \tilde{y}_k(0) = 0. \end{aligned} \quad (10)$$

From (8) and (9) it follows that

$$\tilde{y}_0(\tau) = 0, \quad \bar{y}_0(0) = \eta. \quad (11)$$

Now with the above initial condition Eqs. (6) can be solved for $k = 0$ yielding $\bar{x}_0(0)$ and, according to (10), $\tilde{x}_0(0)$. For $k > 0$ the procedure is similar: Eq. (8b) yields $\tilde{y}_k(0)$ in terms of $\tilde{y}_m(0)$ and $\tilde{x}_m(0)$, $m = 0,1,...,k-1$, and from (9) gives $\bar{y}_k(0)$. This allows us to solve (6) and, finally, to find $\tilde{x}_k(0)$ and solve (8a).

The analysis of the system (1) given in Ref. 5 was based upon the theory of evolution equations in Banach spaces.[2,6,7] For completeness and clarity we present here also the formulation, close to the original one, of the main results of Ref. 5 in the form of Assumption 1, resulting in Lemmas 2–4 and Theorem 5.

Let $\mathscr{X}$ be a complex Banach space with the norm $\|\ \|$ and $\mathscr{B}$ be the Banach space of bounded operators in $\mathscr{X}$ with the norm defined by the same symbol.

Assumption 1: The operator A_0 with domain D_A is an infinitesimal generator of a strongly continuous semigroup. The operator functions $A_1(t)$, $B(t)$, $P(t)$, and $Q(t)$ with values from $\mathscr{B}$ are $(n+1)$ times continuously differentiable on $[0,t_0]$ where $t_0 > 0$, in the sense of norm in $\mathscr{B}$. The quasisemigroups generated by the function $A(t)$ defined for $z \in D_A$ with the equation

$$A(t)z = A_0 z + A_1(t)z,$$

and by the function $B(t)$ are, respectively, $U_A(t,s)$ satisfying on the triangle

$$\Omega = \{(t,s): 0 \leqslant s \leqslant t \leqslant t_0\},$$

the inequality

$$\|U_A(t,s)\| \leqslant \exp(\bar{\gamma}_A(t-s)),$$

and $U_B(t,s)$ satisfying on Ω the inequality

$$\|U_B(t,s)\| \leqslant \exp(\bar{\gamma}_B(t-s)).$$

The constant $\bar{\gamma}_A$ is negative.

The function $(1/\epsilon)A(t)$ generates the quasisemigroup $U_A^{(\epsilon)}(t,s)$ satisfying on Ω the inequality

$$\|U_A^{(\epsilon)}(t,s)\| \leqslant \exp\left(\bar{\gamma}_A \frac{(t-s)}{\epsilon}\right).$$

The functions $q(t)$ and $r(t)$ with values from $\mathscr{X}$ are $(n+1)$ times strongly continuously differentiable on $[0,t_0]$.

Lemma 2: A singularly perturbed system of evolution equations (1) with the initial conditions $x(0) = \mu \in D_A$, $y(0) = \eta \in \mathscr{X}$ and all the functions satisfying Assumption 1, has for any $\epsilon > 0$ unique strongly differentiable solutions $\{x(t),y(t)\}$.

Lemma 3: The system of outer asymptotic equations (6) with all the functions satisfying Assumption 1 has unique strongly differentiable solutions $\{\bar{x}_0(t),...,\bar{x}_n(t),\bar{y}_0(t),...,\bar{y}_n(t)\}$ for any initial conditions such that

$$\bar{y}_k(0) = \bar{\eta}_k \in \mathscr{X}, \quad k = 0,1,...,n.$$

Lemma 4: Let the operator functions $A(t)$, $B(t)$, $P(t)$, and $Q(t)$ satisfy Assumption 1. The system of inner asymptotic equations (8) has unique strongly differentiable solutions

$$\{\tilde{x}_0(t/\epsilon),...,\tilde{x}_n(t/\epsilon),\tilde{y}_0(t/\epsilon),...,\tilde{y}_n(t/\epsilon)\}$$

for any initial conditions put on $\tilde{x}_k(t)$ such that

$$\tilde{x}_k(0) = \tilde{\mu}_k \in D_A, \quad k = 0,1,...,n,$$

and with the requirement (9).

The functions $\tilde{x}_k(t)$ and $\tilde{y}_k(t)$ satisfy the inequalities

$$\|\tilde{x}_k(t/\epsilon)\| \leqslant \exp(\gamma_0 t/\epsilon) \sum_{j=0}^{2k} M_j^{(k)}(t/\epsilon)^j,$$

$$\|\tilde{y}_k(t/\epsilon)\| \leqslant \exp(\gamma_0 t/\epsilon) \sum_{j=0}^{2(k-1)} N_j^{(k)}(t/\epsilon)^j,$$

$$k = 0,1,...,n,$$

where $M_j^{(k)}$ and $N_j^{(k)}$ are some positive constants. A negative parameter γ_0 is defined by the semigroup $G_0(t/\epsilon)$ generated by $A^{(0)}$ and satisfying for $t \in [0,t_0]$ the inequality

$$\|G_0(t/\epsilon)\| \leqslant \exp(\gamma_0 t/\epsilon).$$

Theorem 5: The asymptotic solutions of order n, $\{x^{(n)}(t),y^{(n)}(t)\}$ defined in (4), (5), and (7), tend in the norm to the exact solutions $\{x(t),y(t)\}$ of the singularly perturbed system of equations (1) uniformly on $[0,t_0]$ faster than ϵ^n. In other words, for each $\delta > 0$ there exists $\epsilon > 0$ such that for all $t \in [0,t_0]$:

$$\epsilon^{-n}\|x(t) - x^{(n)}(t)\| < \delta,$$

$$\epsilon^{-n}\|y(t) - y^{(n)}(t)\| < \delta.$$

The described singular perturbation approach to (1) has two features inconvenient from the practical point of view. First, both the outer and inner equations are to be solved successively since they contain the lower order solutions in the source terms. Second, due to the coupling through the initial conditions (10) in each order the inner equation (8b) should be solved before the system of the outer equations (6).

These properties usually limit the practical applications of the singular perturbation approach to (1) to cases when ϵ is small enough to ensure the sufficient accuracy of the low order approximations. Then, since the inner solutions are negligible except when $t = O(\epsilon)$ their role reduces in practice to providing the initial conditions for the outer equations. Thus it would be desirable to get the initial conditions for the outer solutions without solving the inner equations. We show how it is possible and give explicit formula for the low order approximations in the following section.

Further we propose how to avoid the successive solving of the outer equations. The presented algorithm can, in principle, be applied to any order. However, as we shall see in the example of the lengthy terms in the second order, the complexity grows rapidly with the order of approximation. For this reason the second order is the last included in this paper.

The modified outer equations obtained here together with the initial conditions can be solved in zero, first, or second order independently and yield the approximate solutions correct in the desired order of ϵ for asymptotic ($t > \epsilon$) times.

2. THE INITIAL CONDITIONS FOR THE OUTER EQUATIONS IN THE LOW ORDER APPROXIMATION

Let us try to calculate the first three initial conditions $\bar{y}_k(0)$, $k = 0,1,2$, needed for the outer equations (6). Fortunately, as we shall see, one can obtain the explicit expressions for $\bar{y}_k(0)$ although in the general case it is not possible to get such expressions for the inner solutions for an arbitrary $t \in [0,t_0]$.

$$\bar{x}_0(0) = -(A^{(0)})^{-1}(P^{(0)}\eta + q(0)).$$

It follows from the inequality (2) that $(A^{(0)})^{-1}$ exists and is bounded, and moreover[6] Eq. (8a) for $k = 0$:

$$\epsilon \frac{d}{dt}\tilde{x}_0(t/\epsilon) = A^{(0)}\tilde{x}_0(t/\epsilon),$$

(where the substitution $t = \epsilon\tau$ is made), with

$$\tilde{x}_0(0) = \mu - \bar{x}_0(0),$$

has the solution

$$\tilde{x}_0(t/\epsilon) = G_0(t/\epsilon)\tilde{x}_0(0), \tag{12}$$

where the semigroup $G_0(\tau)$ fulfills the inequality

$$\|G_0(\tau)\| \leqslant \exp(\gamma_0\tau), \quad \gamma_0 < 0.$$

In the first order (8b) gives for $\tilde{y}_1(t/\epsilon)$

$$\epsilon \frac{d}{dt}\tilde{y}_1(t/\epsilon) = Q^{(0)}\tilde{x}_0(t/\epsilon).$$

The requirement (9) immediately gives

$$\tilde{y}_1(t/\epsilon) = -(1/\epsilon)\int_t^\infty dt' Q^{(0)}\tilde{x}_0(t'/\epsilon),$$

which yields after inserting (12)

$$\tilde{y}_1(t/\epsilon) = Q^{(0)}(A^{(0)})^{-1}G_0(t/\epsilon)\tilde{x}_0(0), \tag{13}$$

and from (10)

$$\bar{y}_1(0) = -Q^{(0)}(A^{(0)})^{-1}\tilde{x}_0(0). \tag{14}$$

To obtain $\bar{y}_2(0) = -\tilde{y}_2(0)$ one needs $\tilde{x}_1(t/\epsilon)$. Equation (8a) for $k = 1$,

$$\epsilon \frac{d}{dt}\tilde{x}_1(t/\epsilon) = A^{(0)}\tilde{x}_1(t/\epsilon) + \frac{t}{\epsilon}A^{(1)}\tilde{x}_0(t/\epsilon) + P^{(1)}\tilde{y}_1(t/\epsilon),$$

has the solution

$$\begin{aligned}\tilde{x}_1(t/\epsilon) = {} & G_0(t/\epsilon)\tilde{x}_1(0) \\ & + \frac{1}{\epsilon^2}\int_0^t dt'\, t' G_0((t-t')/\epsilon)A^{(1)}G_0(t'/\epsilon)\,\tilde{x}_0(0) \\ & + \frac{1}{\epsilon}\int_0^t dt'\, G_0((t-t')/\epsilon)P^{(1)}Q^{(0)}\,(A^{(0)})^{-1} \\ & \times G_0(t'/\epsilon)\tilde{x}_0(0).\end{aligned} \tag{15}$$

Equation (8b) for $k = 2$ and (9) gives

$$\tilde{y}_2(t/\epsilon) = -\frac{1}{\epsilon}\int_t^\infty dt'\,\{(t'/\epsilon)Q^{(1)}\,\tilde{x}_0(t'/\epsilon) + Q^{(0)}\,\tilde{x}_1(t'/\epsilon) + B^{(0)}\,\tilde{y}_1(t'/\epsilon)\},$$

and after inserting (12), (13), and (15)

$$\begin{aligned}\tilde{y}_2(t/\epsilon) = {} & -\frac{1}{\epsilon}\left[\frac{1}{\epsilon}Q^{(1)} + B^{(0)}Q^{(0)}\,(A^{(0)})^{-1}\right] \\ & \times \int_t^\infty dt'\, G_0(t'/\epsilon)\tilde{x}_0(0) \\ & - \frac{1}{\epsilon}Q^{(0)}\int_t^\infty dt'\, G_0(t'/\epsilon)\tilde{x}_1(0) - \frac{1}{\epsilon^3}Q^{(0)} \\ & \times \int_t^\infty dt' \int_0^{t'} dt''\, t''\, G_0((t'-t'')/\epsilon)A^{(1)} \\ & \times G_0(t''/\epsilon)\,\tilde{x}_0(0) \\ & - \frac{1}{\epsilon^2}Q^{(0)}\int_t^\infty dt' \int_0^{t'} dt''\, G_0((t'-t'')/\epsilon) \\ & \times P^{(1)}Q^{(0)}\,(A^{(0)})^{-1}G_0(t''/\epsilon)\tilde{x}_0(0).\end{aligned} \tag{16}$$

For $t = 0$ the integrations in the double integrals can be performed analytically. To see this let us put $t = 0$ and change the order of integration in the last two terms in (16), which yields

$$\begin{aligned}& -\frac{1}{\epsilon^3}Q^{(0)}\int_0^\infty dt''\, t'' \int_{t''}^\infty dt'\, G_0((t'-t'')/\epsilon)A^{(1)}G_0(t''/\epsilon) \\ & \times \tilde{x}_0(0) - \frac{1}{\epsilon^2}Q^{(0)}\int_0^\infty dt'' \int_{t''}^\infty dt'\, G_0((t'-t'')/\epsilon) \\ & \times P^{(1)}Q^{(0)}\,(A^{(0)})^{-1}G_0(t''/\epsilon)\tilde{x}_0(0).\end{aligned}$$

Now using the simple formulas

$$\int_0^\infty dt'\, G_0(t'/\epsilon) = -\epsilon\,(A^{(0)})^{-1},$$

$$\int_0^\infty dt'\, t' G_0(t'/\epsilon) = \epsilon^2(A^{(0)})^{-2},$$

we obtain from (16):

$$\begin{aligned}\bar{y}_2(0) = {} & [B^{(0)}Q^{(0)} - Q^{(1)} + Q^{(0)}(A^{(0)})^{-1} \\ & \times (A^{(1)} - P^{(1)}Q^{(0)})] \\ & \times (A^{(0)})^{-2}\tilde{x}_0(0) + Q^{(0)}\,(A^{(0)})^{-1}\tilde{x}_1(0).\end{aligned} \tag{17}$$

We get the initial value $\bar{x}_1(0) = -\bar{x}_1(0)$ from (6a) for $k=1$ using the expression (14) for $\bar{y}_1(0)$ and calculating $(d/dt)\bar{x}_0(0)$ from (6a) for $k=0$.

In such a way the initial conditions for $\bar{y}_0(t)$, $\bar{y}_1(t)$, and $\bar{y}_2(t)$ are expressed explicitly with μ and η and with the known functions and their derivatives taken at $t=0$.

3. THE OUTER ASYMPTOTIC EQUATIONS FOR $\bar{y}^{(k)}(t)$

Let us eliminate $\bar{x}_k(t)$ in (6b) using (6a). To calculate the derivatives $(d/dt)\bar{x}_m(t)$, $m=0,1$ in the source we differentiate the same expression for $x_m(t)$ from (6a) substituting for $(d/dt)\bar{y}_m(t)$ the right-hand side of (6a) with $\bar{x}_m(t)$ in terms of $\bar{y}_m(t)$. Then we multiply the resulting equations for $\bar{y}_k(t)$ by ϵ^k and sum up getting the following equations for $\bar{y}^{(k)}(t)$ as defined in (5), $k=0,1,2$:

$$\frac{d}{dt}\bar{y}^{(0)}(t) = C^{(0)}(t)\bar{y}^{(0)}(t) + s^{(0)}(t), \tag{18a}$$

$$\frac{d}{dt}\bar{y}^{(1)}(t) = C^{(0)}(t)\bar{y}^{(1)}(t) + \epsilon\delta C^{(1)}(t)\bar{y}^{(0)}(t) + s^{(1)}, \tag{18b}$$

$$\frac{d}{dt}\bar{y}^{(2)}(t) = C^{(0)}(t)\bar{y}^{(2)}(t) + \epsilon\delta C^{(1)}(t)\bar{y}^{(1)}(t) + \epsilon^2\delta C^{(2)}(t)\bar{y}^{(0)}(t) + s^{(2)}(t), \tag{18c}$$

where

$$C^{(0)}(t) = -Q(t)A(t)^{-1}P(t) + B(t),$$
$$s^{(0)}(t) = -Q(t)A(t)^{-1}q(t) + r(t),$$
$$\delta C^{(1)}(t) = -Q(t)A(t)^{-1}H(t),$$
$$H(t) = \frac{d}{dt}(A(t)^{-1}P(t)) + A(t)^{-1}P(t)C^{(0)}(t),$$
$$s^{(1)}(t) = s^{(0)}(t) - \epsilon Q(t)A(t)^{-1}v^{(0)}(t),$$
$$\delta C^{(2)}(t) = -Q(t)A(t)^{-1}\left[\frac{d}{dt}(A(t)^{-1}H(t)) + A(t)^{-1} \times P(t)\delta C^{(1)}(t) + A(t)^{-1}H(t)C^{(0)}(t)\right],$$
$$s^{(2)}(t) = s^{(0)}(t) + \epsilon Q(t)A(t)^{-1}\left[v^{(1)}(t) + \epsilon A(t)^{-1} \times H(t)s^{(0)}(t) + \epsilon\frac{d}{dt}(A(t)^{-1}v^{(0)}(t))\right],$$
$$v^{(i)}(t) = \frac{d}{dt}(A(t)^{-1}q(t)) + A(t)^{-1}P(t)s^{(i)}(t), \qquad i=0,1.$$

The corresponding expressions for $\bar{x}^{(k)}(t)$, $k=0,1,2$, are

$$\bar{x}^{(0)}(t) = -A(t)^{-1}(P(t)\bar{y}^{(0)}(t) + q(t)), \tag{19a}$$

$$\bar{x}^{(1)}(t) = -A(t)^{-1}(P(t)\bar{y}^{(1)}(t) + \epsilon H(t)\bar{y}^{(0)}(t) + q(t) + \epsilon v^{(0)}(t)), \tag{19b}$$

$$\bar{x}^{(2)}(t) = -A(t)^{-1}\Bigg\{P(t)\bar{y}^{(2)}(t) + \epsilon H(t)\bar{y}^{(1)}(t) + \epsilon^2\left[\frac{d}{dt}(A(t)^{-1}H(t)) + A(t)^{-1} \times (P(t)\delta C^{(1)}(t) + H(t)C^{(0)}(t))\right]\bar{y}^{(0)}(t) + q(t) + \epsilon v^{(1)}(t) + \epsilon^2 \times \left[\frac{d}{dt}(A(t)^{-1}v^{(0)}(t)) + H(t)s^{(0)}(t)\right]\Bigg\},$$

Applying the similar procedure of multiplying by ϵ^k and summing up to (11), (14), and (17) we get the initial conditions to supplement Eqs. (18).

$$\bar{y}^{(0)}(0) = \eta,$$
$$\bar{y}^{(1)}(0) = \eta - \epsilon Q^{(0)}(A^{(0)})^{-1}[\mu + (A^{(0)})^{-1}(P^{(0)}\eta + q(0))],$$
$$\bar{y}^{(2)}(0) = \eta - \epsilon Q^{(0)}(A^{(0)})^{-1}[\mu + (A^{(0)})^{-1} \times (P^{(0)}\bar{y}^{(1)}(0) + \epsilon H(0)\eta] + q(0) + \epsilon v^{(0)}(0)] + \epsilon^2[B^{(0)}Q^{(0)} - Q^{(1)} + Q^{(0)}(A^{(0)})^{-1}(A^{(1)} - P^{(1)}Q^{(0)})](A^{(0)})^{-2} \times [\mu + (A^{(0)})^{-1}(P^{(0)}\eta + q(0))]. \tag{20}$$

4. THE MODIFIED OUTER EQUATIONS

We now show that if in the right-hand side of (18b) $\bar{y}^{(0)}(t)$ is substituted by $\bar{y}^{(1)}(t)$ then the additional error introduced by such modification is of the order of $O(\epsilon^2)$.

Let us introduce the modified outer equation for $\bar{y}^{(k)}_{\rm mod}(t)$, $k=0,1,2$,

$$\frac{d}{dt}\bar{y}^{(k)}_{\rm mod}(t) = C^{(k)}(t)\bar{y}^{(k)}_{\rm mod}(t) + s^{(k)}(t), \tag{21}$$

where

$$C^{(1)}(t) = C^{(0)}(t) + \epsilon\delta C^{(1)}(t),$$
$$C^{(2)}(t) = C^{(1)}(t) + \epsilon^2\delta C^{(2)}(t).$$

Let the modified solutions to (21) fulfill the same initial conditions as

$$\bar{y}^{(k)}_{\rm mod}(0) = \bar{y}^{(k)}(0), \quad k=0,1,2. \tag{22}$$

Since $C^{(0)}(t)$ is a bounded operator function strongly continuously differentiable on $[0,t_0]$ it generates for t, t' such that $t_0 \geqslant t \geqslant t' \geqslant 0$ a strongly continuous evolution operator $V^{(0)}(t,t')$ such that[2]

$$\|V^{(0)}(t,t')\| \leqslant \exp(\bar{\gamma}(t-t')), \quad \bar{\gamma}\in\mathbb{R}^1. \tag{23}$$

The solution to the outer equation (18b) is

$$\bar{y}^{(1)}(t) = V^{(0)}(t,0)\bar{y}^{(1)}(0) + \int_0^t dt'\, V^{(0)}(t,t')(\epsilon\delta C^{(1)}(t')y^{(0)}(t') + s^{(1)}(t')). \tag{24}$$

On the other hand $\bar{y}^{(1)}_{\rm mod}(t)$ satisfies the Volterra equation

$$\bar{y}^{(1)}_{\rm mod}(t) = V^{(0)}(t,0)\bar{y}^{(1)}(0) + \int_0^t dt'\, V^{(0)}(t,t')(\epsilon\delta C^{(1)}(t')\bar{y}^{(1)}_{\rm mod}(t') + s^{(1)}(t')). \tag{25}$$

Subtracting (24) from (25) we get the Volterra equation for $\delta\bar{y}^{(1)}(t) \equiv \bar{y}^{(1)}_{\rm mod}(t) - \bar{y}^{(1)}(t)$:

$$\delta\bar{y}^{(1)}(t) = \int_0^t dt'\, V^{(0)}(t,t')\epsilon\delta C^{(1)}(t')\delta\bar{y}^{(1)}(t') + m(t), \tag{26}$$

where

$$m(t) = \int_0^t dt'\, V^{(0)}(t,t')\epsilon\delta C^{(1)}(t')(\bar{y}^{(1)}(t') - \bar{y}^{(0)}(t')).$$

We have

$$\|\bar{y}^{(1)}(t) - \bar{y}^{(0)}(t)\| = O(\epsilon), \quad t\in[0,t_0],$$

and since $\delta C^{(1)}(t)$ and $V^{(0)}(t,t')$ are bounded for $t_0 \geqslant t \geqslant t' \geqslant 0$:

$$\|m(t)\| = O(\epsilon^2); \quad t\in[0,t_0].$$

Solving (26) by iteration immediately gives the estimation

$$\|\delta\bar{y}^{(1)}(t)\| = O(\epsilon^2), \quad t\in[0,t_0].$$

The similar consideration gives for $k = 2$

$$\|\bar{y}^{(2)}_{\rm mod}(t) - \bar{y}^{(2)}(t)\| = O(\epsilon^3), \quad t\in[0,t_0].$$

We may now substitute $\bar{y}^{(k)}_{\rm mod}(t)$ in (19b) and (19c) for the lower order $\bar{y}^{(m)}(t)$, $m = 0,\dots,k-1$; and define $\bar{x}^{(k)}_{\rm mod}(t)$:

$$\begin{aligned}
\bar{x}^{(0)}_{\rm mod}(t) &= \bar{x}^{(0)}(t) = -A(t)^{-1}(P(t)\bar{y}^{(0)}_{\rm mod}(t) + q(t)),\\
\bar{x}^{(1)}_{\rm mod}(t) &= -A(t)^{-1}[(P(t) + \epsilon H(t))\bar{y}^{(1)}_{\rm mod}(t)\\
&\quad + q(t)\epsilon v^{(0)}(t)],\\
\bar{x}^{(2)}_{\rm mod}(t) &= -A(t)^{-1}\Bigg\{\Bigg[P(t) + \epsilon H(t)\\
&\quad + \epsilon^2\frac{d}{dt}(A(t)^{-1}H(t)) + \epsilon^2 A(t)^{-1}(P(t)\delta C^{(1)}(t)\\
&\quad + H(t)C^{(0)}(t))\Bigg]\bar{y}^{(2)}_{\rm mod}(t) + q(t) + \epsilon v^{(1)}(t)\\
&\quad + \epsilon^2\left[\frac{d}{dt}(A(t)^{-1}v^{(0)}(t)) + H(t)s^{(0)}(t)\right]\Bigg\}.
\end{aligned} \tag{27}$$

It follows immediately that

$$\|\bar{x}^{(k)}_{\rm mod}(t) - \bar{x}^{(k)}(t)\| = O(\epsilon^{k+1}), \quad k = 1,2.$$

The above results can be formulated in the form of the following theorem:

Theorem 6: Let Assumption 1 be fulfilled. The modified outer solutions $\{\bar{x}^{(n)}_{\rm mod}(t),\bar{y}^{(n)}_{\rm mod}(t)\}$, $n = 0,1,2$, being the solutions to (21) and (27) with the initial conditions (20), tend in the norm to the outer asymptotic solutions $\{\bar{x}^{(n)}(t),\bar{y}^{(n)}(t)\}$, given in (5) and (6), uniformly on $[0,t_0]$ faster than ϵ^n.

It immediately follows that the modified outer solutions of the nth order approximate the exact solutions to (1) $\{x(t),y(t)\}$ with the error of the order of $O(\epsilon^{n+1})$ if only $t \gg \epsilon$. The last condition follows from neglecting the inner solutions for $t = O(\epsilon)$, and has nothing to do with the modification of the outer equations. To achieve the $O(\epsilon^{n+1})$ accuracy also in the initial layer one could take $x^{(n)}_{\rm mod}(t) = \bar{x}^{(n)}_{\rm mod}(t) + \tilde{x}^{(n)}(t)$, $y^{(n)}_{\rm mod}(t) = \bar{y}^{(n)}_{\rm mod}(t) + \tilde{y}^{(n)}(t)$ but usually this region is of little practical importance.

The advantage of the modified outer equations is that they do not contain the lower order solutions in the source terms and hence can be solved in each order ($k = 0,1,2$) independently.

Such systems of equations as (1) are usually encountered in kinetic problems where fast, rapidly decaying transients appear. Then the method permits separation of these transients from the processes which are dominant for long times.

In the next section two examples of application are briefly sketched. The first is connected with the Boltzmann equation in the kinetic theory of gases; the second with the reactor kinetics equations in the neutron transport theory.

5. EXAMPLES OF APPLICATIONS

A. Application for linearized Boltzmann equation for gases. Connection with the Champman–Enskog method

Let us first note that for $q(t) = r(t) = 0$ the system (1) can be written in the form[8]

$$\epsilon\frac{d}{dt}z(t) = \mathscr{A}(t)z(t) + \epsilon\mathscr{B}(t)z(t), \tag{28}$$

where we denote formally $z(t) = \binom{x(t)}{y(t)}$,

$$\mathscr{A}(t) = \begin{pmatrix} A(t), P(t)\\ 0\ ,\ 0\end{pmatrix},$$

$$\mathscr{B}(t) = \begin{pmatrix} 0\ ,\ 0\\ Q(t), B(t)\end{pmatrix}.$$

The characteristic feature of (28) is that for each $t\in[0,t_0]$ the operator $\mathscr{A}(t)$ generates a subspace corresponding to its zero eigenvalue since there exists

$$z_0(t) = \begin{pmatrix} A(t)^{-1}P(t)y_0(t)\\ y_0(t)\end{pmatrix}$$

such that[8]

$$\mathscr{A}(t)z_0(t) = 0.$$

The form of (28) is similar to that of the linearized Boltzmann equation in the kinetic theory of gases[9,10]:

$$\frac{\partial}{\partial t}f + \mathbf{v}\frac{\partial}{\partial \xi}f + \frac{1}{\epsilon}Lf = 0, \tag{29}$$

where $f(\xi,\mathbf{v},t)$ is introduced as the small perturbation from the equilibrium solution, and the distribution function takes the form

$$\begin{aligned}
F(\xi,\mathbf{v},t) &= F_0(|\mathbf{v}|)(1 + f(\xi,\mathbf{v},t)),\\
F_0(|\mathbf{v}|) &= (2\pi)^{-3/2}\exp(-|\mathbf{v}|^2),
\end{aligned} \tag{30}$$

with $\xi,\mathbf{v},t$ being dimensionless position, velocity, and time, respectively. L denotes the collision operator and the small parameter ϵ in (29) indicates that the collision time, as measured on the scale of the characteristic time of the macroscopic (fluid) changes in the distribution function, is a small quantity.

The operator $\mathbf{v}\partial/\partial\xi$ is unbounded in realistic functional spaces and it is not possible to apply our theory directly to (29). To avoid this difficulty let us, following Ref. 11, assume for the moment that the operator $\mathbf{v}\partial/\partial\xi$ is approximated by a bounded operator D, for instance by the discretization in the ξ variable. Then we have instead of (29)

$$\frac{\partial}{\partial t}f + Df + \frac{1}{\epsilon}Lf = 0. \tag{31}$$

It is known[10] that N, the space of the solution to the equation

$$Lf_0 = 0, \tag{32}$$

is spanned by the summational invariants

$$f_0 = \alpha_0 + \boldsymbol{\alpha}_1\cdot m\mathbf{v} + \alpha_2\cdot\tfrac{1}{2}m|\mathbf{v}|^2, \tag{33}$$

where α_0, $\alpha_1^{(i)}$, $i = 1,2,3$; α_2 do not depend on $\mathbf{v}$. Following Ref. 11, one can define $\mathscr{P}$ as the projection operator which projects on N, and the operator $\mathscr{R}$ as $I - \mathscr{P}$. Defining

$$\mathscr{P} f = h,$$

$$\mathscr{R} f = g,$$

one gets from (31) the following systems of singularly perturbed equations:

$$\frac{\partial}{\partial t} h = -\mathscr{P} D \mathscr{P} h - \mathscr{P} D \mathscr{R} g, \tag{34}$$

$$\epsilon \frac{\partial}{\partial t} g = -(\mathscr{R} L \mathscr{R} + \epsilon \mathscr{R} D \mathscr{R}) h - \epsilon \mathscr{R} D \mathscr{P} g.$$

Comparing (1) and (34) we observe that now the operators are ϵ-dependent, the first order-$\epsilon \mathscr{R} D \mathscr{P}$ plays the role of $P(t)$ in (1) and $A(t)$ is to be substituted by the sum $-(\mathscr{R} L \mathscr{R} + \epsilon \mathscr{R} D \mathscr{R})$.

However this additional regular dependence on ϵ is such that it does not change significantly the procedure of the asymptotic expansion in the singular perturbation method. This expansion was obtained in Ref. 11, where the above described procedure, corresponding to the Hilbert expansion method, was introduced and investigated.

The regularly perturbed terms on the right-hand side of (34) can be immediately included in the formalism of the modified singular perturbation method of Sec. 4. This observation is based on the fact that to derive (20), (21), and (27) we may allow the operators $A(t)$, $B(t)$, $P(t)$, and $Q(t)$ to depend regularly on ϵ.

We should have only zero order terms in $A(t)$ different from zero and such that there exists the inverse operator, which holds in the present situation. So we can put formally in (20), (21), and (27) $g(t)$, $h(t)$ instead of $x(t)$ and $y(t)$; the operators $-\mathscr{R} L \mathscr{R} - \epsilon \mathscr{R} D \mathscr{R}$, $-\mathscr{P} D \mathscr{P}$, $-\epsilon \mathscr{R} D \mathscr{P}$, $-\mathscr{P} D \mathscr{R}$ instead of $A(t)$, $B(t)$, $P(t)$, and $Q(t)$, respectively (there are no sources so $q(t) = r(t) = 0$). The effect of the regular perturbation is that several terms are of higher order and may be omitted which, together with the fact that the operators are time-independent, simplifies the equations. For instance, in the first order we obtain from (21) and (27)

$$\frac{\partial}{\partial t} \bar{h}^{(1)}_{\text{mod}}(t) = -(\mathscr{P} D \mathscr{P} - \epsilon \mathscr{P} D \mathscr{R} (\mathscr{R} L \mathscr{R})^{-1} \mathscr{R} D \mathscr{P}) \times \bar{h}^{(1)}_{\text{mod}}(t), \tag{35a}$$

$$\bar{g}^{(1)}_{\text{mod}}(t) = -\epsilon (\mathscr{R} L \mathscr{R})^{-1} \mathscr{R} D \mathscr{P} \bar{h}^{(1)}_{\text{mod}}(t). \tag{35b}$$

These equations were obtained in the first order perturbation by Mika,[12] who considered the Chapman–Enskog procedure of a general order applied to the evolution equations of the type (31). In the second order the same agreement holds. In fact, if the operators in (1) are time-independent and there are no sources, then we see that (27) gives the relation sufficient for the applicability of the Chapman–Enskog method:

$$\bar{x}^{(n)}_{\text{mod}}(t) = C \bar{y}^{(n)}_{\text{mod}}(t), \quad n = 0,1,2,$$

where C is a time-independent bounded operator. It means that with the accuracy of the order $O(\epsilon^{n+1})$, $\bar{x}^{(n)}_{\text{mod}}(t)$ depends on time only via $\bar{y}^{(n)}_{\text{mod}}(t)$.

From the above results it follows that the modified singular perturbation method of solving the system (1) may be viewed as an application of a procedure of the Chapman–Enskog type to the evolution equations with time-dependent operators.

B. Applications in the neutron transport theory

The transport of neutrons in material media is described by the linear Boltzmann equation. First, let us suppose that the dominant physical effect in the interaction of the neutrons with the medium is the scattering, while the absorption and free streaming are of smaller importance. In this situation the analysis would be similar to that of part A (see Refs. 11 and 12).

Now let us consider another situation in the neutron transport theory when the singularly perturbed term appears. It is in the nuclear reactor kinetics where the equations describing the neutron angular flux $\Psi(\mathbf{r},\mathbf{v},t)$ an the precursor densities $C^{(i)}(\mathbf{r},t)$ $i = 1,...,J$ may be written in the form[13,14]:

$$\begin{aligned}
\epsilon \frac{1}{v} \frac{\partial}{\partial t} \Psi(\mathbf{r},\mathbf{v},t)
&= -(\mathbf{\Omega} \nabla + \Sigma) \Psi(\mathbf{r},\mathbf{v},t) \\
&\quad + \int d\mathbf{v}' \sum(\mathbf{r},\mathbf{n}' \to \mathbf{v}) \Psi(\mathbf{r},\mathbf{v},t) \\
&\quad + \frac{1}{4\pi} \int d\mathbf{v}'(1-\beta) \mathscr{X}(v) v \sum_f (\mathbf{r},\mathbf{v}',t) \Psi(\mathbf{r},\mathbf{v}',t) \\
&\quad + \frac{1}{4\pi} \sum_{i=1}^{J} \lambda_i \mathscr{X}_i(v) C^{(i)}(\mathbf{r},t) + s(\mathbf{r},\mathbf{v},t),
\end{aligned} \tag{36}$$

$$\frac{\partial}{\partial t} C^{(i)}(\mathbf{r},t) = -\lambda_i C^{(i)}(\mathbf{r},t) + \beta_i \int d\mathbf{v}' v \sum_f (\mathbf{r};\mathbf{v},t) \Psi(\mathbf{r},\mathbf{v},t).$$

The parameter ϵ, which at the end of our procedure will be put equal to 1, is introduced to indicate that the average lifetime of prompt neutrons (roughly speaking of the order of the time period needed for neutrons to travel the distance between two fission reactions) is a small quantity as compared with the precursors lifetime (of the order of $1/\lambda_i$). Usually the ratio of these two times is about 10^{-2} for thermal reactors and reaches 10^{-5} for fast reactors, where the neutron velocity is much larger. Hendry[13] and Hendry and Bell[14] were the first who applied the technique of singular perturbation to (36). Their approach, however, was based on the matched asymptotic expansion and was different from the method developed in Ref. 5 and presented in this paper in Sec. 1. The analysis of the applications of the latter method in the reactor kinetics can be found in Ref. 15.

The simplest model in reactor kinetics—the so called one-point model—shows explicitly the presence of the two time scales mentioned. The equations of this model are the following:[15–17]

$$\epsilon \frac{\Lambda}{\beta} \frac{d}{dt} N(t) = (\rho(t) - 1) N(t) + \sum_{i=1}^{J} f^{(i)} y^{(i)}(t) + \epsilon \frac{\Lambda}{\beta} q(t), \tag{37a}$$

$$\frac{1}{\lambda_i} \frac{d}{dt} y^{(i)}(t) = N(t) - y^{(i)}(t), \quad i = 1,...,J, \tag{37b}$$

where $N(t)$ is the relative neutron density, $\rho(t)$ is the reactivity insertion, $y^{(i)}(t)$ are proportional to the delayed neutron precursors densities, $f^{(i)} = \bar{\beta}_i/\bar{\beta}$, with $\bar{\beta}_i$ the fraction of the ith group of delayed neutrons, $\bar{\beta} = \Sigma_{i=1}^{J}\bar{\beta}_i$; and Λ is called the prompt neutrons lifetime. For details of the one point approximation and the relation between (36) and (37) we refer the reader to [Ref. 18, Chap. 9].

We immediately observe that the modified singular perturbation method may be applied to the system (37). The only difference with (1) is that in (37) the source term is of first order which, according to the remark of the part A, shifts only some terms to the higher order. The details of the derivation of the modified asymptotic equations and initial conditions can be found in Refs. 16 and 17.

In the zero order the procedure gives the known prompt-jump approximation and the higher orders yield the corresponding corrections.

The modified kinetics equations for the nth order $\bar{y}^{(i)}_{n\mathrm{mod}}(t)$ and $\bar{N}_{n\mathrm{mod}}(t)$ are the following:

$$\frac{1}{\lambda_i}\frac{d}{dt}\bar{y}^{(i)}_{n\mathrm{mod}}(t) = -\frac{1}{a(t)}\left(\sum_{i=1}^{J} f_n^{(i)}(t)\bar{y}^{(i)}_{n\mathrm{mod}}(t) + k_n(t)\right) - \bar{y}^{(i)}_{n\mathrm{mod}}(t), \tag{38a}$$

$$\bar{N}_{n\mathrm{mod}}(t) = -\frac{1}{a(t)}\left(\sum_{i=1}^{J} f_n^{(i)}(t)\bar{y}^{(i)}_{n\mathrm{mod}}(t) + k_n(t)\right), \quad n = 0,1,2, \tag{38b}$$

where

$$a(t) = \rho(t) - 1,$$

$$f_0^{(i)}(t) = f^{(i)},$$

$$f_1^{(i)}(t) = f_0^{(i)}(t) + \epsilon\frac{\Lambda}{\bar{\beta}}\left[\frac{d}{dt}\left(\frac{1}{a(t)} - \frac{\bar{\lambda} + a(t)\lambda_i}{a(t)^2}\right)\right]f^{(i)},$$

$$f_2^{(i)}(t) = f_1^{(i)}(t) + \epsilon^2\left(\frac{\Lambda}{\bar{\beta}}\right)^2\left[\frac{1}{2}\frac{d^2}{dt^2}\left(\frac{1}{a(t)^2}\right) - \frac{1}{a(t)}\left(5\frac{\bar{\lambda}}{a(t)} + 3\lambda_i\right)\frac{d}{dt}\left(\frac{1}{a(t)}\right) + \frac{1}{a(t)^2} \times \left(2\frac{\bar{\lambda}^2}{a(t)^2} + 2\frac{\bar{\lambda}\lambda_i}{a(t)} + \frac{\bar{\lambda}^2}{a(t)} + \lambda_i^2\right)\right]f^{(i)},$$

$$\bar{\lambda} = \sum_{i=1}^{J} f^{(i)}\lambda_i,$$

$$\bar{\lambda}^2 = \sum_{i=1}^{J} f^{(i)}\lambda_i^2,$$

$$k_0(t) = 0,$$

$$k_1(t) = \epsilon\frac{\Lambda}{\bar{\beta}}q(t),$$

$$k_2(t) = k_1(t) + \epsilon^2\left(\frac{\Lambda}{\bar{\beta}}\right)^2\left[\frac{d}{dt}\left(\frac{q(t)}{a(t)}\right) - \frac{\bar{\lambda}q(t)}{a(t)^2}\right].$$

The advantage of (38) over (37) is that (38a) are not stiff and may be solved by standard finite difference methods with time steps of the order of $1/\lambda_i$. The smaller the $(\Lambda/\bar{\beta})/(1/\lambda_{\max})$, the better the accuracy given by the approximated equations (38), while at the same time serious numerical difficulties appear in solving the original system (37). The results of numerical tests and the expressions for the initial conditions for (38a) are given in Refs. 16, 17.

ACKNOWLEDGMENT

The author would like to thank Professor J. Mika for helpful discussions and critical remarks.

[1] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall., Englewood Cliffs, N. J., 1971).
[2] S. G. Krein, *Linear Differential Equations in Banach Space* (American Mathematical Society, Providence, R. I., 1971).
[3] J. Mika, J. Math. Phys. **15**, 892–8 (1974).
[4] J. Mika, Ann. Polon. Math. XXXII, 43–57 (1976).
[5] J. Mika, J. Math. Anal. Appl. **58**, 189–201 (1977).
[6] N. Dunford and J. Schwartz, *Linear Operators, Part I: General Theory* (Interscience, New York, 1958).
[7] J. Kisyński, Studia Math. **23**, 285–328 (1964).
[8] A. Vasilieva, B. Butusov, *Singularly Perturbed Equations in Critical Cases* (Isdatelstvo Moskovskovo Universiteta, Moscow, 1978) (in Russian).
[9] H. Grad, Phys. Fluids, **6**, 147–81 (1963).
[10] H. Grad, "Principles of the Kinetic Theory of Gases," in Vol. XII, Sec. 26 of *Handbuch der Physik*, edited by S. Flügge (Springer–Verlag, Berlin, 1958).
[11] J. Mika, Nukleonika, **23**, 221-35 (1978).
[12] J. Mika "A New Asymptotic Expansion Method for Singularly Perturbed Evolution Equations," internal report (1979).
[13] W. L. Hendry, J. Math. Phys. **11**, 1743 (1970).
[14] W. L. Hendry and G. I. Bell, Nucl. Sci. Eng. **35**, 240–8 (1969).
[15] J. Mika, Nukleonika, **21**, 573–87 (1976).
[16] T. Bleński, A. Gadomski, and J. Mika, Nucl. Sci. Eng. **66**, 277–83 (1978).
[17] T. Bleński, Atomkernenergie **31**, 232–7 (1978).
[18] G. I. Bell and S. Glasstone, *Nuclear Reactor Theory* (Van Nostrand, Princeton, N. J., 1970).

Classical mechanics, the diffusion (heat) equation and the Schrödinger equation on a Riemannian manifold

David Elworthy
Mathematics Institute, University of Warwick, Coventry CV4 7AI, England
Aubrey Truman
Mathematics Department, Heriot-Watt University, Currie, Edinburgh EH14 4As, Scotland

(Received 16 July 1980; accepted for publication 31 July 1980)

We consider the limiting case $\lambda \to 0$ of the Cauchy problem,

$$\frac{\partial g_\lambda(x,t)}{\partial t} = \tfrac{1}{2}\lambda \Delta_x g_\lambda(x,t) + \frac{V(x)}{\lambda} g_\lambda(x,t),$$

with $g_\lambda(x,0) = \exp\{-S_0(x)/\lambda\}T_0(x)$; V, S_0 being real-valued functions on N; T_0 a complex-valued function on N; V, S_0, T_0 being independent of λ; Δ_x being the Laplace–Beltrami operator on N, some complete Riemannian manifold. We prove some new results relating the limiting behavior of the solution to the above Cauchy problem to the solution of the corresponding classical mechanical problem

$$\frac{D^2 Z(s)}{\partial s^2} = -\nabla_Z V[Z(s)], \quad s \in [0,t],$$

with $Z(t) = x$ and $\dot{Z}(0) = \nabla S_0(Z(0))$. One of our results is equivalent to the fact that for short times Schrödinger quantum mechanics on the Riemannian manifold N tends to classical Newtonian mechanics on N as $\hbar$ tends to zero.

PACS numbers: 02.50.Ey, 02.40.Ky, 03.65.Sq

INTRODUCTION

In this paper we consider the limiting case $\lambda \to 0$ of the Cauchy problem

$$\frac{\partial g_\lambda(x,t)}{\partial t} = \tfrac{1}{2}\lambda \Delta_x g_\lambda(x,t) + \frac{V(x)}{\lambda} g_\lambda(x,t),$$

with $g_\lambda(x,0) = \exp\{-S_0(x)/\lambda\}T_0(x)$, Δ_x being the Laplace–Beltrami operator for some complete Riemannian manifold N, V being a real-valued potential on N; T_0, S_0 being functions on N independent of λ, S_0 being real valued and T_0 complex-valued, for both real and pure imaginary values of λ.

When $\lambda = i\hbar$ is pure imaginary, the above equation is a Schrödinger equation and the initial data corresponds to an initial particle density $\rho(x) = |T_0(x)|^2$ and to a limiting value of the probability current $j_{\hbar=0}(x) = \rho(x)\nabla S_0(x)$. Here we present some new relationships between the limiting cases of the above Cauchy problems and the solution of the Newton equation on N

$$\frac{D^2 Z(s)}{\partial s^2} = -\nabla_Z V[Z(s)], \quad s \in [0,t],$$

with $Z(t) = x$ and $\dot{Z}(0) = \nabla S_0(Z(0))$. These relationships include as a special case the result that, for small times, in the curved space background of the Riemannian manifold N, quantum mechanics tends to classical mechanics as $\hbar$ tends to zero.

To be specific, when $\lambda = \mu^2 (\mu > 0)$, let N be a complete n-dimensional Riemannian manifold with Laplace–Beltrami operator Δ and with a smooth map $V: N \to \mathbb{R}$ bounded above. Assume that N is stochastically complete i.e., its Brownian motion is defined for all time. This is true, for example, if the Ricci curvature of N is bounded below, in particular if N is compact. For smooth maps $T_0: N \to \mathbb{C}$, $S_0: N \to \mathbb{R}$, set

$$g_0^\mu(x) = T_0(x)\exp\{-S_0(x)/\mu^2\}, \quad x \in N, \tag{1}$$

and let $g^\mu: N \times [0,\infty] \to \mathbb{R}$ be the minimal solution of the diffusion equation on N, $g^\mu = g^\mu(x,t)$:

$$\frac{\partial g^\mu(x,t)}{\partial t} = \frac{\mu^2}{2}\Delta_x g^\mu(x,t) + \frac{V(x)}{\mu^2} g^\mu(x,t), \tag{2}$$

with $g^\mu(x,0) = g_0^\mu(x)$, $x \in N$.

For the case when N is the n-dimensional Euclidean space $\mathbb{R}^n$, under suitable conditions on S_0, T_0, and V, Truman[1] has obtained an expansion of the Wiener path integral expression for g^μ in powers of μ and has given a corresponding expansion, the quasiclassical expansion, for the Feynman path integral representation of the solution to the Schrödinger equation obtained by replacing μ^2 by $(i\hbar)$ in Eqs. (1) and (2). Using this expansion he was able to relate the limiting behavior as $\mu \to 0$ of both the diffusion equation and the Schrödinger equation to the behavior of the corresponding classical mechanical system.

Here we follow the same method and using some of the ideas of Eells and Elworthy[2,3] we obtain a "quasiclassical expansion" for the solution g^μ of the above Cauchy problem. This result is summarized in Theorem 1E. An important application of this expansion is that it suggests corresponding expansions for the solutions of the corresponding Schrödinger equation in terms of Feynman path integrals which could be computable. This expansion is given in Sec. 5, Eqs. (75), (77), and (78).

Thus, for instance, if we consider a mechanical system with a classical Lagrangian, $\mathscr{L}_{CL}(q,\dot{q}) = 2^{-1}\sum_{\alpha,\beta=1}^n g_{\alpha\beta}(q)\dot{q}^\alpha\dot{q}^\beta - V(q)$, with $g_{\alpha\beta}$, for $\alpha, \beta = 1, 2, \ldots, n$, being the metric tensor of the Riemannian manifold N, we obtain a formal Feynman integral expression for the corresponding quantum mechanical propagator as a power series in $\hbar$. Most previous functional integral expressions for this curved space propagator have been afflicted with infinities or ambiguities which for the quasiclassical expansion do not seem to be a problem.

 0022-2488/81/102144-23$01.00

biguities which for the quasiclassical expansion do not seem to be a problem.

The program of computing the terms in this quasiclassical expansion has already been carried out (heuristically) up to the Wentzel–Kramers–Brillouin (WKB) approximation by De Witt–Morette, Maheshwari and Nelson.[4] Their argument is based on analogy with the diffusion equation theory which we present in Secs. 1 and 2. This mathematical background is also a necessary preliminary to carrying out the calculations to higher order. In these sections we have freely used results about the stochastic development from Eells and Elworthy.[2,3]

In the DeWitt, Maheshwari, and Nelson paper[4] the Feynman integrals are defined by "prodistributions", whereas in the present work we examine Feynman's original definition of the integral—the Feynman map $\mathscr{F}$. This is defined for the Schrödinger equation on a Riemannian manifold by exploiting known results for the corresponding diffusion (heat) equation. This is done in Sec. 4. There we give a formula, Eq. (71), for the finite-dimensional integrals involved. It is in terms of the Van–Vleck determinant and we compare it with the Pauli, Van–Vleck, De Witt propagator. Our result for the short-time propagator is different from that usually employed by physicists. It is, however, consistent with the earlier results of DeWitt in Ref. 5 when one takes into consideration the fact that our Schrödinger equation does not involve any curvature terms. We also make the obvious conjecture, Sec. 4D, concerning the convergence of the finite-dimensional integrals. In Sec. 5 we use a Cameron–Martin type theorem, which is proved in Ref. 6, for the Feynman map to give a mathematical justification for the computation in Ref. 4 of the first term of the quasiclassical expansion.

The conditions under which this quasiclassical approximation is valid are discussed for the diffusion equation and the Schrödinger equation in Secs. 3 and 6. For the diffusion equation we examine the approach of Schilder, Donsker, and Varadhan [see Ref. 7] to give some information, Theorem 3C, about the quasiclassical expansion. For the Schrödinger equation using a theorem of Gaffney we give a simple proof of the validity of the WKB approximation for short time: Theorem 6F. This theorem establishes the result that for short times in the curved space background of a Riemannian manifold Schrödinger's quantum mechanics tends to classical Newtonian mechanics as $\hbar$ tends to zero. Similar results for the diffusion equation and for arbitrary times are discussed in Refs. 8 and 9.

I. THE QUASICLASSICAL REPRESENTATION FOR THE DIFFUSION (HEAT) EQUATION ON A RIEMANNIAN MANIFOLD

A. The Feynman-Kac formula for a Riemannian manifold

We shall use the description of Brownian motion on N via the "stochastic development". Let $\pi: O(N) \to N$ denote the orthonormal frame bundle of N. An element $u_0 \in \pi^{-1}(x_0)$ can be considered as an isometry $u_0: R^n \to T_{x_0}N$ onto the tangent space $T_{x_0}N$ to N at x_0. Here for $e=(e^1, \ldots, e^n)$ in R^n, $u_0(e) \in T_{x_0}N$ is the vector with components $(e^1, \ldots, e^n)$ relative to the frame $u_0 \in \pi^{-1}(x_0)$. The Levi-Civita connection of N determines a map

$$X: O(N) \times R^n \to TO(N),$$

which trivializes the horizontal tangent bundle to $O(N)$: if $(u,e) \in O(N) \times R^n$, then $X(u,e)$ is the unique horizontal vector in $T_uO(N)$ with

$$\pi^1 X(u,e) = u(e) \tag{3}$$

for the derivative map π^1 of the projection $\pi^1: TO(N) \to TN$.

Now choose a fixed time $t>0$ and consider Brownian motion ω on R^n defined up to the time t. We can take this to be defined by Wiener measure γ on the space $\Omega = C_0(R^n)$ of continuous paths $\omega: [0,t] \to R^n$, $\omega(0)=0$.

For a fixed $x_0 \in N$ and $u_0 \in \pi^{-1}(x_0)$, let $u^\mu: [0,t] \times \Omega \to O(N)$, $\mu>0$, be the solution of the stochastic differential equation (in the Stratanovich sense)

$$du^\mu = X(u^\mu, \mu d\omega), \tag{4}$$

with $u^\mu(0,\omega)=u_0$, $\omega \in \Omega$. Define $x^\mu: [0,t] \times \Omega \to N$, $\mu>0$, by $x^\mu = \pi u^\mu$. Then for $\mu=1$ the process x^μ would represent Brownian motion on N and u^μ its "horizontal lift". Stretching notation we have in some sense

$$dx^\mu(s,\omega) = u^\mu(s,\omega)\mu d\omega(s), \tag{5}$$

with $x^\mu(0,\omega)=x_0$, $\omega \in \Omega$, $\mu>0$.

The mapping $C_0(R^n) \to C_{x_0}(N)$, $\omega \to x^1(\cdot, \omega)$ defined almost everywhere and with values in the space of continuous paths $\alpha: [0,t] \to N$ with $\alpha(0)=x_0$ is the "stochastic development". For a smooth path ω, Eq. (5) can be interpreted as an ordinary differential equation and $x^1(\cdot,\omega)$ is exactly the Cartan development of $(u_0 \circ \omega)$: $[0,t] \to T_{x_0}N$.

The solution g^μ to Eqs. (1) and (2) is given by the Feynman–Kac formula

$$g^\mu(x_0,t) = \int_\Omega \exp\left\{\mu^{-2}\int_0^t V[x^\mu(s,\omega)]\,ds\right\} g_0^\mu[x^\mu(t,\omega)] \times d\gamma(\omega). \tag{6}$$

Now suppose $Z: [0,t] \to N$ satisfies the classical equation

$$\frac{D^2 Z(s)}{\partial s^2} = -\nabla V[Z(s)], \tag{7}$$

with $Z(0)=x_0$ and $\dot{Z}(t) = -\nabla S_0[Z(t)]$. Then there is a unique smooth path $\sigma(\cdot)$ in R^n which has Z as the development of $(u_0 \circ \sigma)$.

For this path σ, let $v^\mu: [0,t] \times \Omega \to O(N)$, $\mu>0$, satisfy the time-dependent stochastic differential equation

$$dv^\mu = X(v^\mu, \mu d\omega + \dot{\sigma} ds), \tag{8}$$

with $v^\mu(0,\omega)=u_0$, $\omega \in \Omega$. In particular $v^0(\cdot,\omega)$ can be taken to be the horizontal lift of Z, all $\omega \in \Omega$. Define $y^\mu: [0,t] \times \Omega \to N$ by $y^\mu = \pi v^\mu$. Then $y^0(\cdot,\omega)$ is independent of ω and identical to $Z(\cdot)$. Again, in some sense we have

$$dy^\mu(s,\omega) = \mu v^\mu(s,\omega) d\omega(s) + v^\mu(s,\omega)\dot{\sigma}(s) ds, \tag{9}$$

with $y(0,\omega)=x_0$, $\omega \in \Omega$, $\mu>0$.

(Later we shall be differentiating y^μ with respect to μ at $\mu=0$. This is some sense differentiating the stochastic development at σ.)

By the Girsanov–Cameron–Martin formula the measures $y^\mu(\gamma)$ and $x^\mu(\gamma)$ induced on $C_{x_0}(N)$ by y^μ and x^μ are equivalent with the Radon–Nikodym derivative at $y^\mu(\cdot,\omega)$:

$$\frac{dx^\mu(\gamma)}{dy^\mu(\gamma)}=\exp\left\{-\mu^{-1}\int_0^t\langle\dot\sigma(s),d\omega(s)\rangle-\tfrac{1}{2}\mu^{-2}\int_0^t|\dot\sigma(s)|^2\,ds\right\}. \tag{10}$$

Thus, for $\mu>0$,

$$g^\mu(x_0,t)=\int_\Omega\exp\left\{\mu^{-2}\int_0^t V[y^\mu(s,\omega)]\,ds-\mu^{-1}\int_0^t\langle\dot\sigma(s),\,d\omega(s)\rangle-2^{-1}\mu^{-2}\int_0^t[\dot\sigma(s)]^2\,ds\right\}\times g_0^\mu[y^\mu(t,\omega)]d\gamma(\omega). \tag{11}$$

Now, following the quasiclassical representation method, set

$$S=S_0[Z(t)]+2^{-1}\int_0^t|\dot Z(s)|^2ds-\int_0^t V[Z(s)]\,ds. \tag{12}$$

Substituting the expression for g_0^μ into Eq. (11) and using $|\dot Z(s)|=|\dot\sigma(s)|$,

$$\exp\{\mu^{-2}S\}g^\mu(x_0,t)=\int_\Omega T_0[y^\mu(t,\omega)]\exp\{A(\mu,\omega)\}\,d\gamma(\omega), \tag{13}$$

where

$$A(\mu,\omega)=\mu^{-2}\int_0^t\{V[y^\mu(s,\omega)]-V[Z(s)]\}ds-\mu^{-2}\{S_0[y^\mu(t,\omega)]-S_0[Z(t)]\}-\mu^{-1}\int_0^t\langle\dot\sigma(s),d\omega(s)\rangle. \tag{14}$$

B. Expansion in powers of μ

Now y^μ is infinitely differentiable with respect to μ in probability. In fact by results of Baxendale and Malliavin for a class of manifolds N, including compact N, there exists a version which is almost surely C^∞ in μ. Let $\delta y^\mu:[0,t]\times\Omega\to TN$ be the first derivative with respect to μ and let δ^2y^μ be the covariant derivative

$$\delta y^\mu=\frac{\partial y^\mu}{\partial\mu},\quad \delta^2y^\mu=\frac{D\delta y^\mu}{\partial\mu}, \tag{15}$$

$\delta^2y^\mu:[0,t]\times\Omega\to TN$. Inductively set

$$\delta^p y^\mu=\frac{D}{\partial\mu}\delta^{p-1}y^\mu,\quad p\geq 2, \tag{16}$$

$\delta^p y^\mu:[0,t]\times\Omega\to TN$.

The Taylor expansions in μ about $\mu=0$ of $V[y^\mu(s,\omega)]$ and $S[y^\mu(t,\omega)]$ give

$$\begin{aligned}A(\mu,\omega)=\mu^{-2}\int_0^t\{&\mu dV[Z(s)]\delta y(s)+2^{-1}\mu^2\nabla dV[Z(s)]\\&\times(\delta y(s),\delta y(s))+2^{-1}\mu^2dV[Z(s)]\delta^2y(s)\}ds\\&-\mu^{-2}\{\mu^{-1}dS_0[Z(t)]\delta y(t)+2^{-1}\mu^2\nabla dS_0[Z(t)]\\&\times(\delta y(t),\delta y(t))+2^{-1}\mu^2dS_0[Z(t)]\delta^2y(t)\}\\&-\mu^{-1}\int_0^t\langle\dot\sigma(s),d\omega(s)\rangle+R_1(\mu,\omega)+R_2(\mu,\omega),\end{aligned} \tag{17}$$

where δy and δ^2y denote δy^0 and δ^2y^0, while

$$\begin{aligned}R_1(\mu,\omega)=2^{-1}\mu\int_0^1(1-\theta)^2\int_0^t\{&\nabla^2dV[y^{\theta\mu}(s)]\\&\times(\delta y^{\theta\mu}(s),\delta y^{\theta\mu}(s),\delta y^{\theta\mu}(s))\\&+\nabla dV[y^{\theta\mu}(s)](\delta^2y^{\theta\mu}(s),\delta y^{\theta\mu}(s))\\&+2\nabla dV[y^{\theta\mu}(s)](\delta y^{\theta\mu}(s),\delta^2y^{\theta\mu}(s))\\&+dV[y^{\theta\mu}(s)]\delta^3y^{\theta\mu}(s)\}ds\,d\theta\end{aligned} \tag{18}$$

and

$$\begin{aligned}R_2(\mu,\omega)=-2^{-1}\mu\int_0^1(1-\theta)^2\{&\nabla^2dS_0[y^{\theta\mu}(t)]\\&\times(\delta y^{\theta\mu}(t),\delta y^{\theta\mu}(t),\delta y^{\theta\mu}(t))\\&-\nabla dS_0[y^{\theta\mu}(t)](\delta^2y^{\theta\mu}(t),\delta y^{\theta\mu}(t))\\&-2\nabla dS_0[y^{\theta\mu}(t)](\delta y^{\theta\mu}(t),\delta^2y^{\theta\mu}(t))\\&+dS_0[y^{\theta\mu}(t)]\delta^3y^{\theta\mu}(t)\}d\theta.\end{aligned} \tag{19}$$

Since σ is C^2, partial integration yields

$$\begin{aligned}\int_0^t\langle\dot\sigma(s),d\omega(s)\rangle&=\langle\dot\sigma(t),\omega(t)\rangle-\int_0^t\langle\ddot\sigma(s),\omega(s)\rangle ds\\&=\langle\dot Z(t),v^0(t)\omega(t)\rangle-\int_0^t\langle\ddot Z(s),v^0(s)\omega(s)\rangle\,ds\\&=-dS_0[Z(t)]v^0(t)\omega(t)+\int_0^t dV[Z(s)]v^0(s)\omega(s)\,ds\end{aligned} \tag{20}$$

by Eq. (7), giving

$$\begin{aligned}A(\mu,\omega)=\mu^{-1}\int_0^t &dV[Z(s)]\{\delta y(s)-v^0(s)\omega(s)\}ds\\&-\mu^{-1}dS_0[Z(t)]\{\delta y(t)-v^0(t)\omega(t)\}\\&+2^{-1}\int_0^t\nabla dV[Z(s)](\delta y(s),\delta y(s))\,ds\\&-2^{-1}\nabla dS_0[Z(t)](\delta y(t),\delta y(t))\\&+2^{-1}\int_0^t dV[Z(s)]\delta^2y(s)\,ds-2^{-1}dS_0[Z(t)]\delta^2y(t)\\&+R_1(\mu,\omega)+R_2(\mu,\omega).\end{aligned} \tag{21}$$

C. Formulas for δy

The formulas involving δy and δ^2y which we will obtain next, will be proved again later while obtaining additional information. However, some readers may find the following derivation more pleasant.

Consider a smooth $\alpha\in C_0(R^n)$. By solving Eq. (8) with $\omega=\alpha$ as an ordinary differential equation we get a smooth path $v^\mu(\cdot,\alpha)$ in $O(N)$ projecting to a smooth path $y^\mu(\cdot,\alpha)$ in N. With this convention

$$\frac{d}{ds}y^\mu(s,\alpha)=\mu v^\mu(s,\alpha)\dot\alpha(s)+v^\mu(s,\alpha)\dot\sigma(s), \tag{22}$$

whence

$$\frac{D}{\partial\mu}\frac{d}{ds}y^\mu(s,\alpha)=v^\mu(s,\alpha)\dot\alpha(s)+\mu\frac{D}{\partial\mu}[v^\mu(s,\alpha)\dot\alpha(s)]+\frac{D}{\partial\mu}[v^\mu(s,\alpha)\dot\sigma(s)], \tag{23}$$

giving

$$\frac{D}{\partial s}\delta y(s,\alpha) = v^0(s,\alpha)\dot{\alpha}(s) + \frac{D}{\partial \mu}[v^\mu(s,\alpha)\dot{\sigma}(s)]\,|_{\mu=0}\,. \qquad (24)$$

Now, if R denotes the curvature tensor (with the sign conventions of Kobayashi and Nomizu[10]), for $0 \le s_0 \le t$,

$$\frac{D}{\partial s}\frac{D}{\partial \mu}[v^\mu(s,\alpha)\dot{\sigma}(s_0)]\,|_{\mu=0} = \frac{D}{\partial \mu}\frac{D}{\partial s}[v^\mu(s,\alpha)\dot{\sigma}(s_0)]\,|_{\mu=0} + R[\dot{Z}(s),\delta y(s,\alpha)]v^0(s,\alpha)\dot{\sigma}(s_0)$$
$$= R[\dot{Z}(s),\delta y(s,\alpha)]v^0(s,\alpha)\dot{\sigma}(s_0)\,, \qquad (25)$$

since $v^\mu(\cdot,\alpha)\dot{\sigma}(s_0)$ is just a parallel translation of $\dot{\sigma}(s_0)$ along $y^\mu(\cdot,\alpha)$. Denoting $v^0(s,\alpha)$ by $v^0(s)$ and working in the parallel translated orthonormal frame $v^0(s)$, the above equation is equivalent to

$$\frac{d}{ds}\left\{v^0(s)^{-1}\frac{D}{\partial\mu}[v^\mu(s,\alpha)\dot{\sigma}(s_0)]\right\}\Big|_{\mu=0} = v^0(s)^{-1}R[\dot{Z}(s),\delta y(s,\alpha)]v^0(s)\dot{\sigma}(s_0)\,, \qquad (26)$$

or

$$\frac{d}{ds}\left\{v^0(s)^{-1}\frac{D}{\partial\mu}[v^\mu(s,\alpha)\dot{\sigma}(s)]\right\}\Big|_{\mu=0} = v^0(s)^{-1}R[\dot{Z}(s),\delta y(s,\alpha)]v^0(s)\dot{\sigma}(s) + v^0(s)^{-1}R[\dot{Z}(s),\delta y(s,\alpha)]v^0(s)\ddot{\sigma}(s)$$
$$= \frac{d}{ds}\int_0^s v^0(r)^{-1}R[\dot{Z}(r),\delta y(r,\alpha)]v^0(r)\dot{\sigma}(s)\,dr\,. \qquad (27)$$

Combining this with Eq. (24) finally yields

$$\frac{d}{ds}\{v^0(s)^{-1}\delta y(s,\alpha) - \alpha(s)\} = \int_0^s v^0(r)^{-1}R[\dot{Z}(r),\delta y(r,\alpha)]v^0(r)\dot{\sigma}(s)\,dr\,. \qquad (28)$$

This equation is true for all smooth paths α and even for piecewise C^1 paths α. We can approximate our Brownian motion paths ω by such α so that as the given approximations converge to ω the solutions $y^\mu(r,\alpha)$ converge almost surely, uniformly for $r \in [0,t]$ and uniformly together with all derivatives in $\mu \in [0,1]$, to a version of $y^\mu(r,\omega)$. (For compact N Malliavin's regularization, or piecewise linear approximations, will do. For noncompact N we can use an additional limit over compact domains of N. Essentially this is Malliavin's transfer principle.[11] However, if we are prepared to work in more detail as below there is no need for these deep theorems.) In any case we have:

Proposition 1C: The process $[\omega(s) - v^0(s)^{-1}\delta y(s)]$ has almost surely C^2 sample paths with

$$\frac{d}{ds}[v^0(s)^{-1}\delta y(s) - \omega(s)] = \int_0^s v^0(r)^{-1}R[\dot{Z}(r),\delta y(r)]v^0(r)\dot{\sigma}(s)\,dr\,. \qquad (29)$$

Since $v^0(r)^{-1}R[\dot{Z}(r),\delta y(r)]v^0(r): T_{y^0(r)}N \to T_{y^0(r)}N$ is skew-symmetric for each r, so is its integral with respect to r. Consequently, applying both sides of above equation to $\dot{\sigma}(s)$, we arrive at

$$\left\langle\frac{D}{\partial s}[\delta y(s) - v^0(s)\omega(s)],\dot{Z}(s)\right\rangle = 0\,,\quad 0 \le s \le t\,. \qquad (30)$$

This enables us to prove the following lemma.

Lemma 1C: $A(\mu,\omega)$

$$= 2^{-1}\int_0^t \nabla dV[Z(s)](\delta y(s),\delta y(s))\,ds - 2^{-1}\nabla dS_0[Z(t)](\delta y(t),\delta y(t))$$
$$+ 2^{-1}\int_0^t dV[Z(s)]\,\delta^2 y(s)\,ds - 2^{-1}\,dS_0[Z(t)]\,\delta^2 y(t) + R_1(\mu,\omega) + R_2(\mu,\omega)\,. \qquad (31)$$

Proof: The coefficient of μ^{-1} in $A(\mu,\omega)$ is just

$$-\int_0^t \langle\ddot{Z}(s),\delta y(s) - v^0(s)\omega(s)\rangle\,ds + \langle\dot{Z}(t),\delta y(t) - v^0(t)\omega(t)\rangle \qquad (32)$$

which vanishes on integration by parts.

D. Formulas for $\delta^2 y$

We must now examine the terms in $\delta^2 y$. To do this we return to the case of our smooth path α. Using Eq. (22) we obtain

$$\frac{\partial}{\partial\mu}\left|\frac{\partial}{\partial s}y^\mu(s,\alpha)\right|^2 = \frac{\partial}{\partial\mu}\,|\mu\dot{\alpha}(s) + \dot{\sigma}(s)|^2$$
$$= 2\langle\dot{\alpha}(s),\mu\dot{\alpha}(s) + \dot{\sigma}(s)\rangle\,. \qquad (33)$$

Consequently

$$\frac{\partial^2}{\partial\mu^2}\left|\frac{\partial y^\mu}{\partial s}(s,\alpha)\right|^2 = 2\,|\dot{\alpha}(s)|^2\,. \qquad (34)$$

On the other hand

$$\frac{\partial^2}{\partial\mu^2}\left|\frac{\partial y^\mu}{\partial s}(s,\alpha)\right|^2 = 2\frac{\partial}{\partial\mu}\left\langle\frac{D}{\partial s}\delta y^\mu(s,\alpha),\frac{\partial y^\mu}{\partial s}(s,\alpha)\right\rangle$$
$$= 2\left|\frac{D}{\partial s}\delta y^\mu(s,\alpha)\right|^2 + 2\left\langle\frac{D}{\partial s}\delta^2 y^\mu(s,\alpha),\frac{\partial y^\mu}{\partial s}(s,\alpha)\right\rangle$$
$$+ 2\left\langle R\left[\delta y^\mu(s,\alpha),\frac{\partial y^\mu}{\partial s}(s,\alpha)\right]\delta y^\mu(s,\alpha),\dot{Z}(s)\right\rangle\,. \qquad (35)$$

Setting $\mu = 0$, Eqs. (34) and (35) give

$$\frac{d}{ds}\langle \delta^2 y(s,\alpha), \dot Z(s)\rangle - \langle \delta^2 y(s,\alpha), \ddot Z(s)\rangle$$

$$= -\left|\frac{D}{\partial s}\delta y(s,\alpha)\right|^2$$

$$+ |\dot\alpha(s)|^2 - \langle R[\delta y(s,\alpha), \dot Z(s)]\delta y(s,\alpha), \dot Z(s)\rangle$$

$$= -\left|\frac{d}{ds}[\alpha(s) - v^0(s,\alpha)^{-1}\delta y(s,\alpha)]\right|^2$$

$$+ 2\left\langle \frac{d}{ds}[\alpha(s) - v^0(s,\alpha)^{-1}\delta y(s,\alpha)], \dot\alpha(s)\right\rangle$$

$$- \langle R[\delta y(s,\alpha), \dot Z(s)]\delta y(s,\alpha), \dot Z(s)\rangle . \tag{36}$$

We can integrate both sides of the last equation with respect to s to give an equation valid for all sufficiently smooth α. As before we can deduce

Lemma D: Almost surely

$$\langle \delta^2 y(t), \dot Z(t)\rangle + \int_0^t \langle \delta^2 y(s), \nabla V[Z(s)]\rangle\, ds$$

$$= 2\int_0^t \left\langle \frac{d}{ds}[\omega(s) - v^0(s)^{-1}\delta y(s)], d\omega(s)\right\rangle$$

$$- \int_0^t \left|\frac{d}{ds}[\omega(s) - v^0(s)^{-1}\delta y(s)]\right|^2 ds$$

$$- \int_0^t \langle R[\delta y(s), \dot Z(s)]\delta y(s), \dot Z(s)\rangle\, ds . \tag{37}$$

Consequently we obtain:

$$A(\mu,\omega) = \int_0^t \left\langle \frac{d}{ds}[\omega(s) - v^0(s)^{-1}\delta y(s)], d\omega(s)\right\rangle$$

$$- \frac{1}{2}\int_0^t \left|\frac{d}{ds}[\omega(s) - v^0(s)^{-1}\delta y(s)]\right|^2 ds$$

$$- \frac{1}{2}\int_0^t \langle R[\delta y(s), \dot Z(s)]\delta y(s), \dot Z(s)\rangle$$

$$+ \frac{1}{2}\int_0^t \nabla dV[Z(s)](\delta y(s), \delta y(s))\, ds$$

$$- \frac{1}{2}\nabla dS_0[Z(t)](\delta y(t), \delta y(t)) + R_1(\mu,\omega) + R_2(\mu,\omega) . \tag{38}$$

E. The quasiclassical expansion for the heat equation

Now Proposition 1C shows that δy is a linear function of ω. We would like to convert our basic formula, Eq. (13), to an integral over the space of vector fields along Z furnished with the Gaussian measure $\delta y(\gamma)$. However, it is simpler and comes to essentially the same thing to make the change of variables on $C_0(R^n)$,

$$\theta : C_0(R^n) \to C_0(R^n) ,$$

$$\theta(\omega)(s) = v^0(s)^{-1}\delta y(s,\omega) . \tag{39}$$

Setting $\theta_0 = \theta - \mathrm{I}$, where I is the identity map, Proposition 1C gives an explicit expression for θ_0^{-1} as an iterated Volterra operator. Restricted to the reproducing kernel Hilbert space $L_0^{2,1}(R^n)$ of γ, it follows that θ_0 determines a trace-class operator with no nonzero eigenvalues. Consequently the Fredholm determinant, $\det\theta$, on $L_0^{2,1}(R^n)$ is such that $\det\theta = 1$. Consequently by the Cameron–Martin formula, $\theta^{-1}(\gamma) \approx \gamma$, with

$$\frac{d\theta^{-1}}{d\gamma}(\gamma)(\omega) = \exp\left\{- \int_0^t \left\langle \frac{d}{ds}[\theta(\omega)(s) - \omega(s)], d\omega(s)\right\rangle \right.$$

$$\left. - 2^{-1}\int_0^t \left|\frac{d}{ds}[\theta(\omega)(s) - \omega(s)]\right|^2 ds\right\}. \tag{40}$$

(Alternatively we could use the Cameron–Martin–Girsanov formula, in which case we would not have to bother about the Fredholm determinant.) Applying this change of variable to the right-hand side of Eq. (13), using Eq. (38), we obtain:

Theorem 1E: Let Z satisfy the classical equations of motion

$$\frac{D^2 Z}{\partial s^2}(s) = -\nabla V[Z(s)], \quad 0 \le s \le t , \tag{41}$$

with $Z(0) = x_0$ and $\dot Z(t) = -\nabla S_0[Z(t)]$. Let $g^\mu(x_0, t)$ be the solution of the equation

$$\frac{\partial g^\mu}{\partial t} = \left[\frac{\mu^2}{2}\Delta_{x_0} + \frac{V}{\mu^2}\right] g^\mu ,$$

with $g^\mu(\cdot, 0) = \exp\{\mu^{-2}S_0(\cdot)\}T_0(\cdot)$, for smooth T_0, S_0, V with V bounded above. Then if

$$S = S_0[Z(t)] + 2^{-1}\int_0^t |\dot Z(s)|^2 ds - \int_0^t V[Z(s)]\, ds ,$$

we have

$$\exp\{\mu^{-2}S\} g^\mu(x_0, t) = \int_\Omega T_0[z^\mu(t,\omega)] \exp\{B(\mu,\omega)\}\, d\gamma(\omega) , \tag{42}$$

where $z^\mu(s,\omega) = y^\mu(s, \theta^{-1}\omega)$, $0 \le s \le t$, $\mu \ge 0$, *and*

$$B(\mu,\omega) = 2^{-1}\int_0^t \nabla dV[Z(s)](v^0(s)\omega(s), v^0(s)\omega(s))\, ds$$

$$- 2^{-1}\nabla dS_0[Z(t)](v^0(t)\omega(t), v^0(t)\omega(t))$$

$$+ 2^{-1}\int_0^t \langle R[v^0(s)\omega(s), \dot Z(s)]v^0(s)\omega(s), \dot Z(s)\rangle\, ds$$

$$+ \rho(\mu,\omega) , \tag{43}$$

with $\rho(\mu,\omega) = R_1(\mu, \theta^{-1}\omega) + R_2(\mu, \theta^{-1}\omega)$, as given by Eqs. (18) and (19). In particular $\rho(\mu,\omega) \to 0$ and $z^\mu(t,\omega) \to Z(t)$, almost surely, as $\mu \to 0$.

Theorem 1E gives the quasiclassical representation for the diffusion (heat) equation on a Riemannian manifold N. The analogous (but now formal) expression for the Schrödinger equation on the Riemannian manifold N can be written down almost immediately and is given in a subsequent section. The above quasiclassical representations are formal power series in $\mu^2 = i\hbar$ with coefficients which are either Feynman or Wiener integrals. We shall evaluate the leading terms in these series.

F. The expansion rewritten

When we attempt to use Feynman path integrals it will be useful to have a record of what we have done to the integrand of our original path integral (1). We will also need to introduce a (positive) variance parameter β. For this let γ_B denote Wiener measure with variance parameter β on $C^0(R^n)$. Also let A be the translation of $C^0(R^n)$,

$$A(\omega)=\omega-\mu^{-1}\sigma$$

and set

$$\Theta=\theta\cdot A.$$

Then our work shows that $\Theta(\omega)-\omega$ is always C^2 and

$$\exp\left(\frac{\mu^{-2}S}{\beta}\right)\exp\left\{\frac{\mu^{-2}}{\beta}\int_0^t V(x^\mu(s,\omega))\,ds-\frac{\mu^{-2}}{\beta}S_0(x^\mu(t,\omega))\right\}$$

$$=\exp\left\{\frac{-1}{\beta}\int_0^t\left\langle\frac{d}{ds}(\Theta(\omega)(s)-\omega(s)),d\omega(s)\right\rangle\right.$$

$$\left.-\frac{1}{2\beta}\int_0^t\left|\frac{d}{ds}(\Theta(\omega)(s)-\omega(s))\right|^2 ds\right\}$$

$$\times\exp\left\{\frac{1}{\beta}B(\mu,\Theta(\omega))\right\},\quad \mu>0 \tag{44}$$

for γ-almost all $\omega\in C^0(R^n)$, and for γ_β-almost all ω if x^μ is interpreted using γ_β. It also holds for all $\omega\in L_0^{2,1}(R^n)$, when x^μ is interpreted via the classical solution to (4).

2. STOCHASTIC DIFFERENTIAL GEOMETRY

A. The derivative process of a stochastic dynamical system

In this section we review some general methods of stochastic differential geometry making applications to the quasiclassical expansion.

In order to examine δy^μ and $\delta^2 y^\mu$ in more detail, first consider a general stochastic differential equation

$$dz=Xd\omega+Wds \tag{45}$$

on a manifold M. Here X: $M\times R^n\to TM$ and the (possibly time dependent) vector field W are suitably smooth. For $e\in R^n$ let $S(s,m)e$ be the solution of the ordinary differential equation on M

$$\frac{dy}{ds}=X(y(s),e),\quad y(0)=m.$$

For the standard basis $e_1,\dots,e_n$ of R^n let X^i denote the vector field $X(-,e_i)$ and set $S^i(s,m)=S(s,m)e_i$. Then for a C^2 map $f:M\to R$ the Itô formula for a solution x of (45), assumed nonexplosive, can be written

$$f(x(t))=f(x(0))+\sum_{i=1}^N\int_0^t\frac{d}{dr}f\circ S^i(r,x(s))\big|_{r=0}dw^i(s),$$

$$+\frac{1}{2}\sum_{i=1}^n\int_0^t\frac{d^2}{dr^2}f\circ S^i(r,x(s))\big|_{r=0}ds+\int_0^t df\cdot W(x(s))\,ds, \tag{46}$$

where $\omega^i(s)=\langle\omega(s),e_i\rangle$ is the ith component of $\omega(s)$.

From (X,W) we can form a stochastic dynamical system $(\delta X,\delta W)$ on TM by taking

$$\delta X:TM\times\mathbb{R}^n\to TTM$$

to be defined by $\delta X^i=\delta X(\ ,e_i)=S\circ TX^i:TM\to TTM$, $i=1,\dots,n$, where

$$S:T^2M\to T^2M$$

is the symmetry map given over an open set U of $\mathbb{R}^m$ as

$$S:U\times\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}^m\to U\times\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}^m,$$

$$S(x,u,v,\omega)=(x,v,u,\omega).$$

The vector field δW on TM is defined the same way:

$$\delta W=S\circ TW.$$

If we let $\delta S(s,v_0)e$ be the solution of the O.D.E. (ordinary differentiation) on TM

$$dv/ds=\delta X(v(s),e),\quad v(0)=v_0,$$

we see

$$\delta S(s,v_0)e=T[S(s,-)e]v_0 \tag{47}$$

considering $S(s,-)e$ as a diffeomorphism of M.

In essentially the same way, using the analysis in Gikman and Skorohod,[12] Chapter 1, the solutions of $(\delta X,\delta W)$ on TM can be considered as derivatives in probability with respect to the initial conditions of solutions of (X,W) on M.

B. Brownian motion, geodesics, and the Laplace-Beltrami operator

Now we return to the situation and notation of Chapter 1, so $M=O(N)$ and X is given by the Levi-Civita connection. We shall first identify the deterministic flows $S^i,\delta S^i$ on $O(N)$, $TO(N)$ together with their projections onto N and $TO(N)$. For $a\in O(N)$ set $\gamma^i(s,a)=\pi\circ S^i(s,a)$. Then, immediately from the fact that $\pi\circ X(a,e_i)=ae_i$:

Lemma 2B: The curves $\gamma^i(-,a)$ are geodesics in N with

$$\frac{d\gamma^i}{ds}(0,a)=ae_i.$$

The following theorem identifies the infinitesimal generators of the processes x^μ and y^μ, in particular showing that x^1 is a "Brownian motion" on N. It is needed to verify the expression (6) for the solution g^μ of our basic partial differential equation.

Proposition 2B: Let $f:N\to R$ be C^2, then almost surely

$$f(x^\mu(t))=f(x_0)+\int_0^t df(x^\mu(s))u^\mu(s)\,d\omega(s)+\tfrac{1}{2}\mu^2\int_0^t\Delta f(x^\mu(s))\,ds$$

and

$$f(y^\mu(t))=f(x_0)+\int_0^t df(y^\mu(s))v^\mu(s)\,d\omega(s)+\tfrac{1}{2}\mu^2\int_0^t\Delta f(y^\mu(s))\,ds$$

$$+\tfrac{1}{2}\int_0^t df(y^\mu(s))v^\mu(s)\dot\sigma(s)\,ds.$$

Proof: We have to apply Itô's formula (46) to $f\circ\pi:O(N)\to R$ acting on u^μ and v^μ.

By the Lemma

$$\frac{d}{dr}f\circ\pi\circ S^i(r,a)\big|_{r=0}=ae_i,\quad a\in O(N),$$

and, at $r=0$,

$$\frac{d^2}{dr^2}f\circ\pi\circ S^i(r,a)=\frac{d^2}{dr^2}[f\circ\gamma^i(r,a)]$$

$$=\frac{d}{dr}[df(\gamma^i(r,a))\dot\gamma(r,a)]$$

$$=\nabla df(\pi a)(ae_i,ae_i).$$

Since

$$\Delta f(m)=\sum_{i=1}^n\nabla df(m)(ae_i,ae_i),\quad m=\pi a,$$

the result follows.

C. Brownian motion and Jacobi fields

Recall that a vector field J along along a geodesic γ in N is a *Jacobi field* if it satisfies the differential equation

$$D^2J/dt^2 + R[J,\dot{\gamma}]\dot{\gamma} = 0.$$

For $a \in O(N)$ and $l \in T_aO(N)$ we have the field $\delta S^i(-,l)$ along $S^i(-,l)$, $i=1,\ldots,n$. Set

$$J^i(s,l) = T\pi \circ \delta S^i(s,l)$$

so $J^i(-,l)$ is a field along $\gamma^i(-,a)$.

Let $\tilde{\omega}$ denote the connection form of N; it is a smooth 1-form on $O(N)$ with values in the Lie algebra $o(N)$ of $O(N)$, considered as the space of skew adjoint linear endomorphisms of R^n.

Lemma 2C: $J^i(-,l)$ is a Jacobi field along $\gamma^i(-,a)$ with $J^i(\theta,l) = T\pi(l)$ and

$$\frac{D}{\partial s}J^i(s,l)\Big|_{s=0} = a\tilde{\omega}(l)e_i.$$

Proof: Let hl denote the horizontal component of l and consider the path g in $O(n)$ given by

$$g(s) = \exp s\tilde{\omega}(l).$$

Let ϕ be a horizontal path in $O(N)$ with $\phi(o) = a$ and $\phi'(o) = hl$. Then using the right action of $O(n)$ on $O(N)$

$$\frac{d}{dr}[\phi(r)\cdot g(r)]\Big|_{r=0} = l.$$

Consequently, using (47),

$$J^i(s,l) = T\pi \circ \delta S^i(s,l)$$
$$= \frac{\partial}{\partial r}\gamma^i(s,\phi(r)\cdot g(r))\Big|_{\mu=0}.$$

The standard computation shows now that $J^i(-,l)$ is a Jacobi field (see for example,[13] Lemma 14.3, but remember the different sign convention of the curvature tensor). For the initial conditions, it is clear that $J^i(0,l) = T\pi(l)$, while at $s=0, r=0$

$$\frac{D}{\partial s}J^i(s,l) = \frac{D}{\partial s}\frac{\partial}{\partial r}\gamma^i(s,\phi(r)\cdot g(s))$$
$$= \frac{D}{\partial r}\frac{\partial}{\partial s}\gamma^i(s,\phi(r)\cdot g(s))$$
$$= \frac{D}{\partial r}[\phi(r) \circ g(r)e_i]$$
$$= \phi(0)\frac{d}{dr}g(r)\,e_i$$
$$= a\tilde{\omega}(l)e_i,$$

as required.

D. Formulas for δy

We can now obtain equations for δy^μ. Recall that δy^μ is the projection by $T\pi: TO(N) \to TN$ of the $TO(N)$-valued process δv^μ obtained from v^μ by differentiating with respect to μ. We see δv^μ is given by the time-dependent equation

$$dl = \mu\,\delta X d\omega + \delta X\dot{\sigma}(t)dt + \check{X}d\omega(s), \tag{48}$$

where for $l \in T_aO(N)$ and $e \in R^n$, $\check{X}(l)e$ is the vertical vector in $T_lTO(N)$ corresponding to $X(a,e), a \in O(N)$.

We will use the parallelization of $TO(N)$:

$$TO(N) \to O(N) \times \mathbb{R}^n \times O(n),$$
$$l \to (a, \theta(l), \tilde{\omega}(l)), \quad l \in T_aO(N),$$

where θ is the fundamental 1-form

$$\theta: TO(N) \to \mathbf{R}^n,$$
$$\theta(l) = a^{-1}T\pi(l).$$

Let $\underline{\Omega}$ denote the curvature form of the connection: so $\underline{\Omega}$ is the $\theta(n)$-valued 2-form on $O(N)$ defined by

$$\underline{\Omega}(l_1,l_2) = d\tilde{\omega}(hl_1,hl_2),$$

where h is the orthogonal projection. Consider the "principal part"

$$[\delta X^i]^\frown: O(N)\times\mathbb{R}^n\times o(n) \to TO(N)\times\mathbb{R}^n\times o(n), i=1,2,\ldots$$

of the vector field induced from δX^i by our parallelizing diffeomorphism. We have

Proposition 2D: For $(a,\xi,A) \in O(N)\times R^n\times o(n)$ and $i=1,\ldots,n$

$$[\delta X^i]^\frown(a,\xi,A) = (X^i(a), Ae_i, 2\underline{\Omega}(X^i(a),X(a)\xi)).$$

Proof: Set $F = R^n \times o(n)$ and for $b \in O(N)$ define $\psi(b): T_bO(N) \to F$ by

$$\psi(b)(k) = (\theta(k), \tilde{\omega}(k)).$$

Set

$$l = \psi(a)^{-1}(\xi, A),$$

so

$$\xi = \theta(l) \text{ and } A = \tilde{\omega}(l).$$

We will work over a chart (U,ψ) for $O(N)$ about α, considering U as an open subset of F. In this representation ψ is a map:

$$\psi: U \to L(F;\mathbb{R}^n\times o(n))$$

and

$$\delta X^i: U\times F \to U\times F\times F\times F$$

is given by

$$\delta X^i(a,l) = (a, l, X_0^i(a), DX_0^i(a)l),$$

where

$$X_0^i: U \to F$$

is the principal part of X^i. Consequently

$$[\delta X^i]^\frown(a,\xi,A) = \Big(X^i(a), D\psi(a)(X_0^i(a))l + \psi(a)DX_0^i(a)l\Big).$$

Now in terms of the exterior derivative $d\psi$ of ψ

$$D\psi(a)(X_0^i(a))l = 2d\psi(X^i(a),l) + D\psi(a)(l)X_0^i(a);$$

also

$$D\psi(a)(l)X_0^i(a) = -\psi(a)DX_0^i(a)l,$$

because

$$\psi(b)X_0^i(b) = (e_i, o), b \in U.$$

Therefore

$$[\delta X^i]^\frown(a,\xi,A) = (X_0^i(a), 2d\psi(X^i(a),l)). \tag{49}$$

Next we apply the structure equations of the connection: for $l_1, l_2 \in T_aO(N)$

$$d\tilde{\omega}(l_1, l_2) = -\tfrac{1}{2}[\tilde{\omega}(l_1), \tilde{\omega}(l_2)] + \underline{\Omega}(l_1, l_2),$$

$$d\theta(l_1, l_2) = \tfrac{1}{2}(\tilde{\omega}(l_2)\theta(l_1) - \tilde{\omega}(l_1)\theta(l_2)),$$

giving

$$2d\tilde{\omega}(X^i(a), l) = 2\underline{\Omega}(X^i(a), l) = 2\underline{\Omega}(X^i(a), X^i(a)\xi)$$

[since $hl = X(a)\xi$] and

$$2d\theta(X^i(a), l) = A e_i$$

which we can substitute in (49) to complete the proof.

Now let $[\check{X}^i]^\wedge : O(N)\times R^n \times o(n) \to TO(N)\times R^n \times o(n)$ be the "principal part" of the vector field corresponding to $\check{X}(-)e_i$.

Lemma 2D(*i*): *For* $(a, \xi, A) \subset O(N)\times R^n \times O(n)$ *and* $i = 1, \ldots, n$,

$$[\check{X}^i]^\wedge(a, \xi, A) = (O_a, e_i, O),$$

where $O_a \in T_aO(N)$ is the zero vector.

Proof: By the defining property of $\check{X}^i$ we have

$$[\check{X}^i]^\wedge(a, \xi, A) = (O_a, \theta X^i(a), \tilde{\omega}X^i(a)).$$

But

$$\theta X^i(a) = e_i \text{ and } \tilde{\omega}X^i(a) = 0, \quad a \in O(N), //.$$

Lemma 2D(*ii*): *If* $f: M \to N$ is a C^2-map and for some $\alpha, \beta \in T_mM, \gamma \in T_\beta TM$ is the vertical vector corresponding to α, then $TTf(\gamma)$ is the vertical vector in TTN corresponding to $Tf(\alpha)$

Proof: Immediate by taking charts. //

Let $K: TN \oplus TN \to R$ denote the *Ricci curvature* tensor of N:

$$K(v_1, v_2) = \mathrm{tr}[v \to R(v, v_1)v_2].$$

Theorem 2D: For $u \geq 0$ set

$$\xi^\mu(s) = \theta\circ\delta v^\mu(s) : \Omega \to \mathbb{R}^\mu$$

and

$$A^\mu(s) = \tilde{\omega}\circ\delta v^\mu(s) : \Omega \to o(n), \quad 0 \leq s \leq t.$$

Then the processes ξ^μ and A^μ satisfy the following equations:

$$\xi^\mu(s) = \omega(s) + \mu\int_0^s A^\mu(r)\,d\omega(r) + \int_0^s A^\mu(r)\sigma(r)\,dr$$
$$- 2^{-1}\mu^2\int_0^s V^\mu(r)^{-1}\circ K(\delta y^\mu(r), -)^{\#}\,dr, \tag{50}$$

$$A^\mu(s) = \mu\int_0^s v^\mu(r)^{-1}\circ R[v^\mu(r)d\omega(r), \delta y^\mu(r)]\circ v^\mu(r)\,dr$$
$$+ \int_0^s v^\mu(r)^{-1}\circ R[v^\mu(r)\dot{\sigma}(r), \delta y^\mu(r)]\circ v^\mu(r)\,dr$$
$$+ 2^{-1}\mu^2\int_0^s v^\mu(r)^{-1}\circ\sum_{i=1}^n R[v^\mu(r)e_i, v^\mu(r)A^\mu(r)e_i]\circ v^\mu(r)\,dr$$
$$+ 2^{-1}\mu^2\int_0^s v^\mu(r)^{-1}\circ L(\delta y^\mu(r))\circ v^\mu(r)\,dr, \tag{51}$$

where, if $\delta \in T_{y_0}N$ some $y_0 \in N$,

$$K(\delta, -)^{\#} \in T_{y_0}N$$

and

$$L(\delta): T_{y_0}N \to T_{y_0}N$$

are defined by

$$\langle K(\delta, -)^{\#}, v_1\rangle = K(\delta, v_1)$$

and

$$\langle L(\delta)v_1, v_2\rangle = \nabla K(v_2)(v_1, \delta) - \nabla K(v_1)(v_2, \delta), v_1, v_2 \in T_{y_0}N.$$

Proof: In order to apply the Itô formula as given above (46) let $S^{\mu i}(-, a)$ and $\delta S^{\mu i}(-, l)$ be the solution curves starting from $a \in O(N)$ and $l \in T_aO(N)$ respectively of the vector fields μX^i and $\mu\,\delta X^i + X(-)e_i$. By Lemma 2D(i) and Proposition 2D

$$\frac{d}{ds}[\theta\circ\delta S^{\mu i}(s, l)] = \mu\tilde{\omega}(\delta S^{\mu i}(s, l))e_i + e_i$$

and

$$\frac{d}{ds}[\tilde{\omega}\circ\delta S^{\mu i}(s, l)] = 2\mu\underline{\Omega}(X^i(S^{\mu i}(s, a)), \delta S^{\mu i}(s, l))$$
$$= \mu S^{\mu i}(s, a)^{-1}\circ R(S^{\mu i}(s, a)e_i,$$
$$T\pi\circ S^{\mu i}(s, l))\circ S^{\mu i}(s, a)$$

[recall that R corresponds to 2Ω]. Thus at $s = 0$

$$\frac{d^2}{ds^2}[\theta\circ\delta S^{\mu i}(s, l)] = \mu^2 a^{-1}\circ R[ae_i, T\pi(l)]\; ae_i$$

and

$$\frac{d^2}{ds^2}[\tilde{\omega}\circ\delta S^{\mu i}(s, l)] = \mu a^{-1}\circ\frac{D}{\partial s}R[S^{\mu i}(s, a)e_i, T\pi\circ\delta S^{\mu i}(s, l)]\big|_{s=0}\circ a$$
$$= \mu a^{-1}\circ\nabla R(\mu a e_i)(ae_i, T\pi(l))\circ a$$
$$+ \mu a^{-1}\circ R[ae_i, a\tilde{\omega}(l)(\mu e_i)]\circ a,$$

since $(D/ds)S^{\mu i}(s, a)e_i = 0$ because $S^{\mu i}(s, a)e_i$ is horizontal and since $(D/\partial s)T\pi\circ\delta S^{\mu i}(s, l)|_{s=0} = a\tilde{\omega}(l)(\mu e_i) + ae_i$ which can be seen using Lemma 2D(ii) to obtain the term ae_i from $\check{X}$ and Lemma 2C for the other term. The Itô formula therefore gives

$$\theta\circ\delta v^\mu(s) = \int_0^s d\omega(r) + \mu\int_0^s \tilde{\omega}(\delta v^\mu(r))\,d\omega(r)$$
$$+ \int_0^s \tilde{\omega}(\delta v^\mu(r))\dot{\sigma}(r)\,dr$$
$$+ 2^{-1}\mu^2\sum_{i=1}^n\int_0^s v^\mu(r)^{-1}$$
$$\circ R[v^\mu(r)e_i, \delta y^\mu(r)]\circ v^\mu(r)e_i\,dr$$

and

$$\tilde{\omega}\circ\delta v^\mu(s) = \mu\int_0^s v^\mu(r)^{-1}\circ R[v^\mu(r)d\omega(r), \delta y^\mu(r)]\circ v^\mu(r)\,dr$$
$$+ 2^{-1}\mu^2\int_0^s v^\mu(r)^{-1}\circ\sum_{i=1}^n \nabla R(v^\mu(r)e_i)$$
$$\times(v^\mu(r)e_i, \delta y^\mu(r))\circ v^\mu(r)\,dr$$
$$+ 2^{-1}\mu^2\int_0^s v^\mu(r)$$
$$\circ\sum_{i=1}^n R[v^\mu(r)e_i, v^\mu(r)\tilde{\omega}(\delta v^\mu(r)e_i)]\circ v^\mu(r)\,dr$$
$$+ \int_0^s v^\mu(r)^{-1}\circ R[v^\mu(r)\dot{\sigma}(r), \delta y^\mu(r)]\circ v^\mu(r)\,dr.$$

Finally, writing $f_i = v^\mu(r)e_i, i=1,\ldots,n, \delta=\delta y^\mu(r)$, and taking $v_1, v_2 \in T_y\mu_{(r)}$, the second Bianchi identity gives

$$\begin{aligned}\nabla K(v_1)(v_2,\delta) &= \sum_i \langle \nabla R(v_1)(f_i,v_2)\delta, f_i\rangle \\ &= -\sum_i \langle \nabla R(v_2)(v_1,f_1)\delta, f_i\rangle \\ &\quad -\sum_i \langle \nabla R(f_i)(v_2,v_1)\delta, f_i\rangle \\ &= \nabla K(v_2)(v_1,\delta) + \sum_i \langle \nabla R(f_i)(f_i,\delta)v_2, v_1\rangle .\end{aligned}$$

Thus $\sum_{i=1}^n \nabla R(f_i)(f_i,\delta) = L(\delta)$.//

Equations (50) and (51) for ξ^μ, A^μ can be differentiated in μ to give equations for the derivatives in probability with respect to μ, $\delta\xi^\mu$ and δA^μ, of ξ^μ and A^μ. In fact the process can be iterated to obtain arbitrarily high derivatives; although the resulting lengthy equations do not seem very illuminating, at least for a general manifold N. However the following formula will be useful:

$$\delta\xi^0(s) = \int_0^s A^0(r)\,d\omega(r) + \int_0^s \delta A^0(r)\dot\sigma(r)\,dr. \tag{52}$$

E. Formulas for $\delta^2 y$

To relate $\delta^2 y$ to $\delta\xi^0$ we need another lemma:

Lemma 2E: If $\alpha:(a,b)\to O(N)$ is differentiable and $\zeta\in R^n$ then

$$\frac{D}{\partial t}\alpha(t)\zeta = \alpha(t)\tilde\omega(\dot\alpha(t))\zeta .$$

Proof: Write $\alpha(t)=\beta(t)\cdot\phi(t)$ for

$$\phi:(a,b)\to o(n)$$

and β horizontal

$$\beta:(a,b)\to O(N).$$

Then

$$\dot\alpha(t) = \dot\beta(t)\cdot\phi(t) + \beta(t)\cdot\dot\phi(t)$$

and so, as in Kobayashi and Nomizu (10) page 69

$$\tilde\omega(\dot\alpha(t)) = 0 + \phi(t)^{-1}\dot\phi(t).$$

But

$$\begin{aligned}\frac{D}{\partial t}\alpha(t)\zeta &= \beta(t)\frac{d}{dt}(\beta(t)^{-1}\alpha(t)\zeta) \\ &= \beta(t)\frac{d}{dt}\phi(t). // \\ &= (\beta(t)\cdot\phi(t))\phi(t)^{-1}\dot\phi(t). //\end{aligned}$$

In particular this gives, for $\zeta\in R^n$, differentiating for μ rather than t:

$$\frac{D}{\partial\mu}v^\mu(s)\zeta = v^\mu(s)A^\mu(s)\zeta . \tag{53}$$

From this and (52) we have immediately:

Proposition 2E:

$$\delta^2 y(s) = v^0(s)\Big\{A^0(s)\xi^0(s) + \int_0^s A^0(r)\,d\omega(r) + \int_0^s \delta A^0(r)\dot\sigma(r)\,dr\Big\}. // \tag{54}$$

F. The formulas of section 1

We can now confirm the formulas of Sec. 1. First by (50) and (51) we have

$$\begin{aligned}\frac{d}{ds}(\xi^0(s) - \omega(s)) &= A^0(s)\dot\sigma(s) \\ &= \int_0^s v^0(r)^{-1}\circ R[Z(r), \delta y^0(r)]v^0(r)\dot\sigma(s)\,dr ,\end{aligned} \tag{55}$$

giving (29).

Next, to obtain (21), note that by (54) and the skew-adjointness of $\delta A^0(r)$:

$$\begin{aligned}&\int_0^s \langle \delta^2 y(r) - v^0(r)A^0(r)\xi^0(r), \ddot Z(r)\rangle\,dr \\ &= \langle \delta^2 y(s) - v^0(s)A^0(s)\xi^0(s), \dot Z(s)\rangle \\ &\quad - \int_0^s \langle A^0(r)\,d\omega(r), \dot\sigma(r)\rangle \\ &\quad - \int_0^s \langle \delta A^0(r)\dot\sigma(r), \dot\sigma(r)\rangle\,dr \\ &= \langle \delta^2 y(s), \dot Z(s)\rangle - \langle A^0(s)\xi^0(s), \dot\sigma(s)\rangle \\ &\quad + \int_0^s \langle d\omega(r), A^0(r)\dot\sigma(r)\rangle\,dr ,\end{aligned}$$

hence

$$\begin{aligned}&\langle \delta^2 y(t), \dot Z(t) + \int_0^t \langle \delta^2 y(s), \nabla V(Z(s))\rangle\,ds \\ &= -\int_0^t \langle A^0(s)\xi^0(s), \ddot\sigma(s)\rangle\,ds + \langle A^0(t)\xi^0(t), \dot\sigma(t)\rangle \\ &\quad - \int_0^t \langle d\omega(s), A^0(s)\dot\sigma(s)\rangle\,ds ,\end{aligned}$$

however

$$\begin{aligned}&\int_0^t \langle A^0(s)\xi^0(s), \ddot\sigma(s)\rangle\,ds \\ &= \int_0^t \langle A^0(s)(\xi^0(s) - \omega(s)), \ddot\sigma(s)\rangle\,ds \\ &\quad + \int_0^t \langle A^0(s)\omega(s), \ddot\sigma(s)\rangle\,ds .\end{aligned}$$

Therefore

$$\begin{aligned}&\int_0^t \langle A^0(s)\xi^0(s), \ddot\sigma(s)\rangle\,ds = \langle A^0(t)(\xi^0(t) - \omega(t)), \dot\sigma(t)\rangle \\ &\quad - \int_0^t \langle \dot A^0(s)(\xi^0(s) - \omega(s)), \dot\sigma(s)\rangle\,ds \\ &\quad + \int_0^t \langle \frac{d}{ds}(\xi^0(s) - \omega(s)), A^0(s)\dot\sigma(s)\rangle ds + \langle A^0(t)\omega(t), \dot\sigma(t)\rangle \\ &\quad - \int_0^t \langle \dot A^0(s)\omega(s), \dot\sigma(s)\rangle\,ds - \int_0^t \langle A^0(s)d\omega(s), \dot\sigma(s)\rangle\,ds .\end{aligned}$$

Formula (37) follows using (55) and (51).

3. CLASSICAL PATHS AND FLOWS; THEIR RELEVANCE FOR THE DIFFUSION (HEAT) EQUATION

In this section we discuss a theorem due to Varadhan[7] and its applications to the quasiclassical expansion. This leads naturally to a discussion of when the classical flow induces a diffeomorphism of the configuration-space manifold, a question which we partly answer in Theorem 3J. These results will be used in the last section of the paper.

A. Varadhan's theorems

We now give the argument used by Varadhan. His argument dovetails very nicely with the results already obtained from the quasiclassical expansion. For the Schrödinger equation this "Laplace method" would have to be replaced by some method of "stationary phase". Varadhan's method is based on his more abstract results, particularly on Theorem 3.1 of Ref. 7. We first quote a special case of this:

Theorem 3A: Let Γ be a regular topological space with a sequence $\{P_n\}_{n=1,2,\ldots}$ of Borel probability measures on Γ, $\{a_n\}_{n=1,2,\ldots}$ a positive real sequence and $I:\Gamma \to \mathbb{R}^+\cup\{+\infty\}$ a map satisfying

(i) $a_n \to \infty$ as $n\to\infty$,

(ii) for all closed subsets C of Γ

$$\overline{\lim_{n\to\infty}}\, a_n^{-1}\ln[P_n(C)] \leq -\inf_{z\in C} I(z),$$

(iii) for all open subsets U of Γ

$$\overline{\lim_{n\to\infty}}\, a_n^{-1}\ln[P_n(U)] \geq -\inf_{z\in U} I(z),$$

together with

I(i) I is lower semicontinuous on Γ i.e. if $z_0\in\Gamma$

$$\underline{\lim_{z\to z_0}}\, I(z) \geq I(z_0),$$

I(ii) for each $m\in\mathbb{R}$, $\{z\in\Gamma\,|\,I(z)\leq m\}$ is compact in Γ. Suppose now $F:\Gamma\to\mathbb{R}\cup\{-\infty\}$ is bounded above and upper semicontinuous. Then

$$\overline{\lim_{n\to\infty}}\, a_n^{-1}\ln\left\{\int_\Gamma \exp[a_n F(z)]\,dP_n(z)\right\} \leq \sup_{z\in\Gamma}[F(z)-I(z)].$$

For our situation we take $\Gamma = C_{x_0}(N)$, the space of continuous paths $z:[0,t],\ 0\to N, x_0$. Define $\{P^\mu\}_{\mu>0}$ to be the measures on Γ determined by $\{x^\mu\}_{\mu>0}$, $P^\mu = x^\mu(\gamma)$; and define $I:\Gamma\to\mathbb{R}^+\cup\{\infty\}$ by

$$I(z) = 2^{-1}\int_0^t |\dot z(s)|^2\,ds,\quad z\in L^{2,1}([0,t];N)\cap\Gamma,$$

$$=\infty,\ \text{otherwise}.$$

The sequence $\{a_\mu\}_{\mu>0}$ is then taken to be $a_\mu = \mu^{-2}$, $\mu>0$. Conditions I(i) and I(ii) are well known to hold.

For our application of the theorem we need to verify (ii) and (iii) in the form:

$$\text{(ii)}'\ \overline{\lim_{\mu\to0}}\,\mu^2\ln[P^\mu(C)] \leq -\inf_{z\in C} I(z).$$

$$\text{(iii)}'\ \underline{\lim_{\mu\to0}}\,\mu^2\ln[P^\mu(U)] \geq -\inf_{z\in U} I(z),$$

for all open subsets U and closed subsets C of Γ. For N complete these can be found in the Appendix to Molchanov's paper.[8] Then we deduce:

Corollary 3A: If $F: C_{x_0}(N)\to\mathbb{R}\cup\{-\infty\}$ is bounded above and upper semicontinuous, then

$$\overline{\lim_{\mu\to0}}\,\mu^2\ln\left\{\int_\Omega \exp[\mu^{-2}F(x^\mu)]\,d\gamma\right\} \leq \sup_{z\in C_{x_0}(N)}[F(z)-I(z)].$$

Definition: For $S_0:N\to\mathbb{R}$ and $V:N\to\mathbb{R}$, for the moment assumed only to be measurable, define the "action" functional $\mathscr{S}: C_{x_0}(N)\to\mathbb{R}\cup\{\infty\}$ by

$$\mathscr{S}(z) = S_0[z(t)] + I(z) - \int_0^t V[z(s)]\,ds.$$

B. A result from the calculus of variations

We now need a standard sort of result from the calculus of variations:

Lemma 3B: Assume S_0, V are lower and upper semicontinuous respectively. Then $\mathscr{S}$ is lower semicontinuous on $C_{x_0}(N)$. Moreover, if V is bounded above, and $\{z_i\}_{i=1,2,\ldots}$ is a sequence in $C_{x_0}(N)$ such that

(i) $\{S_0[z_i(t)]\}_{i=1,2,\ldots}$ is bounded below,

(ii) $\{\mathscr{S}(z_i)\}_{i=1,2,\ldots}$ is bounded above,

then $\{z_i\}_{i=1,2,\ldots}$ has a convergent subsequence in $C_{x_0}(N)$.

Proof: The lower semicontinuity of $\mathscr{S}$ comes immediately from that of I. Also, if S_0, V and $\{z_i\}_{i=1,2,\ldots}$ behave as above then $\{I(z_i)\}_{i=1,2,\ldots}$ is bounded above and the subsubconvergence of $\{z_i\}_{i=1,2,\ldots}$ follows from I(ii).//

C. Convergence to the classical paths

We now give the main result of this section:

Theorem 3C: Suppose S_0, V, T_0 are measurable real-valued functions on N and $Z:[0,t], 0\to N, x_0$ is continuous. Assume

(i) Z is an absolute minimum for $\mathscr{S}$ on $C_{x_0}(N)$,

(ii) if $\tilde Z$ is an absolute minimum for $\mathscr{S}$ then either $\tilde Z(t)\notin \operatorname{supp} T_0$ or $\tilde Z(t) = Z(t)$,

(iii) V is upper semicontinuous and bounded above on N,

(iv) S_0 is lower semicontinuous on N and bounded below on supp T_0.

For $\delta>0$, let $\Gamma(Z,\delta) = \{z\in C_{x_0}(N)\,|\,\sup_{0\leq s\leq t} d(z(s),Z(s)) \geq \delta\}$, where d is a metric on N. Then, for the measure $P^\mu = x^\mu(\gamma)$, on $C_{x_0}(N)$,

$$\lim_{\mu\to0}\int_{\Gamma(Z,\delta)} \exp\left\{\mu^{-2}\left[\mathscr{S}(Z) - S_0[z(t)] + \int_0^t V[z(s)]\,ds\right]\right\}$$

$$\times\, T_0[z(t)]\,dP^\mu(z) = 0.$$

Proof: Let $J(\mu)$ denote the integral in question and set

$$\Gamma_0 = \{z\in\Gamma(Z,\delta)\,|\,z(t)\in\operatorname{supp}T_0\}.$$

Following Varadhan's proof of Theorem 3.5 in Ref. 7, define $F:\Gamma\to\mathbb{R}\cup\{-\infty\}$ by

$$F(z) = \mathscr{S}(Z) - S_0[z(t)] + \int_0^t V[z(s)]\,ds,\quad z\in\Gamma_0,$$

$$= -\infty\ \text{otherwise}.$$

By the hypotheses F is bounded above and upper semicontinuous on $C_{x_0}(N)$ since Γ_0 is closed. Applying Corollary 3A

$$\overline{\lim_{\mu\downarrow 0}}\mu^2 \ln J(\mu) \leq \sup_{z\in C_{x_0}(N)} [F(z)-I(z)] = \sup_{z\in\Gamma_0}[\mathscr{S}(Z)-\mathscr{S}(z)] = k\ ,$$

for some $k \leq 0$, by hypothesis (i).

Suppose $k=0$. Then there exists $\{z_i\}_{i=1,2,\ldots}$ in Γ_0 with $\mathscr{S}(z_i)\to\mathscr{S}(Z)$. By the Lemma we can assume $\{z_i\}_{i=1,2,\ldots}$ converges in $C_{x_0}(N)$ to some path z and then by lower semicontinuity

$$\mathscr{S}(z) \leq \lim_{i\to\infty}\mathscr{S}(z_i) = \mathscr{S}(Z)\ .$$

But $z\in\Gamma_0$, so that $z(t)\in\operatorname{supp} T_0$ and we have a contradiction to hypothesis (ii). Hence, $k<0$ and

$$\lim_{\mu\downarrow 0}\ln J(\mu) = -\infty\ ,$$

proving the theorem. //

[For an introduction to Donsker–Varadhan theory see Ref. 14.]

D. The minimum of the action functional $\mathscr{S}$

We reimpose the assumption that S_0 and V are C^2. Let $\mathscr{P}$ denote the Hilbert manifold of $L^{2,1}$ paths on N starting at x_0

$$\mathscr{P} = L^{2,1}_{x_0}([0,t];N)\ .$$

This is a Hilbert manifold. The tangent space $T_z\mathscr{P}$ at a point z of $\mathscr{P}$ can be identified with the space of $L^{2,1}$ vector fields ξ along z vanishing at 0: that is those $\xi:[0,t]\to TN$ such that $\xi(s)\in T_{z(s)}N$ and $\xi(0)=0$.

There is a natural inner product on $T_z\mathscr{P}$:

$$\langle\xi_1,\xi_2\rangle_z = \int_0^t \langle\dot{\xi}_1(s),\dot{\xi}_2(s)\rangle ds\ .$$

We shall consider $\mathscr{S}$ restricted to $\mathscr{P}$, $\mathscr{S}:\mathscr{P}\to\mathbb{R}$,

$$\mathscr{S}(z) = S_0[z(t)] + 2^{-1}\int_0^t |\dot{z}(s)|^2 ds - \int_0^t V[z(s)]ds\ .$$

Proposition 3D: A path Z in $\mathscr{P}$ is a critical point for $\mathscr{S}$ iff Z is C^2 and satisfies

$$\frac{D^2 Z(s)}{\partial s^2} = -\nabla V[Z(s)]\ ,\quad s\in[0,t]\ ,$$

with

$$\dot{Z}(t) = -\nabla S_0[Z(t)]\ .$$

The path is a nondegenerate minimum iff also the bilinear form $\mathscr{H}$ on $T_z\mathscr{P}$ is positive definite, where

$$\mathscr{H}(\xi,\eta) = \langle\xi,\eta\rangle_z + \int\langle R[\dot{Z}(s),\xi(s)]\dot{Z}(s),\eta(s)\rangle ds - \int_0^t \nabla dV[Z(s)](\xi(s),\eta(s))ds + \nabla dS_0[Z(t)](\xi(t),\eta(t))\ . \tag{56}$$

Remark: The form $\mathscr{H}$ is the Hessian of $\mathscr{S}$ at Z.

Proof: If $Z\in\mathscr{P}$ and $\eta\in T_Z\mathscr{P}$, then

$$d\mathscr{S}(\eta) = \langle\nabla S_0[Z(t)],\eta(t)\rangle + \int_0^t\langle\dot{Z}(s),\dot{\eta}(s)\rangle ds - \int_0^t\langle\nabla V[Z(s)],\eta(s)\rangle ds\ . \tag{57}$$

Partial integration yields

$$d\mathscr{S}(\eta) = \langle\nabla S_0[Z(t)] - \int_0^t\nabla V[Z(s)]ds,\eta(t)\rangle + \int_0^t\langle\dot{Z}(s) + \int_0^s\nabla V[Z(s')]ds',\dot{\eta}(s)\rangle ds\ .$$

For Z to be a critical point of $\mathscr{S}$ it follows that $(\dot{Z}(-)+\int_0^t\nabla V[Z(s)]ds)$ is orthogonal in $T_Z\mathscr{P}$ to $\{\eta\in T_Z\mathscr{P}\,|\,\eta(t)=0\}$.

This means that it is a ray i.e., differentiable with parallel covariant derivative. Consequently Z is C^2 and a different integration by parts of (57) gives:

$$d\mathscr{S}(\eta) = \langle\nabla S_0[Z(t)]+\dot{Z}(t),\eta(t)\rangle - \int_0^t\langle\ddot{Z}(s)+\nabla V[Z(s)],\eta(s)\rangle ds \tag{58}$$

From this we see $(\ddot{Z}(-)+\nabla V[Z(-)])$ vanishes in L^2 and the characterization of critical points follows.

For the nondegenerate minimum condition we must examine the next derivative of $\mathscr{S}$, the Hessian, at the critical point Z. For $\xi,\eta\in T_Z\mathscr{P}$ set, for $0\leq\alpha\leq 1$,

$$Z_\alpha(s) = \exp_{Z(s)}[\alpha\xi(s)]$$

and

$$\eta_\alpha(s) = \nabla_2\{\exp_{Z(s)}[\alpha\xi(s)]\}(\eta(s))\ ,$$

where ∇_2 refers to differentiations with respect to the second variable [in this case $\alpha\xi(s)$] so that $\nabla_2(\exp_{Z(s)}(\alpha\xi(s))): T_{Z(s)}N\to T_{Z_\alpha(s)}N$. In fact, though not strictly relevant, for each s the map $\alpha\mapsto\eta_\alpha(s)$ is a Jacobi field along $Z_\alpha(s)$ with

$$\eta_\alpha(s) = \eta(s)\ ,\quad \frac{D}{\partial\alpha}\eta_\alpha(s)\Big|_{\alpha=0} = 0\ .$$

See for example Eliasson, Ref. 15, Theorem 3.3, or our discussion in Lemma 2C. Note also that $(\partial/\partial\alpha)Z_\alpha(s) = \nabla_2[\exp(\alpha\xi(s))(\xi(s))]$.

What we are doing is equivalent to taking an exponential chart, as in Ref. 15, for $\mathscr{P}$ about the path Z. We now compute the second derivative of $\mathscr{S}$ in that chart. Assuming temporarily that ξ,η are C^2 and using Eq. (58),

$$\begin{aligned}\mathscr{H}(\xi,\eta) &= \frac{D}{\partial\alpha}[d\mathscr{S}(\eta_\alpha)]\Big|_{\alpha=0}\\ &= \frac{D}{\partial\alpha}\Big\{\langle\nabla S_0(Z_\alpha(t))+\dot{Z}_\alpha(t),\eta_\alpha(t)\rangle - \int_0^t\langle\ddot{Z}_\alpha(s)+\nabla V[Z_\alpha(s)],\eta_\alpha(s)\rangle ds\Big\}\Big|_{\alpha=0}\\ &= \langle\nabla^2 S_0(Z(t))\xi(t)+\dot{\xi}(t),\eta(t)\rangle - \int_0^t\langle\ddot{\xi}(s) - R[\dot{Z}(s),\xi(s)]\dot{Z}(s) + \nabla^2 V[Z(s)]\xi(s),\eta(s)\rangle ds\\ &= \nabla dS_0(Z(t))(\xi(t),\eta(t)) + \langle\xi,\eta\rangle_Z + \int_0^t\langle R[\dot{Z}(s),\xi(s)]\dot{Z}(s),\eta(s)\rangle ds - \int_0^t\nabla dV(Z(s))(\xi(s),\eta(s))ds\ .\end{aligned}$$

Approximating general elements of $T_Z\mathscr{P}$ by C^2 elements we get the required result.

E. Nondegenerate minimum of $\mathscr{L}$

From now on we let Z be a critical point of $\mathscr{S}$ with $\mathscr{H}$ the Hessian at Z. There is a self-adjoint $A : T_Z\mathscr{P} \to T_Z\mathscr{P}$ with $\mathscr{H}(\xi,\eta) = \langle A\xi,\eta\rangle_Z$, all ξ,η in $T_Z\mathscr{P}$, and $\mathscr{H}$ is called *nondegenerate* if A is an isomorphism. However A is easily seen to be a compact perturbation of the identity: so applying the Fredholm theory $\mathscr{H}$ is nondegenerate iff $\mathscr{H}(\xi,\eta)=0$ for all $\eta \in T_Z\mathscr{P}$ implies $\eta = 0$.

Lemma 3E: Nondegeneracy of $\mathscr{H}$ is equivalent to the vanishing identically of ξ whenever ξ is C^2 and satisfies

$$\frac{D^2\xi(s)}{\partial s^2} - R[\dot{Z}(s), \xi(s)]\dot{Z}(s) + \nabla^2 V[Z(s)]\xi(s) = 0, \quad s \in [0,t], \tag{59}$$

with

$$\xi(0) = 0 \text{ and } \dot{\xi}(t) = -\nabla^2 S_0(Z(t))\,\xi(t)\ . \tag{60}$$

Proof: On integrating by parts Eq. (56), for $\xi,\eta \in T_Z\mathscr{P}$, we have

$$\begin{aligned}\mathscr{H}(\xi,\eta) = &\langle\{\nabla^2 S_0(Z(t))\xi(t), \eta(t)\rangle + \Big\langle \int_0^t \{R[\dot{Z}(s), \xi(s)]\dot{Z}(s)\\ &+ \nabla^2 V[Z(s)]\xi(s)\}ds, \eta(t)\\ &+ \int_0^t \Big\langle \dot{\xi}(s) - \int_0^s \{R[\dot{Z}(s'), \xi(s')]\dot{Z}(s')\\ &+ \nabla^2 V[Z(s')]\xi(s')\}ds', \dot{\eta}(s)\Big\rangle ds\,.\end{aligned}$$

The lemma is then proved by using the argument of the previous proof to observe that if $\mathscr{H}(\xi,\eta)=0$, for all $\eta \in T_Z\mathscr{P}$ then ξ is C^2 and satisfies the given boundary value problem.//

We can apply the Morse index theory to characterize the nondegenerate minima of $\mathscr{S}$.

Proposition 3E: Assume also that V is C^3. The Hessian $\mathscr{H}$ is positive definite on $T_Z\mathscr{P}$ iff whenever ξ is a C^2 vector field along Z satisfying, for some $0 \le s_0 \le t$,

$$\frac{D^2\xi(s)}{\partial s^2} - R[\dot{Z}(s), \xi(s)]\dot{Z}(s) + \nabla^2 V[Z(s)]\xi(s) = 0,$$

$$s_0 \le s \le t\,,$$

$$\dot{\xi}(t) = \nabla^2 S_0(Z(t))\xi(t)\,, \quad \xi(s_0) = 0$$

then

$$\xi(s) \equiv 0, \quad s_0 \le s \le t\,.$$

Proof: Let $V[0,t]$ denote the Banach space of C^2 vector fields along Z with C^2 norm. In the terminology of Ref. 16 the family of forms $\Omega(s)$ on $V[0,t]$ given by

$$\Omega(s)(\xi,\eta) = \langle\xi(s), \dot{\eta}(s)\rangle + \langle R[\dot{Z}(s), \dot{\xi}(s)]\dot{Z}(s), \eta(s)\rangle$$
$$- \nabla dV[Z(s)](\xi(s), \eta(s)), s \in [0,t]\,,$$

is a Sturm form on $V[O,t]$ of order 1. (The assumption that V is C^3 is needed here.) Set $V_0[0,t] = \{\xi \in V[0,t] \mid \xi(0) = 0\}$. It follows from Theorem 3.1 of Ref. 16 and the previous lemma that $\mathscr{H}$ has zero index on $V_0[0,t]$ iff the vanishing of η occurs as described.

Now suppose that these two equivalent conditions hold. If there existed $\xi \in T_Z\mathscr{P}$ with $\mathscr{H}(\xi,\xi) < 0$, since $V_0[0,t]$ is dense in $T_Z\mathscr{P}$ we could find $\bar{\xi} \in V_0[0,t]$ with $\mathscr{H}(\xi,\xi) < 0$, which is impossible. Therefore, $\mathscr{H}$ has zero index on $T_Z\mathscr{P}$. But by the lemma it is nondegenerate. Consequently it is positive definite on $T_Z\mathscr{P}$ as required.//

F. Classical flows and diffeomorphisms

Let $x[a,A,s]$ denote the solution of the classical equations of motion

$$\frac{D^2 x(s)}{\partial s^2} = -\nabla V[x(s)]$$

with $x[a,A,0] = a \in N$, $\dot{x}[a,A,0] = A \in T_a N$. We assume $x[a,A,s]$ exists for $s \in [0,t]$. Define $\Phi_s : N \to N$ by

$$\Phi_s(a) = x[a, \nabla S_0(a), s]\,.$$

First we shall investigate the derivative of Φ_s. For $A \in T_a N$ and $q:(-1,1) \to N$ with $q(0) = a$, $\dot{q}(0) = A$,

$$\frac{D^2}{\partial s^2}\frac{\partial}{\partial s'}\Phi_s(q(s')) = \frac{D}{\partial s'}\frac{D^2}{\partial s^2}\Phi_s(q(s'))\,,$$
$$+R\left[\frac{\partial}{\partial s}\Phi_s(q(s'))\,,\ \frac{\partial}{\partial s'}\Phi_s(q(s'))\right]\frac{\partial}{\partial s}\Phi_s(q(s'))\,.$$

Taking $s=0$, gives

$$\frac{D^2}{\partial s^2}T_a\Phi_s(A) + \nabla^2 V(T_a\Phi_s(A)) - R[\dot{\Phi}_s(a), T_a\Phi_s(A)]\dot{\Phi}_s(a) = 0\,.$$

There are also the initial conditions

$$T_a\Phi_0(A) = A\,, \quad \frac{D}{\partial s}T_a\Phi_0(A) = \nabla^2 S_0(A)\,.$$

Proposition 3F: The map Φ_{s_0} is a diffeomorphism in a neighborhood of $a \in N$ iff the following condition C_{s_0} holds:

C_{s_0}: The solution K_s with $K_s \in L(T_a N, T_{\Phi_s(a)}N)$ of

$$\frac{D^2}{\partial s^2}K_s(-) - R[\dot{\Phi}_s(a), K_s(-)]\dot{\Phi}_s(a) + \nabla^2 V(K_s(-)) = 0 \tag{61}$$

with

$$K_0(A) = A \text{ and } \dot{K}_0(A) = \nabla^2 S_0(A)\,, \text{ all } A \in T_a N\,, \tag{62}$$

is nonsingular at time s_0.

Consequently Φ_s is a diffeomorphism, for $0 \le s \le t$, of some neighborhood of a (independent of s) iff C_s holds for $s \in [O,t]$ i.e. iff K_s is nonsingular for $s \in [0,t]$.

Proof: The first assertion follows from the inverse function theorem from our equations for $T_a\Phi_s$.

Suppose now that C_s holds for $s \in [0,t]$. For a given t, let U be a precompact neighborhood of a such that Φ_{s_0} is a diffeomorphism of a neighborhood of $\bar{U}$. Since $s \to \Phi_s|\bar{U}$ is continuous in the C^1 topology it follows by openness of diffeomorphisms in the C^1 topology that Φ_s is a diffeomorphism of U for s in some neighborhood of s_0. By the compactness of $[0,t]$ there is therefore a neighborhood of a independent of s as required.

G. Nondegenerate minimality and the local diffeomorphism property

We now have the main geometrical result of this section.

Theorem 3G: Suppose S_0 is C^2 and V is C^3. Then the map Φ_s, $s\in[0,t]$, is a diffeomorphism of a certain neighborhood of a point $a\in N$ with $\Phi_t(a)=x_0$ iff

$$Z(s)=\Phi_{t-s}(a), \quad 0\leq s\leq t$$

is a nondegenerate minimum of the functional

$$\mathscr{S}: L^{2,1}_{x_0}([0,t],N)\to\mathbb{R}.$$

Proof: Assume Z is a nondegenerate minimum of the functional $\mathscr{S}$. Then, in the notation of Proposition 3D, $\mathscr{H}$ is positive definite. If condition C_s of the previous proposition does not hold, for some $s\in[0,t]$, there exists a solution K of (61) (62) with

$$K_s(A)=0, \quad \text{some } A\neq 0.$$

Setting $\xi(s)=K_{t-s}(A)$, $s\in[0,t]$, we get a contradiction to Proposition 3E.

Conversely suppose C_s holds for $s\in[0,t]$. Take a basis $A_1,A_2,\ldots,A_n$ for T_aN and set $\xi_i(s)=K(t-s)A_i$ with K as before. Then

$$\xi_i(t)=A_i,$$

$$\dot{\xi}_i(t)=-\nabla^2S_0(A_i), \quad i=1,2,\ldots,n,$$

and $\{\xi_i(s)\}_{i=1,2,\ldots,n}$ are linearly independent for each $s\in[O,t]$. However, the solutions ξ of Eq. (59) are determined by the values of $\xi(t)$ and $\dot{\xi}(t)$ and so there are at most n linearly independent solutions of Eq. (59) with $\dot{\xi}(t)=-\nabla^2S_0(\xi(t))$. Consequently any such solution must have the form

$$\xi(s)=\sum_{i=1}^{n} r_i\xi_i(s), \quad r_i\in\mathbb{R},$$

and these vanish identically or not at all. Thus $\mathscr{H}$ is positive definite by Proposition 3E.

H. A global inverse function theorem

We go on to find conditions on s which will ensure that Φ_s is a diffeomorphism; first we give a standard global inverse function theorem.

Lemma 3H: Let M and N be connected Riemannian manifolds with $\dim M=\dim N$. Suppose $\Phi: M\to N$ is smooth and, for some $\kappa>0$, for all $v\in TM$

$$|T\Phi(v)|\geq\kappa|v|.$$

Then if M is complete Φ is a covering map and N is complete. If also Φ is homotopic to a homeomorphism then Φ is a diffeomorphism.

Proof: Define $\langle\,,\rangle_{\tilde{M}}$ on M by $\langle u,v\rangle_{\tilde{M}}=\langle T_m\Phi(u),T_m\Phi(v)\rangle_N$, where $\langle\,,\rangle_N$ is the Riemannian scalar product on TN. Then, using an obvious notation, by hypothesis

$$|v|^{\sim}\geq\kappa|v|, \quad v\in TM.$$

Let d and $\tilde{d}$ be the distance functions on M associated with $\langle\,,\rangle_M$ and $\langle\,,\rangle_{\tilde{M}}$ respectively. We first show that the above inequality implies that $(M,\tilde{d})$ is complete. For, if $x,y\in M$,

$$\tilde{d}(x,y)=\inf\left\{\int_0^1|\dot{\sigma}(s)|^{\sim}ds\,|\,\sigma(0)=x,\ \sigma(1)=y,\ \sigma a C^1 \text{ path}\right\}$$

$$\geq\inf\left\{\kappa\int_0^1|\dot{\sigma}(s)|ds\,|\,\sigma(0)=x,\ \sigma(1)=y,\ \sigma a C^1\text{ path}\right\}$$

$$=\kappa d(x,y).$$

Hence, for the complete manifold (M,d), if $\{x_n\}$ is Cauchy in $(M,\tilde{d})$ then $\{x_n\}$ is Cauchy in (M,d) and consequently is convergent in (M,d), hence convergent in the topology of M and so convergent in $(M,\tilde{d})$.

Secondly by definition of $\langle\,,\rangle^{\sim}$, the map $\Phi:(M,\langle\,,\rangle^{\sim})\to N$ is an isometry in the sense that $T_m\Phi: T_mM\to T_{\Phi(m)}N$ is an isometry. This implies by the completeness of M that N is complete and Φ is surjective. To see this set $U=\Phi(M)$. Since, by the local inverse function theorem, U is open in N, we can treat U as a submanifold in its own right. For $m\in M$, $\tilde{v}\in T_mM$, let $v=T_m\Phi(\tilde{v})$. If, for the exponential map of $\langle\,\rangle^{\sim}$, $\tilde{\gamma}(t)=\exp_m(t\tilde{v})$, then $\gamma(t)=\Phi(\tilde{\gamma}(t))$ is a geodesic in N with $\dot{\gamma}(0)=v$. Hence,

$$\gamma(t)=\exp_{\Phi(m)}(tv).$$

Consequently geodesics go on for all time in U i.e., U is geodesically complete. Therefore, U is metrically complete. It follows that U is closed in N and, by connectedness, $U=N$, so N is complete and Φ is surjective.

By using a standard result (see e.g., M. Berger Ref. 17, p. 239 Prop. VII 5.1) since Φ is a surjective isometry and M is complete if follows that Φ is a covering map.

Suppose now that Φ is homotopic to a homeomorphism and $\Phi(m)=n$. Then $\Phi_*:\pi_1(M,m)\approx\pi_1(N,n)$. Let $m,m'\in\Phi^{-1}(n)$ with m,m' distinct. Take σ in M with $\sigma(0)=m$ and $\sigma(1)=m'$. Then $m\neq m'$ implies $\Phi(\sigma)$ does not lie in $\Phi_*[\sigma]$, which is a contradiction. Hence Φ is injective.

I. A lemma on second order linear ordinary differential equations

Next consider a matrix equation

$$\ddot{K}(s)+P(s)K(s)=0, \tag{63}$$

with

$$K(0)=\mathrm{I},\ \dot{K}(0)=Q \tag{64}$$

for $K(s)$ in the space $L(\mathbb{R}^n;\mathbb{R}^n)$ of $n\times n$-matrices, where $P:[0,t)\to L(\mathbb{R}^n,\mathbb{R}^n)$ is continuous and $Q\in L(\mathbb{R}^n;\mathbb{R}^n)$. Let $\|P(s)\|$ denote the operator norm of $P(s)$ and let $\Lambda(Q)$, $\lambda(Q)$ denote respectively the maximum and minimum eigenvalues of Q.

Lemma 3I: Let K satisfy (63) and (64) where Q is symmetric. For real numbers $\lambda,\Lambda,\rho,\rho\neq 0$, set

$$M(\lambda,\Lambda,\rho)(s)=2+\lambda s+\Lambda(s-\rho^{-1}\sinh(\rho s))-\cosh\rho s.$$

Suppose $\Lambda(Q)\leq\Lambda$ and $\lambda(Q)\geq\lambda$ while $\|P(s)\|\leq\rho^2$ for $0\leq s\leq t$. Let $s=\tau$ be the least positive solution of the equation

$$(1+\lambda s)M(\lambda,\Lambda,\rho)(s)=0. \tag{65}$$

Then $K(s)^{-1}$ exists on the interval $[0,\min(t,\tau))$ and satisfies $\|K(s)^{-1}\|\leq M(\lambda,\Lambda,\rho)(s)^{-1}$.

Proof: We can assume $t<\tau$.

First consider $A\in\mathrm{GL}(n)$ and $B\in L(\mathbb{R}^n;\mathbb{R}^n)$. If $\|B\|<\|A^{-1}\|^{-1}$ the inverse of $(A+B)$ exists and

$$\|(A+B)^{-1}\| = \|(1+A^{-1}B)^{-1}A^{-1}\| = \left\|\left(\sum_0^\infty (-1)^n(A^{-1}B)^n\right)A^{-1}\right\|$$

$$\leqslant \sum \|A^{-1}B\|^n\|A^{-1}\| = (1-\|A^{-1}B\|)^{-1}\|A^{-1}\|$$

$$\leqslant (\|A^{-1}\|^{-1} - \|B\|)^{-1}.$$

Take $A = A(s) = 1 + sQ$ and $B = B(s) = \int_0^s ds' \int_0^{s'} P(r)K(r)dr$, so that $K(s) = A(s) + B(s)$. Now $1 + \lambda s > 0$ for $0 \leqslant s \leqslant t < \tau$; consequently $A(s) \in \mathrm{GL}(n)$, $0 \leqslant s \leqslant t$. Also

$$\|B(s)\| \leqslant \rho^2 \int_0^s ds' \int_0^{s'} \|K(r)\|\,dr$$

$$= \rho^2 k(s),$$

where

$$\ddot{k}(s) = \|K(s)\|, \quad k(0) = \dot{k}(0) = 0.$$

To estimate $k(s)$ consider $c(s)$ defined by

$$\ddot{c}(s) = \|A(s)\| + \rho^2 c(s), \quad c(0) = \dot{c}(0) = 1.$$

We have

$$\frac{d^2}{ds^2}(c(s) - k(s)) \geqslant \rho^2(c(s) - k(s)).$$

It follows [e.g. from Caplygin's method[18]] that

$$k(s) \leqslant c(s), \quad s \geqslant 0.$$

Consequently, provided

$$\rho^2 c(s) < \|A(s)^{-1}\|^{-1} = 1 + s\lambda(Q) \tag{66}$$

then $K(s)^{-1}$ exists and

$$\|K(s)^{-1}\| \leqslant (\|A(s)^{-1}\|^{-1} - \|B\|)^{-1}$$

$$\leqslant (1 + s\lambda - \rho^2 c(s))^{-1}. \tag{67}$$

However since

$$\ddot{c}(s) - \rho^2 c(s) = 1 + \Lambda s,$$

with

$$c(0) = \dot{c}(0) = 0,$$

we have

$$c(s) = \rho^{-2}\cosh\rho s + \Lambda\rho^{-3}\sinh\rho s - \rho^{-2}(1 + \Lambda s)$$

giving

$$M(\lambda, \Lambda, \rho)(s) = 1 + s\lambda - \rho^2 c(s).$$

Inequality (66) is therefore true whenever $M(\lambda, \Lambda, \rho)(s) > 0$, which holds by assumption if $0 \leqslant s \leqslant t$. Since (67) reduces to $\|K(s)\|^{-1} \leqslant M(\lambda, \Lambda, \rho)(s)^{-1}$ the lemma is proved. // In some sense the last theorem is a best possible as the following example illustrates. It will correspond to the case $V \equiv 0$, $M = \mathbb{R}$.

Example (One-dimensional): Suppose $P \equiv 0$ and $Q < 0$. Taking $|\rho|$ arbitrarily small in the lemma yields that $K(s)^{-1}$ exists on $[0, -Q^{-1})$ and satisfies $K(s)^{-1} \leqslant (1 + Qs)^{-1}$ there. Since $K(s) = 1 + Qs$ this is exact.

J. A criterion for Φ_s to be a diffeomorphism

Theorem 3J: Let ρ^2 be an upper bound over $(s, a) \in [0, t] \times M$ of the operator norm of $\nabla^2 V(\cdot) - R(\dot{\Phi}_s(a), \cdot)\,\dot{\Phi}_s(a) : T_{\Phi_s(a)}M \to T_{\Phi_s(a)}M$. Let λ, Λ be lower and upper bounds of the set of the eigenvalues of the operators $\nabla^2 S_0(a) : T_a M \to T_a M$, $a \in M$. Assume λ, Λ and ρ to be finite, $\rho \neq 0$.

Let $s = \tau$ be the least positive solution of the equation

$$(1 + \lambda s)(2 + \lambda s + \Lambda(s - \rho^{-1}\sinh\rho s) - \cosh\rho s) = 0. \tag{68}$$

Then $\Phi_s : M \to M$ is a diffeomorphism for $0 \leqslant s \leqslant \min(\tau, t)$.

Proof: The proof is immediate from Propositon 3F and Lemmas 3H and 3I. //

Corollary 3J: If M is compact, or more generally if $\nabla V, \nabla^2 V, \nabla S_0, \nabla^2 S_0$ and R are all uniformly bounded on M, then there exists $\tau > 0$ so that $\Phi_s : M \to M$ is a diffeomorphism for $0 \leqslant s \leqslant \tau$.

4. FEYNMAN MAPS AND THE SCHRÖDINGER EQUATION ON A RIEMANNIAN MANIFOLD

A. Time reversed paths

When dealing with Feynman integrals it is conventional and more convenient to consider paths ending at a given point. However for Wiener integrals and stochastic calculus the paths are usually taken to start at that point. We shall abide by these conventions in order to facilitate comparison with previous work. For a fixed "ending time" it is easy to go from one convention to the other by the transformation J where

$$J(\alpha)(s) = \alpha(t - s), \quad 0 \leqslant s \leqslant t$$

for a path α defined on $[0, t]$.

Set $H = {}_0L^{2,1}(R^n)$, the Sobolev space of paths $\alpha[0, t] \to R^n$ of the form $\alpha(s) = \int_s^t \beta(r)\,dr$ for some L^2 path $\beta : [0, t] \to R^n$, (so α is continuous on $[0, t]$ and differentiable almost everywhere with β as derivative). The space is a Hilbert space for the inner product $\langle \alpha_1, \alpha_2 \rangle = \int_0^t \langle \dot{\alpha}_1(s), \dot{\alpha}_2(s) \rangle\, ds$. Thus J gives an isometry

$$L_0^{2,1}(R^n) \to {}_0L^{2,1}(R^n).$$

B. Definition of the Feynman maps

Let Π be a partition of $[0, t]$ with $\Pi = \{t_0, \ldots, t_m\}$, $0 = t_0 < t_1 < \ldots < t_m = t$.

Set $\Delta_j t = t_j - t_{j-1}$, $\Delta_j \alpha = \alpha(t_j) - \alpha(t_{j-1})$ for $\alpha \in H$.

Define the orthogonal projection $P_\Pi : H \to H$ by

$$P_\Pi(\alpha)(s) = \alpha(t_{j-1}) + (s - t_{j-1})(\Delta_j t)^{-1}\Delta_j \alpha, \quad t_{j-1} \leqslant s \leqslant t_j.$$

Give $P_\Pi H$ its Euclidean structure as a subspace of H. This determines an mn-dimensional Lebesque measure λ_m on it. The projection $P_\Pi(\alpha)$ of the generic path α can be given coordinates $(\Delta_1\alpha, \ldots, \Delta_m\alpha) \in R^n \times \ldots \times R^n$ and then this measure corresponds to $\sqrt{(\Delta_1 t)} \ldots \sqrt{(\Delta_m t)}\, d(\Delta_1\alpha) \ldots d(\Delta_m\alpha)$.

We now introduce the complex Gaussian $e_z : H \to C$ defined by

$$e_z(\alpha) = \exp\left\{\frac{i}{2z}\|\alpha\|^2\right\}, \quad z \in C, \quad z \neq 0.$$

Let $f : H \to E$ be a functional with values in a real linear space E. We shall take E to be a separable Banach space. The Feynman map $\mathfrak{F}^z$ will be defined as follows:

Definition: For a partition Π and for $\mathrm{Im} z \leqslant 0$, $z \neq 0$, define $F_\pi^z(f)$ by

$$F_\pi^z(f) = (2\pi i z)^{-(1/2)mn}(\Delta_1 t)^{-1/2} \ldots (\Delta_m t)^{-1/2} \int_{P_\pi H} (f e_z)\,|P_\pi H\, d\lambda_m,$$

where $|$ denotes restriction (we will usually omit it in future). The value of $(2\pi iz)^{-1/2}$ is taken so that it is continuous on $\mathrm{Im}z \leqslant 0$, $z \neq 0$, and positive for $z=-ib, b>0$. When $\mathrm{Im}z<0$ the integral is given its usual meaning: we require $(fe_z)P_\Pi H$ to be Lebesgue integrable over $P_\Pi H$. For $\mathrm{Im}z=0$ we will take the "oscillatory integral" definition and say it exists provided

$$\lim_{\epsilon\downarrow 0} \int_{P_\Pi H} f(x) e_z(x) \phi(\epsilon x)\, d\lambda_m(x)$$

exists for all $\phi: R^n\times \ldots \times R^n \to \mathbb{C}$ in the Schwarz space $\mathscr{S}(R^{mn})$ of rapidly decreasing functions with $\phi(0)=1$, and provided the limit (the "oscillatory integral") is independent of the choice of ϕ.

The definition of the finite-dimensional integral is very close to that of Tarski[19]; essentially he replaces $\phi(\epsilon x)$ by $\phi_\epsilon(x)$ with $\phi_\epsilon(x)=\exp\{-\frac{1}{2}k\epsilon\|x-x_0\|^2\}$ for $\mathrm{Re}k>0$ and $x_0\in P_\Pi H$. Tarski's definitions for both the finite- and the infinite-dimensional case would do equally well for our purposes. Our definitions are discussed in detail in Ref. 6.

We can now define $\mathscr{F}^z(f)$ by

$$\mathscr{F}^z(f)=\lim_{\Pi} F_\Pi^z(f),$$

whenever the $\mathscr{F}_\Pi^z(f)$ exists for all Π and the limit exists. Strictly speaking by the limit we will mean: for all $\rho>0$, given $\epsilon>0$ there exists $\delta>0$ with

$$|F^z(f)-F_\Pi^z(f)|_E<\epsilon$$

for all Π with mesh $\Pi\equiv\max_j \Delta_j t<\delta$ and $\Delta_j t>\rho$ mesh π all j. However for what we do here we could simply take the limit as mesh $\Pi\to 0$.

Let $\mathscr{F}^z(p\cdot l\cdot, H; E)$ denote the space of functions $f:H\to E$ for which $\mathscr{F}^z(f)$ exists. Set $\mathscr{F}^z(p\cdot l\cdot, H)=F^z(p\cdot l\cdot, H; C)$. When $z=1$, the above definition can be seen to be essentially equivalent to Feynman's original definition of the path integral.

C. Relationship with Wiener integration

When $b>0$, $\mathscr{F}^{-ib}$ coincides with the Wiener integral in the sense that if $f:{}_0C(R^n)\to C$ is bounded and continuous then $\mathscr{F}^{-ib}(f)=E^b(f)$. Here ${}_0C(R^n)$ denotes $J[C_0(R^n)]$ and E^b is the integral with respect to the measure ${}_b\gamma$ induced on ${}_0C(R^n)$ by J from Wiener measure γ_b on $C_0(R^n)$, variance parameter b.

In fact, for $\mu\geqslant 0$, define

$${}^\mu u: [0,t]\times{}_0C(R^n)\to O(N)$$

and

$${}^\mu x: [0,t]\times{}_0C(R^n)\to N$$

by

$${}^\mu u(s,w)=J(u^\mu(-,Jw))(s)$$

and

$${}^\mu x(s,w)=J(x^\mu(-,Jw))(s).$$

Then results of McShane[20], (see also Refs. 2,3,21) imply that $\{{}^\mu x(s,P_\Pi w)\}_m$ converges in ${}_b\gamma$-measure to ${}^\mu x(s,w)$ uniformly in $s\in[0,t]$ for each $b>0, \mu\geqslant 0$. Here, for $b\neq 1$, we have to define ${}^\mu x(s,w)$ in terms of Brownian motion variance parameter b. For N compact the convergence is even uniform in $\mu\in[0,\mu_0]$ any $\mu_0>0$, (3), but we will not use this. It follows by the dominated convergence theorem that if g_u is the solution to

$$\frac{\partial g_b^\mu}{\partial t}=\frac{b\mu^2}{2}\Delta g_b^\mu+\frac{\mu^{-2}}{b}Vg_b^\mu, \tag{69}$$

with $g_b^\mu(x,0)=\exp\{-\mu^{-2}/bS_0(x)\}T_0(x)$ bounded on N and V bounded above then

$$g_b^\mu(x_0,t)=\mathscr{F}^{-ib}\exp\left\{\frac{\mu^{-2}}{b}\int_0^t V({}^\mu x(s,-))\,ds\right\}g_b^\mu({}^\mu x(t,-)), \tag{70}$$

where, as usual we consider ${}^\mu x(s,-):H\to N$ defined by the classical differential equations corresponding to (4). In fact, given some regularity conditions, e.g. if N is compact, this result can be proved directly, without using stochastic differential equations. We would like to have a corresponding result for complex b; however it seems to be too much to hope to get a pointwise solution, so we must work in L^2.

D. A basic conjecture

Let dv be the Riemannian volume element on the complete manifold N. Denote by $L^2(N,dv)$ the corresponding complex Hilbert space. Then Gaffney[22] has shown that Δ, the Laplace–Beltrami operator for N, is essentially self-adjoint on some domain $\mathscr{D}(\Delta)\subset L^2(N,dv)$ with self-adjoint closure $\bar\Delta$ defined on a domain $\mathscr{D}(\bar\Delta)$. Assume that the potential V is $(-\frac{1}{2}\hbar^2\bar\Delta)$-bounded with relative bound less than unity and $\mathscr{D}(V)\supset\mathscr{D}(\bar\Delta)$. This will be true, for example, if V is bounded.

The Kato-Rellich Theorem (Ref. 23, page 162) implies that $H=(-\frac{1}{2}\hbar^2\bar\Delta+V)$ is self-adjoint on $\mathscr{D}(\bar\Delta)$. Also $\bar\Delta\leqslant 0$ and V is bounded above. Consequently for $\mathrm{Im}z \leqslant 0$, we can define the semigroup,

$$Q_t^z: L^2(N,dv)\to L^2(N,dv), \quad t\geqslant 0,$$

by

$$Q_t^z=\exp\{-i\hbar^{-1}H(z)t\},$$

where

$$H(z)=-\tfrac{1}{2}z\hbar^2\Delta+V/z.$$

For $\psi_0\in L^2(N;dv)$ and $\alpha\in H$ we can consider the map $x_0\to\psi_0({}^\mu x(t,\alpha))$. Strictly speaking this depends on a choice of frame u_0 for each x_0: but this can certainly be done in a measurable way, and the actual choice will be irrelevant.

Let $\mathscr{N}$ be a class of Riemmanian manifolds, e.g. $\mathscr{N}=\mathrm{Comp}$, the class of compact manifolds, $\mathscr{N}=\mathrm{Flat}$, the flat manifolds. Consider the following statement $C(z,\mathscr{N})$

$C(z,\mathscr{N})$: "If $\mu=\hbar^{1/2}$, $\psi_0\in L^2(N;dv)$, and $N\in\mathscr{N}$ then

$$\exp-\left\{\frac{i\mu^{-2}}{z}\int_0^t V({}^\mu x(s,-)ds\right\}$$

$$\psi_0({}^\mu x(t,-))\in F^z(p\cdot l\cdot, H; L^2(N;dv)),$$

and

$$\mathscr{F}^z(\psi_0(^\mu x(t,-)) = Q_t^z \psi_0".$$

We make the following obvious conjecture:

Conjecture: Statement $C(z, \text{Comp})$ is true for $\text{Im} z \leqslant 0$, $z \neq 0$. We would also expect it to be true for a wide class of noncompact manifolds.

The truth of $C(-ib, \text{Comp})$ for $b>0$, follows from (70): the dominated convergence theorem shows that $g_b^\mu(x_0, t)$ is the L^2-solution. We can also consider the statement $C(z, \mathscr{N}, \mathscr{V})$ obtained by restricting the potentials V to lie in some class $\mathscr{V}$. When $\mathscr{N}$ consists of Euclidean spaces, $C(z, \mathscr{N}, \mathscr{V})$ is true for z real, and $\mathscr{V}$ the class of Fourier transforms of complex measure of bounded total variation (in fact for the sum of such V and a harmonic oscillator potential) as is shown in (6). See also Tarski[19] for the case of more general bounded V.

Despite the need to work in L^2 we shall nevertheless continue working pointwise for simplicity of exposition.

E. The polygonal approximations to the stochastic development

In order to relate to previous schemes for path integration on curved space we must examine

$$\mathscr{F}_\pi^z\left[\exp\left\{\int_0^t -i\frac{\mu^{-2}}{z} V(^\mu x(s))\, ds\right\} g_0^\mu(^\mu x(0))\right].$$

It will be notationally more convenient to switch back to the $L_0^{2,1}$ picture and so consider paths starting at 0 and x_0. Take $\mu=1$, set $g_0^\mu = g$, and write $k=(2\pi i z)^{-(1/2)mn} \prod_{j=1}^m (\Delta_j t)^{-(1/2)n}$; in fact $\Delta_j t = \Delta t = t/m$ for our purposes. Then our approximation is

$$I_\pi \equiv k \int_{T_{x_0}N\times\cdots\times T_{x_0}N} \exp\left\{\frac{i}{2z}\sum_{j=1}^m \frac{|\Delta_j|^2}{\Delta_j t} - \frac{i}{z}\int_0^t V(\rho(\Delta_1,\ldots,\Delta_m)(s))\, ds\right\}$$

$$\times g(\rho(\Delta_1,\ldots,\Delta_m)(t)) d\Delta_1 \ldots d\Delta_m,$$

where $\rho \equiv \rho(\Delta_1,\ldots,\Delta_m)(\cdot)$ is the path on N given by $x^1(\cdot, u_0(\alpha))$ for $\alpha \equiv \alpha(\Delta_1,\ldots,\Delta_m)$ the polygonal path in $T_{x_0}M$ with $\alpha(0)=0$ and $\Delta_j \alpha = \Delta_j$. As before u_0 is the frame at x_0 used to define x^1. Thus $\rho(\Delta_1,\ldots,\Delta_m)$ is the classical development of α.

Now ρ can be expressed as follows: First

$$\rho(s) = \exp_{x_0}(s\Delta_1/t_1), \quad 0 \leqslant s \leqslant t_1.$$

Let $U(\Delta_1,\ldots,\Delta_j)$ denote parallel translation along ρ from 0 to t_j, so

$$U(\Delta_1,\ldots,\Delta_j)\colon T_{x_0}N \to T_{\rho(t_j)}N$$

is orthogonal. Then if $\rho_j = \rho(t_j)$

$$\rho(s) = \exp_{\rho_{j-1}}\left\{\frac{s-t_{j-1}}{\Delta_j t} U(\Delta_1,\ldots,\Delta_{j-1})\Delta_j\right\} t_{j-1} \leqslant s \leqslant t_j.$$

We will assume that $\exp_{x_0}\colon T_{x_0}N \to N$ is a diffeomorphism and that there is a unique geodesic joining any two points of N. This will be true if N has nonpositive sectional curvatures and is simply connected. Let $\gamma(t_{j+1}, a; t_{j+1}, b)$ denote the geodesic (not parametrized by arc length!) which has $\gamma(t_j)=a$, $\gamma(t_{j+1})=b$, for $a, b\in N$. Let $A(t_j, a; t_{j+1}, b)$ denote its action

$$A(t_j, a; t_{j+1}, b) = \frac{1}{2}\int_{t_j}^{t_{j+1}} |\dot\gamma(s)|^2\, ds - \int_{t_j}^{t_{j+1}} V(\gamma(s))\, ds.$$

Define $\psi\colon N\times N \to R$ by

$$\psi(x,a) = |\det D(\exp_x^{-1})(a)|,$$

where the inner products of T_xN and T_aN are used in computing the determinants of $D(\exp_x^{-1})(a)\colon T_aN \to T_xN$ i.e.

$$\psi(x,a) = \sqrt{g^{-1}(a)}\left|\det\left[\frac{\partial(\exp_x^{-1})^\alpha(a)}{\partial a^\beta}\right]_{\alpha\beta}\right|\sqrt{g^{-1}(x)}$$

for

$$g(x) = \det[g_{\alpha\beta}(x)]_{\alpha\beta}.$$

In the notation of A. Besse (Ref. 24 Sec. 6.2)

$$\psi(x,a) = \theta_x(\exp_x^{-1}a)^{-1}.$$

Consider the transformation

$$\Xi\colon T_{x_0}N\times\ldots\times T_{x_0}N \to N\times\ldots\times N$$

$$\Xi(\Delta_1,\ldots,\Delta_m) = (\rho_1,\ldots,\rho_m)$$

for ρ_j defined as above.

For $j<k$ we see $D_k\rho_j = 0$ where D_k denotes the partial derivative in the kth factor. Consequently the Jacobian determinant $|\det D\,\Xi(\Delta_1,\ldots,\Delta_n)|$ is the product of the Jacobian determinants $\prod_j |\det D_j \rho_j(\Delta_1,\ldots,\Delta_m)|$.

Now $D_j\rho(\Delta_1,\ldots,\Delta_m)$

$$= D\exp_{\rho_{j-1}}\{U(\Delta_1,\ldots,\Delta_{j-1})\Delta_j\} \circ U(\Delta_1,\ldots,\Delta_{j-1}),$$

so

$$|\det D\,\Xi(\Delta_1,\ldots,\Delta_m)| = \prod_j |\operatorname{Det} D\exp_{\rho_{j-1}}\{U(\Delta_1,\ldots,\Delta_{j-1})\Delta_j\}| = \prod_j \psi(\rho_{j-1},\rho_j)^{-1}$$

In particular Ξ is a local diffeomorphism, and hence a diffeomorphism since it is bijective by the assumptions on N. Therefore

$$I_\pi = k\int_{T_{x_0}N\times\cdots\times T_{x_0}N} \exp\left\{\frac{i}{z}\sum_j A(t_{j-1},\rho_{j-1}; t_j, \rho_j)\right.$$

$$\times g(\rho_m)\, d\Delta_1 \ldots d\Delta_m$$

$$= k\int_{N\times\cdots\times N} \exp\left\{\frac{i}{z}\sum_j A(t_{j-1},\rho_{j-1}; t_j,\rho_j)\right\}$$

$$\times g(\rho_m)\prod_j \psi(\rho_{j-1},\rho_j)\, dv(\rho_1)\ldots dv(\rho_m),$$

where v is the measure coming from the Riemmanian volume element on N.

Let $\mathscr{D}$ denote the invariant Van-Vleck determinant:

$$\mathscr{D}(t_{j-1}, x; t_j, y) = \left|\det D_y D_x \int_{t_{j-1}}^{t_j} \tfrac{1}{2}|\dot\gamma(t_{j-1}, x; t_j, y)(s)|^2\, ds\right|$$

so

$$\mathscr{D} = g^{-1/2}(x) D g^{-1/2}(y)$$

in a standard notation. Then

$$\mathscr{D} = 2^{-n}(\Delta_j t)^{-n} |\det D_y D_x d(x,y)^2|.$$

However, for $h \in T_x N$

$$D_x d(x,y)^2(h) = -2\langle \exp_x^{-1}(y), h\rangle$$

so

$$D_y D_x(x,y)^2(l,h) = -2\langle D_y(\exp_x^{-1})(s)(l), h\rangle, \quad l \in T_y N$$

giving the known result

$$\mathscr{D}(t_{j-1}, \rho_{j-1}; t_j, \rho_j) = (\Delta_j t)^{-n}\Psi(\rho_{j-1}, \rho_j).$$

Thus we have finally

$$I_\pi = (2\pi i z)^{-(1/2)nm} \int_{N\times\cdots\times N} \left[\prod_{j=1}^{m} e^{(i/z)A(t_{j-1},\rho_{j-1};t_j,\rho_j)} (\Delta_j t)^{(1/2)n} \mathscr{D}(t_{j-1},\rho_{j-1};t_j,\rho_j)\right. \times g(\rho_m)\, dv(\rho_1)\ldots dv(\rho_m). \tag{71}$$

The integral differs from that obtained using the Pauli, Van-Vleck, De-Witt propagator[5] (which is the principal term in the WKB expansion: cf. equation (78) below) by a factor of $\prod_j \sqrt{\psi(\rho_{j-1},\rho_j)}$ i.e. we have $(\Delta_j t)^{(1/2)n}\mathscr{D}$ instead of $\mathscr{D}^{1/2}$. However note that we have no terms involving the scalar curvatures appearing in our differential equations. For the case $z = -ib$, $b > 0$, we have now *proved* that I_π converges to a solution of the corresponding diffusion equation (45). For more discussion, and historical remarks, about the formula when z is real see Ref. 25.

5. THE QUASICLASSICAL REPRESENTATION FOR THE SCHRÖDINGER EQUATION

A. The quasiclassical expansion in the Feynman map notation

Applying the time reversing transformation J to Eq. (44) we see that

$$\exp\left(-\frac{i}{z}\mu^{-2}S\right)\mathscr{F}_\alpha^z\left[\exp\left\{-\frac{i}{z}\mu^{-2}\int_0^t V(^\mu x(s,\alpha))\,ds + \frac{i}{z}\mu^{-2}S_0(^\mu x(0,\alpha))\right\}T_0(^\mu x(0,\alpha))\right]$$

$$= \mathscr{F}_\alpha^z\left[\exp\left\{\frac{i}{z}\int_0^t\left\langle\frac{d}{ds}(^\nu\Theta(\alpha)(s)-\alpha(s)), \alpha(s)\,ds\right\rangle + \frac{i}{2z}\int_0^t\left|\frac{d}{ds}(^\nu\Theta(\alpha)(s)-\alpha(s))\right|^2 ds\right\}\right.$$

$$\left.\times \exp\left\{-\frac{i}{z}{}^\nu B(\mu, {}^\nu\Theta(\alpha))\right\}T_0(^\mu x(0,\alpha))\right], \quad \mu > 0 \tag{72}$$

provided either side exists. Here the α in $\mathscr{F}_\alpha^z$ is used to show that the "integration" is over α. Also we have set

$${}^\nu\Theta = J\Theta J$$

(${}^\nu\Theta^{-1}$ is given explicitly in (75) below) and

$${}^\nu B(\mu,\alpha) = B(\mu, J_\alpha)$$
$$= \frac{1}{2}\int_0^t \nabla dV(^\nu Z(s))(^0v(s)\alpha(s), {}^0v(s)\alpha(s))\,ds$$
$$= \frac{1}{2}\nabla dS_0(^\nu Z(0))(^0v(0)\alpha(0), {}^0v(0)\alpha(0))$$
$$+ \frac{1}{2}\int_0^t \langle R[^0v(s)\alpha(s), {}^\nu Z(s)]^0 v(s)\alpha(s), {}^\nu Z(s)\rangle\,ds$$
$$+ \rho(\mu, J\alpha) \tag{73}$$

for ρ as described after Eq. (44).

B. Transformation of the Feynman map

Now let $\Psi: H \to H$ be a continuous affine transformation. Set $\Psi(\alpha) = \alpha + \psi(\alpha)$, so ψ is affine. Define

$$e_z^\Psi: H \to \mathbb{C},$$

$$e_z^\Psi(\alpha) = \exp\left\{\frac{i}{2z}\langle\psi(\alpha),\psi(\alpha)\rangle_H + \frac{i}{z}\langle\psi(\alpha),\alpha\rangle_H\right\}$$
$$= \exp\left\{\frac{i}{2z}(|\Psi(\alpha)|^2 - |\alpha|^2)\right\}.$$

First we take $\Psi = {}^\nu\Theta$. The right-hand side of (72) is then just

$$\mathscr{F}_\alpha^z\left[e_z^{{}^\nu\Theta}(\alpha)\exp\left\{-\frac{i}{z}{}^\nu B(\mu,{}^\nu\Theta(\alpha))\right\}T_0(^\mu x(0,\alpha))\right.\ .$$

Just as in Chapter 1, and as in previous works, we are therefore led to consider

$$\mathscr{F}_\alpha^z\left[\exp\left\{-\frac{i}{z}{}^\nu B(\mu,\alpha)\right\}T_0(^\mu x(0,{}^\nu\Theta^{-1}(\alpha))\right],$$

although we must emphasize that we have no reason to believe that the integrand lies in $\mathscr{F}^z(p\cdot l\cdot, H)$. In particular we cannot justify this change of variables when $z \neq -ib$, $b > 0$, (see Refs. 6 and 26 for a discussion of Cameron–Martin formulas for Feynman integrals). However, by analogy with previous works we shall call the resulting expression

$$\exp\left(+\frac{i}{z\hbar}S\right)\mathscr{F}_\alpha^z\left[\exp\left\{-\frac{i}{2z}\left[\int_0^t \nabla dV(^\nu Z(s))\right.\right.\right.$$
$$\times(^0v(s)\alpha(s), {}^0v(s)\alpha(s))\,ds - \nabla dS_0(^\nu Z(0))$$
$$\times(^0v(0)\alpha(0), {}^0v(0)\alpha(0))$$
$$+\int_0^t\langle R[^0v(s)\alpha(s), {}^\nu Z(s)]^0v(s)\alpha(s), {}^\nu Z(s)\rangle\,ds$$
$$\left.\left.\left.+\rho(\hbar^{1/2}, J\alpha)\right]\right\}T_0(^\mu z(0,\alpha)\right] \tag{74}$$

the quasiclassical expansion of the solution $Q_t^z(g_0)$ to

$$\frac{\partial g}{\partial t} = iz\hbar\Delta g - \frac{i}{z\hbar}Vg$$

$$g(0,x) = \exp\left\{\frac{i}{z\hbar}S_0(x)\right\}T_0(x)$$

In (74), ${}^{\mu}z(\cdot,\alpha)$ is the classical development (in the reverse time direction) of the path ${}^{\nu}\Theta^{-1}(\alpha)$:

$$
{}^{\nu}\Theta^{-1}(\alpha)(s)=\hbar^{1/2}\alpha(s)+{}^{\nu}\sigma(s)
-\hbar^{1/2}\int_0^s\int_0^{\eta}{}^{0}v(\xi)^{-1}R[{}^{\nu}Z(\xi),{}^{0}v(\xi)\alpha(\xi)]^{\nu}\dot Z(\xi)d\xi\,d\eta \tag{75}
$$

for ${}^{0}v$ parallel translation along the classical path ${}^{\nu}Z$ which itself is the development of ${}^{\nu}\sigma$. Note that we can write

$$
\rho(\hbar^{1/2},J\alpha)=\hbar^{1/2}r(\hbar^{1/2},\alpha), \tag{76}
$$

where $r(\hbar^{1/2},\alpha)$ is continuous in $\hbar\in[0,\infty)$ for each α.

Using the notation of Chapter 3 observe that ${}^{\nu}Z(s)=\Phi_s(a)$ for $a={}^{\nu}Z(0)=Z(t)$, and that $S=\mathscr{S}({}^{\nu}Z)$. In terms of the Hessian $\mathscr{H}$ for our functional S we obtain from Sec. 3D the following expression for the quasiclassical expansion:

$$
\exp\left(+\frac{i}{z\hbar}S({}^{\nu}Z)\right)
\times\mathscr{F}^{z}_{\alpha}\left[\exp\left\{\frac{i}{2z}\left[\mathscr{H}(\alpha,\alpha)-\int_0^t|\dot\alpha(s)|^2ds-\hbar r(\hbar^{1/2},\alpha)\right]\right\}
\times T_0({}^{\mu}z(0,\alpha))\right]. \tag{77}
$$

C. Evaluation of the leading term

At $\hbar=0$ the integrand in the quasiclassical expansion (77) takes on a well-known form, (see for example Refs. 1, 4, 6), it can be evaluated for suitable t: for real z we require that $\mathscr{H}$ be nondegenerate on $T_z\mathscr{P}$ and otherwise we require that $\mathscr{H}$ be positive definite. Under these conditions we have from Ref. 6

$$
\mathscr{F}^{z}_{\alpha}\left[\exp\left\{\frac{i}{2z}\left[\mathscr{H}(\alpha,\alpha)-\int_0^t|\dot\alpha(s)|^2ds\right]\right\}\right]
=\exp\left\{-\frac{i\pi}{2}\,\mathrm{Ind}\mathscr{H}\right\}|\det K_t|^{-1/2},
$$

where $K_s\in L(T_aN;T_{\nu z(s)}N)$

satisfies Eqs. (61) and (62) viz.,

$$
\frac{D^2}{\partial s^2}K_s(-)-R[{}^{\nu}\dot Z(s),K_s(-)]^{\nu}\dot Z(s)+\nabla^2V(K_s(-))=0, \tag{61}
$$

with

$$
K_0=I,\quad \dot K_0=\nabla^2S_0. \tag{62}
$$

Here $\mathrm{Ind}\mathscr{H}$ is the index of the bilinear form $\mathscr{H}$ i.e., the dimension of a maximal subspace of H on which $\mathscr{H}$ is negative definite (see Ref. 16 for a Morse index theorem in this situation). We will set $I(x_0,t)=\mathrm{Ind}\mathscr{H}$.

The results of Sec. 3F and Corollary 3J can now be applied (for V and S_0 of class C^3 and C^2 respectively) and we can sum up with

Theorem 5C: Suppose ${}^{\nu}Z$ is a nondegenerate critical point of the action functional

$$
\mathscr{S}:\mathscr{P}\to\mathbb{R},
$$

$$
\mathscr{S}(\alpha)=S_0(\alpha(0))+\frac12\int_0^t|\dot\alpha(s)|^2ds-\int_0^tV(\alpha(s))ds.
$$

Assume either that z is real, or that $\mathrm{Im}\,z\leq 0$ and ${}^{\nu}Z$ is a local minimum for $\mathscr{S}$. Then the leading term in the quasiclassical expansion of the solution to

$$
\frac{\partial g}{\partial t}=iz\hbar\nabla g-\frac{i}{z\hbar}Vg,
$$

$$
g(0,x)=\exp\left\{\frac{i}{z\hbar}S_0(x)\right\}T_0(x),
$$

is

$$
\exp\left(+\frac{i}{z\hbar}\mathscr{S}({}^{\nu}Z)\right)
\mathscr{F}^{z}_{\alpha}\left[\exp\frac{i}{2z}\left(\mathscr{H}(\alpha,\alpha)-\int_0^t|\dot\alpha(s)|^2ds\right)T_0({}^{\nu}Z(0))\right],
$$

where $\mathscr{H}$ is the Hessian of $\mathscr{S}$ given by

$$
\mathscr{H}(\xi,\eta)
=\int_0^t\langle\dot\xi(s),\dot\eta(s)\rangle ds+\int_0^t\langle R[{}^{\nu}\dot Z(s),\xi(s)]^{\nu}\dot Z(s),\eta(s)\rangle ds
-\int_0^t\nabla dV({}^{\nu}Z(s))(\xi(s),\eta(s))ds+\nabla dS_0({}^{\nu}Z(t))(\xi(t),\eta(t)).
$$

The Feynman integral exists and on evaluation the leading term becomes

$$
\exp\left\{+\frac{i}{z\hbar}\mathscr{S}({}^{\nu}Z)-\frac{i\pi}{2}I(x_0,t)\right\}|\det T_a\Phi_t|^{-1/2}T_0(a)
$$

for $a={}^{\nu}Z(0)$ and $I(x_0,t)$ the Morse index of ${}^{\nu}Z$. //

N.B. The leading term is only a useful approximation to the solution when ${}^{\nu}Z$ is a unique minimum of $\mathscr{S}$, for $z=ib$, $b>0$, or when ${}^{\nu}Z$ is uniquely determined by ${}^{\nu}Z(t)$ for real z. In other cases we can expect to have to take a sum, or integral, over the relevant classical paths ${}^{\nu}Z$. Useful conditions on S_0, V, R and t are given in Theorem 3J and its Corollary. In fact it is rather misleading to call (74) the quasiclassical expansion except in these circumstances. There is a precise approximation theorem for the Schrödinger equation in the next chapter. For the diffusion equation see Ref. 9, and for general discussions see Refs. 27 or 28 and 29.

D. Special case of a nondegenerate minimum

Suppose now that ${}^{\nu}Z$ a nondegenerate minimum of $\mathscr{S}$. Then $\mathscr{H}$ is positive definite on $T_Z\mathscr{P}$ and so has the form $\mathscr{H}(\xi,\eta)=\langle B\xi,B\eta\rangle_H$ where $B:T_Z\mathscr{P}\to T_Z\mathscr{P}$ is positive definite. In fact B is described explicitly in Sec. 6 of Ref. 26 (for $R=0$). Our putative solution (72) can then be written as

$$
\exp\left(+\frac{i}{z\hbar}S\right)\mathscr{F}^{z}_{\alpha}\left[e^{\Psi}_{z}(\alpha)
\times\exp\left\{-\frac{i}{2z}\hbar r(\hbar^{1/2},{}^{\nu}\Theta(\alpha))\right\}T_0({}^{\mu}x(0,\alpha))\right],
$$

where $\Psi={}^{\nu}\Theta\cdot B$. Suppose for simplicity that $T_0\equiv 1$. Then a formal change of variable, valid for $z=ib$, $b>0$ reduces this to

$$
\exp\left(+\frac{i}{z\hbar}S\right)|T_a\Phi_t|^{-1/2}\mathscr{F}^{z}_{\alpha}\left[\exp\left\{-\frac{i}{2z}\hbar r(\hbar^{1/2},B^{-1}\alpha)\right\}\right]. \tag{78}
$$

6. THE RELEVANCE OF THE CLASSICAL PATHS FOR THE SCHRÖDINGER EQUATION

In this final section we discuss the relevance of the classical paths for the Schrödinger equation on a finite dimensional Riemannian manifold N. In the first subsection we prove some elementary results on the Hamilton–Jacobi and continuity equations in curved space. These are required for our subsequent results. In the second subsection we discuss in what sense the WKB approximation is valid when the relevant classical paths induce a diffeomorphism of the manifold N. We present in this section a simple theorem summarizing our main result for the Schrödinger equation on a Riemannian manifold N. This theorem shows that when the classical paths induce a diffeomorphism of N the first term of the quasiclassical representation is, for small $\hbar$, close (in the L^2 norm) to the actual solution. Most of the results here are well known and can be found in much greater generality in Refs. (27)–(30). However it seemed worth proving them here for our special case, and without having to introduce all the extra notation of those references.

A. The Hamilton–Jacobi equation

We begin with an elementary lemma. We use the notation of the preceding sections and assume that V and S_0 are real valued and C^2. We also assume that $\Phi_t : N \to N$ is a diffeomorphism for $0 \le t \le \tau$. Recall that ${}^{\nu}Z(s) = \Phi_s({}^{\nu}Z(0)) = \Phi_s \Phi_t^{-1}(x_0)$ and that

$$\ddot{\Phi}_s(a) = -\nabla V(\Phi_s(a)) \text{ with } \Phi_0(a) = a, \quad \dot{\Phi}_0(a) = \nabla S_0(a). \tag{79}$$

Lemma 6A: Define

$$S : N \times [0,t] \to \mathbb{R}$$

by

$$S(x,t) = S_0(\Phi_t^{-1}(x)) + \frac{1}{2}\int_0^t |\dot{\Phi}_s \circ \Phi_t^{-1}(x)|^2 ds - \int_0^t V(\Phi_s \circ \Phi_t^{-1}(x)) ds .$$

Then S satisfies the Hamilton–Jacobi equation

$$\tfrac{1}{2}|\nabla S(x,t)|^2 + V(x) + \frac{\partial S}{\partial t}(x,t) = 0, \quad 0 \le t \le \tau \tag{80}$$

with initial condition $S(x,0) = S_0(x)$.

The flow $\{\Phi_t\}_{0 \le t \le \tau}$ is the flow of the time dependent vector field ∇S on N, i.e. it satisfies

$$\frac{\partial}{\partial t}\Phi_t(x) = \nabla S(\Phi_t(x),t). \tag{81}$$

(By ∇ we will always mean the gradient in the space variables.)

Proof: Let $A^* : T_yN \to T_xN$ denote the adjoint of any given linear $A : T_xN \to T_yN$. Then taking the gradients:

$$\nabla S(x,t) = (T\Phi_t^{-1})^* \nabla S_0(\Phi_t^{-1}(x)) + \int_0^t \frac{D}{\partial s}[T(\Phi_s \circ \Phi_t^{-1})^*](\dot{\Phi}_s \circ \Phi_t^{-1}(x)) ds - \int_0^t T(\Phi_s \circ \Phi_t^{-1}) \nabla V(\Phi_s \circ \Phi_t^{-1}(x)) ds .$$

Integrating the second integral by parts gives

$$\nabla S(x,t) = \dot{\Phi}_t \circ \Phi_t^{-1}(x),$$

whence

$$\dot{\Phi}_t(a) = \nabla S(\Phi_t(a),t)$$

proving (81). On the other hand, differentiating for t:

$$\frac{\partial S}{\partial t}(x,t) = dS_0(\dot{\Phi}_t^{-1}(x)) + \tfrac{1}{2}|\dot{\Phi}_t \circ \Phi_t^{-1}(x)|^2 + \int_0^t \left\langle \frac{D}{\partial s} T\Phi_s(\dot{\Phi}_t^{-1}(x)), \dot{\Phi}_s \circ \Phi_t^{-1}(x)\right\rangle ds - V(x) - \int_0^t dV \circ T\Phi_s(\dot{\Phi}_t^{-1}(x)) ds .$$

Integrate the second integral by parts:

$$\int_0^t \left\langle \frac{D}{\partial s} T\Phi_s(\dot{\Phi}_t^{-1}(x)), \dot{\Phi}_s \circ \Phi_t^{-1}(x)\right\rangle ds = \langle T\Phi_t(\dot{\Phi}_t^{-1}(x)), \dot{\Phi}_t \circ \Phi_t^{-1}(x)\rangle - \langle \dot{\Phi}_t^{-1}(x), \nabla S_0(\Phi_t^{-1}(x))\rangle + \int_0^t \langle T\Phi_s(\dot{\Phi}_t^{-1}(x)), \nabla V(\Phi_t^{-1}(x))\rangle ds$$

to obtain (80) on using (81) and observing that $\Phi_t \circ \Phi_t^{-1} = id$ yields

$$T\Phi_t \circ \dot{\Phi}_t^{-1} + \dot{\Phi}_t \circ \Phi_t^{-1} = 0. \tag{82}$$

B. The continuity equation

Our next lemma follows from (81) by a well-known result about flows of vector fields. Define

$$\phi : N \times [0,\tau) \to \mathbb{R}$$

by

$$\phi(x,t) = |\det T_x\Phi_t^{-1}|,$$

where the Riemannian metric is used to give orthonormal bases in T_xN and $T_{\Phi_t^{-1}(x)}N$ so as to compute the determinant of $T_x\Phi_t^{-1} : T_xN \to T_{\Phi_t^{-1}(x)}N$ up to sign.

Lemma 6B: The map ϕ satisfies the continuity equation

$$\frac{\partial\phi}{\partial t}(x,t) + \operatorname{div}[\phi(x,t)\nabla S(x,t)] = 0. \tag{83}$$

Proof: Since $T_x\Phi_{t+s}^{-1} = T_y\Phi_t^{-1} \circ T_x(\Phi_t \circ \Phi_{t+s}^{-1})$ for $y = \Phi_t \circ \Phi_{t+s}^{-1}(x)$ we have

$$\phi(x,t+s) = \phi(y,t)|\det T_x(\Phi_t \circ \Phi_{t+s}^{-1})|.$$

Consequently, at $s = 0$,

$$\frac{\partial}{\partial s}\phi(x,t+s) = \langle \nabla\phi(x,t), T\Phi_t(\dot{\Phi}_t^{-1}(x))\rangle + \phi(x,t)\frac{\partial}{\partial s}\exp \operatorname{tr} \ln T_x(\Phi_t\Phi_{t+s}^{-1}),$$

where we have taken a coordinate system near x in order to make sense of the logarithm.

Now, using (82)

$$\langle \nabla\phi(x,t), T\Phi_t(\dot{\Phi}_t^{-1}(x))\rangle = -\langle \nabla\phi(x,t), \dot{\Phi}_t(\Phi_t^{-1}(x))\rangle = -\langle \nabla\phi(x,t), \nabla S(x,t)\rangle.$$

by Lemma 6A.

Also, at $s=0$,

$$\frac{\partial}{\partial s}\exp \operatorname{tr}\ln T_x(\Phi_t \circ \Phi_{t+s}^{-1})$$
$$=\operatorname{tr}\nabla(T\Phi_t \circ \dot{\Phi}_t^{-1})(x)$$
$$=-\operatorname{tr}\nabla(\dot{\Phi}_t \circ \Phi_t^{-1})(x) \quad \text{by (82)}$$
$$=-\operatorname{tr}\nabla^2 S(x,t).$$

Since

$$\operatorname{div}\phi S=\operatorname{tr}\nabla^2\phi S$$
$$=\langle\nabla\phi,\nabla S\rangle+\phi \operatorname{tr}\nabla^2 S$$

the result follows.

C. Some technical formulas

Our next lemma is a ragbag; we are staying with the same notation.

Lemma 6C:

(i) For C^2 functions $f,g:N\to\mathbb{C}$ we have

$$\nabla(fg)=f\nabla g+2\langle\nabla f,\nabla g\rangle+g\nabla f \tag{84}$$

(ii) $$\left(\frac{i\hbar}{2}\Delta+\frac{V}{i\hbar}\right)\exp\left(\frac{i}{\hbar}S\right)$$
$$=\left(\frac{i}{\hbar}\frac{\partial S}{\partial t}-\tfrac{1}{2}\Delta S \ \exp\frac{i}{\hbar}S\right) \tag{85}$$

(iii) for a C^1 map $\theta:N\to\mathbb{C}$

$$\frac{\partial}{\partial t}\left|\phi^{1/2}\theta \circ \Phi_t^{-1}\right|=-\tfrac{1}{2}\phi^{-1/2}\operatorname{div}(\phi\nabla S)-\phi^{1/2}d\theta \circ T\Phi_t^{-1}(\nabla S). \tag{86}$$

Proof: Part (i) is elementary. For (ii), because the divergence is the trace of the covariant derivative

$$\Delta\exp\left(\frac{i}{\hbar}S\right)=\operatorname{div}\left(\frac{i}{\hbar}\exp\left(\frac{i}{\hbar}S\right)\nabla S\right)$$
$$=\frac{i}{\hbar}\exp\left(\frac{i}{\hbar}S\right)\Delta S-\hbar^{-2}|\nabla S|^2\exp\left(\frac{i}{\hbar}S\right)$$
$$=\left(\frac{i}{\hbar}\Delta S+2\hbar^{-2}V(x)+2\hbar^{-2}\frac{\partial S}{\partial t}\right)\exp\left(\frac{i}{\hbar}S\right)$$

by the Hamilton–Jacobi equation (80). This proves (ii). For (iii) we use the continuity equation (83) to obtain

$$\frac{\partial}{\partial t}\phi^{1/2}=\tfrac{1}{2}\phi^{-1/2}\frac{\partial\phi}{\partial t}$$
$$=-\tfrac{1}{2}\phi^{-1/2}\operatorname{div}(\phi\Delta S).$$

Also by (81) and (82)

$$\frac{\partial}{\partial t}\Phi_t^{-1}(x)=-T\Phi_t^{-1}(\nabla S(x,t))$$

giving

$$\frac{\partial}{\partial t}\theta(\Phi_t^{-1}(x))=-d\theta \circ T\Phi_t^{-1}(\nabla S(x,t))$$

and we have (86).

D. Basic assumptions

We now discuss in a simple case the validity of the WKB approximation arising as the first term in our (quasiclassical representation) for the Schrödinger equations on the n-dimensional Riemannian manifold N. To prove the main result of this section, Theorem 6F, we need to make some simplifying assumptions. These are principally the assumptions of the last section; we recapitulate them here:

Gaffney's results show that when the finite dimensional Riemannian manifold N is complete then $H_0=-\hbar^2\Delta/2$ with some natural domain of definition in $L^2(N,dv)$, the space of functions square integrable with respect to the natural Riemannian volume element dv, is essentially self-adjoint. Hence $\bar{H}_0$, the closure of H_0, is self-adjoint on some domain $\mathcal{D}(\bar{H}_0)$. When the real-valued potential V with $\mathscr{D}(V)\supset\mathscr{D}(\bar{H}_0)$ is $\bar{H}_0$-bounded with relative bound less than unity the Kato–Rellich theorem implies that $H=(\bar{H}_0+V)$ is self-adjoint on $\mathscr{D}(\bar{H}_0)$ and according to Stone's theorem, generates a continuous unitary one parameter group

$$U(t)=\exp(-itH/\hbar),\quad U(t):L^2(N,dv)\to L^2(N,dv).$$

Then putting $\psi_t=\psi(\cdot,t)\in L^2(N,dv)$ we have

$$\psi_t=U(t)\psi_0$$

gives a (weak) solution of the Schrödinger equation in the sense that

$$\lim_{t\to 0}\frac{[U(t)\psi-\psi]}{t}=-\frac{i}{\hbar}H\psi,\quad \psi\in\mathscr{D}(H)=\mathscr{D}(\bar{H}_0).$$

In what follows we shall assume V is at least this well-behaved. Such a V is called a Kato potential in Euclidean space.

We now come to our most restrictive assumption. Let $\Phi_t:N\to N$ be defined as before. We assume that, for each $0\leq t<\tau$, Φ_t is a C^3 diffeomorphism.

Observe that when N is a complete finite dimensional Riemannian manifold Theorem 3J effectively gives us a minimum positive value for τ in this assumption.

E. An auxiliary semigroup

Define the family of unitary operators

$$W_0(t):L^2(N)\to L^2(N)$$

by

$$W_0(t)(\theta)(x)=\exp\left(\frac{i}{\hbar}S(x,t)\right)\phi^{1/2}(x,t)\theta(\Phi_t^{-1}(x)). \tag{87}$$

Then

$$W_0(t)^{-1}(\theta)(x)=\exp\left(\frac{-i}{\hbar}S(\Phi_t(x),t)\right)\phi^{-1/2}(\Phi_t(x),t)\theta(\Phi_t(x)). \tag{88}$$

Also define

$$W(t,s):L^2(N)\to L^2(N)$$

by

$$W(t,s)=W_0(s)^{-1}U(s-t)W_0(t).$$

We can think of this as defined for $0\leq s\leq t<\tau$ although it is defined for all $s,t\in[0,\tau)$.

Proposition 6E: The family $\{W(t,s)\}$ is a (time inhomogeneous) semigroup of unitary operators:

$$W(t,s)W(u,t)=W(u,s).$$

For $\theta: N\to\mathbb{C}$ a C^∞ function of compact support

$$\lim_{s\to 0^+}\frac{W(t+s,t)\theta-\theta}{s}$$
$$=-\tfrac{1}{2}i\hbar\phi^{-1/2}(\Phi_t(\cdot),t)(\Delta(\phi^{1/2}\theta\circ\Phi_t^{-1}))\circ\Phi_t \tag{89}$$

with convergence in $L^2(N)$, for each $t\in[0,\tau)$. (Geometrically: if θ is considered as a scalar density of order $\frac{1}{2}$ then the infinitesimal generator of $\{W(t,s)\}$ is the transport of $-\frac{1}{2}i\hbar\Delta$ by the classical flow.)

Proof: It is clear that we have a unitary semigroup. Part (iii) of Lemma 6C shows that, pointwise on N,

$$\frac{\partial}{\partial t}[W_0(t)\theta]=\left(\frac{i}{\hbar}\frac{\partial S}{\partial t}-\frac{1}{2}\phi^{-1}\operatorname{div}(\phi\nabla S)\right)W_0(t)\theta$$
$$-\exp\left(\frac{i}{\hbar}S\right)\phi^{1/2}d\phi\circ T_x\Phi_t^{-1}(\nabla S)$$
$$=\left(\frac{i}{\hbar}\frac{\partial S}{\partial t}-\tfrac{1}{2}\Delta S-\tfrac{1}{2}\phi^{-1}\langle\nabla\phi,\nabla S\rangle\right)W_0(t)\theta$$
$$-\exp\left(\frac{i}{\hbar}S\right)\phi^{1/2}d\theta\circ T_x\Phi_t^{-1}(\nabla S). \tag{90}$$

However

$$\frac{1}{s}(W_0(t+s)-W_0(t))=\int_0^1\frac{\partial W_0(t+\lambda s)}{\partial(t+\lambda s)}\,d\lambda$$

pointwise on N and θ has compact support. Consequently the joint continuity of $\Phi_t(x)$ on $[0,\tau)\times N$ together with the dominated convergence theorem allows us to interpret the derivative $(\partial/\partial t)[W_0(t)\theta]$ as a derivative in $L^2(N)$.

Continuing in $L^2(N)$ now, and using Eqs. (84), and (87)

$$\frac{\partial}{\partial s}U(-s)W_0(t)\theta\bigg|_{s=0}=-\left(\tfrac{1}{2}i\hbar\Delta+\frac{V}{i\hbar}\right)W_0(t)\theta$$
$$=-\phi^{1/2}(\theta\circ\Phi_t^{-1})\left(\tfrac{1}{2}i\hbar\Delta+\frac{1}{i\hbar}V\right)\exp\left(i\frac{S}{\hbar}\right)$$
$$-i\hbar\left\langle\nabla\exp\left(\frac{iS}{\hbar}\right),\nabla(\phi^{1/2}\theta\circ\Phi_t^{-1})\right\rangle$$
$$-\exp\left(\frac{i}{\hbar}S\right)\tfrac{1}{2}i\hbar\Delta(\phi^{1/2}\theta\circ\Phi_t^{-1}). \tag{91}$$

Also

$$-i\hbar\left\langle\nabla\exp\left(\frac{i}{\hbar}S\right),\nabla(\phi^{1/2}\theta\circ\Phi_t^{-1})\right\rangle$$
$$=\tfrac{1}{2}\langle\nabla S,\phi^{-1}\nabla\phi\rangle W_0(t)\theta+\phi^{1/2}\exp\left(\frac{i}{\hbar}S\right)d\theta\circ T\Phi_t^{-1}(\nabla S). \tag{92}$$

Now

$$\lim_{s\to 0^+}\frac{1}{s}(W(t+s,t)\theta-\theta)$$
$$=W_0(t)^{-1}\left(\frac{\partial}{\partial t}W_0(t)\theta+\frac{\partial}{\partial s}U(-s)W_0(t)\theta\bigg|_{s=0}\right),$$

and if (90) and (91) are substituted into this the result follows on using (92) and (85). //

F. The validity of the WKB approximation when the classical flow induces a diffeomorphism of configuration space

Theorem 6F: Let $\psi_\hbar(x,t)$ be the solution of the Schrödinger equation on the manifold N,

$$\frac{\partial\psi_\hbar}{\partial t}=\frac{1}{2}i\hbar\Delta\psi_\hbar+\frac{1}{i\hbar}V(x)\psi_\hbar$$

with Cauchy data

$$\psi_\hbar(x,0)=\exp\left(\frac{i}{\hbar}S_0(x)\right)T_0(x),$$

where S_0 is real valued and T_0 independent of $\hbar$. Then given the above assumptions on S_0, V, N: in particular assuming Φ_t is a diffeomorphism for $0\le t\le\tau$ we have, for T_0 infinitely differentiable and with compact support

$$\exp\left(-\frac{i}{\hbar}S(x,t)\right)\psi_\hbar(x,t)\to|\det T_x\Phi_t^{-1}|^{1/2}T_0(\Phi_t^{-1}x)$$

in $L^2(N)$ with respect to x, as $\hbar\to 0$, uniformly in $t\in[0,\tau_0)$ any $0<\tau_0<\tau$.

Proof: The proof is a minor modification of the proof of the corresponding result in (1). Let $A(t)$ be the infinitesimal generator of the semigroup $\{W(t,s)\}$ defined above; so

$$A(t)T_0=-\frac{i}{2}\hbar\phi^{1/2}(\Phi_t(\cdot),t)(\Delta(\phi^{1/2}T_0\circ\Phi_t^{-1}))\circ\Phi_t.$$

Set

$$T_t=W(t,0)T_0.$$

Then

$$W(t,0)A(t)T_0=\lim_{s\to 0^+}\frac{1}{s}[W(t,0)W(t+s,t)T_0-T_t]$$
$$=\lim_{s\to 0^+}\frac{1}{s}[T_{t+s}-T_t].$$

In particular the latter limit exists in $L^2(N)$. It follows that the right derivative in t of $\langle T_0,T_t\rangle_{L^2}$ exists and is continuous. Consequently, by general principles (e.g., Ref. 31, Chapter IX Sec. 3) the derivative exists and

$$|\langle T_0,T_t\rangle_{L^2}-\langle T_0,T_0\rangle_{L^2}|$$
$$=\left|\int_0^t\langle T_0,W(s,0)A(s)T_0\rangle ds\right|$$
$$\le\|T_0\|_{L^2}\hbar\int_0^t\left\|\frac{1}{\hbar}A(s)T_0\right\|_{L^2}ds$$

by the Cauchy–Schwarz inequality and the isometric property of $W(s,0)$. Since $(1/\hbar)A(s)T_0$ is independent of $\hbar$ and bounded uniformly in s on $[0,\tau_0]$ we see

$$\lim_{\hbar\to 0^+}\langle T_0,T_t\rangle=\langle T_0,T_0\rangle_{L^2}$$

uniformly in

$$0\le t\le\tau_0.$$

From this, since

$$\langle T_t, T_t\rangle_{L^2} = \langle T_0, T_0\rangle_{L^2},$$

we have

$$\|T_t - T_0\|_{L^2}^2 = 2\langle T_0, T_0\rangle - 2\mathrm{Re}\langle T_0, T_t\rangle_{L^2} \to 0$$

as $\hbar \to 0$, uniformly in $0 \leq t \leq \tau_0$.

Using the definition of T_t this gives

$$\lim_{\hbar\to 0^+} \|W_0(t)T_0 - U(t)W_0(0)T_0\|_{L^2} = 0$$

and the theorem is proved.//

A brief discussion of the fact that this result gives the correct classical limit of quantum mechanics for small times is given in Ref. 36.

ACKNOWLEDGMENTS

Part of the work on this article was done while the first named author was a visiting professor at the Center for Relativity, Department of Physics, University of Texas at Austin, on sabbatical leave from the University of Warwick; and while the second named held a Science Research Council Senior Visiting Fellowship at Princeton University, Mathematics and Physics Departments, arranged by B. Simon and A.S. Wightman. Discussions and correspondence with C. DeWitt-Morette, B. Nelson, and B. Simon were very helpful.

Note added in proof: With reference to the results of Sec. 3, H. Doss has recently pointed out to us that, when there is a unique classical path Z, the Theorems 1 and 5 in his article, Ann. Inst. H. Poincaré, Vol. XVI No. 1, 17–28 (1980), together with inequalities (ii)′ and (iii)′ of Sec. 3A, combine to give equality in Corollary 3A even with lim replaced by lim. This strengthens the results of Sec. 3C to a complete proof of convergence to the WKB term in this case.

[1]A. Truman, "Classical mechanics: the diffusion (heat) equation and the Schrödinger equation," J. Math. Phys. **18**, 2308–15 (1977).

[2]K. D. Elworthy, "Stochastic differential equations on manifolds," notes of seminars by J. Eells and K. D. Elworthy, Department of Mathematics, University of Warwick, Coventry, England.

[3]K. D. Elworthy, "Stochastic dynamical systems and their flows," in *Stochastic Analysis*, edited by A. Friedman and M. Pinsky (Academic, New York, 1978), pp. 79–95.

[4]C. DeWitt-Morette, A. Maheshwari, and B. Nelson, "Path integration in nonrelativistic quantum mechanics," Phys. Rep. **50**, 257–372 (1979).

[5]B. S. DeWitt, "Dynamical theory in curved spaces. I. A review of the classical and quantum action principles," Rev. Mod. Phys. **29**, 337–397 (1957).

[6]K. D. Elworthy and A. Truman, "Feynman maps, Cameron-Martin formulae and the anharmonic oscillator," (in preparation).

[7]S. R. S. Varadhan, "Diffusion processes in a small time interval," Comm. Pure App. Math.**20**, 659–685 (1967).

[8]S. A. Molchanov, "Diffusion processes and Riemannian geometry," Usp. Mat. Nauk, **30**, 3–59 (1975) [Russian Math. Surveys **30**, 1–63 (1975)].

[9]M. A. Pinsky, "Stochastic Riemannian geometry," Probabilistic Analysis and Related Topics **1**, 199–236 (1978).

[10]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. 1 (Interscience, New York, 1963).

[11]P. Malliavin, "Un principe de transfert et son application au calcul des variations," C. R. Acad. Sci. Serie A, Paris, **284**, 187–199 (1977).

[12]I. I. Gikman and A. V. Skorohod, *Stochastic Differential Equations*, Ergebn. Math. Vol. 72 (Springer-Verlag, Berlin, 1972).

[13]J. Milnor, *Morse Theory*, *Annals of Math. Studies* (Princeton U. P., Princeton, N. J., 1963).

[14]B. Simon, *Functional Integration and Quantum Physics* (Academic, New York, 1979).

[15]H. Eliasson, "Geometry of manifolds of maps," J. Differential Geom. 1, 169–194 (1967).

[16]H. Edwards, "A generalised Sturm theorem," Ann. Math. **80**, 22–57 (1964).

[17]M. Berger, "Lectures on Geodesics in Riemannian Geometry," Tata Institute Notes, Bombay (1965).

[18]E. F. Beckenbach and R. Bellman, *Inequalities*, Ergebn. Math. Vol. 30 (Springer, Berlin, 1961).

[19]J. Tarski, "Feynman type integrals defined in terms of general cylindrical approximations," in *Feynman Path Integrals*, edited by S. Albeverio *et al.*, Springer Lecture Notes in Physics, **106**, 254–79 (1978).

[20]E. J. McShane, *Stochastic Calculus and Stochastic Models* (Academic, New York, 1974).

[21]N. Ikeda, S. Nakao, and Y. Yamato, "A class of approximations of Brownian motion," Pub. RIMS, Kyoto University **13**, 285–300 (1977).

[22]M. P. Gaffney, "A special Stoke's theorem for complete Riemannian manifolds," Ann. Math. **60**, 140–5 (1954).

[23]M. Reed and B. Simon, *Fourier Analysis and Self-Adjointness*, *Methods of Modern Mathematical Physics*, Vol. II (Academic, New York, 1975).

[24]A. L. Besse, *Manifolds all of whose Geodesics are Closed*, Ergebn. Math. Vol. 93 (Springer, Berlin, 1978).

[25]Elworthy's Class (C. De Witt-Morette, K. D. Elworthy, B. L. Nelson and G. S. Sammelman), "A stochastic scheme for constructing solutions of the Schrödinger equation," Preprint, Center for Relativity, Department of Physics, University of Texas, Austin, Texas 78712. (To appear in Ann. Inst. H. Poincaré.)

[26]A. Truman, "Polygonal path formulation of the Feynman path integral," in *Feynman Path Integrals*, edited by S. Albeverio *et al.*, Springer Lecture Notes in Physics, **106**, 73–102 (1978).

[27]V. Guillemin and S. Sternberg, Geometric Asymptotics, Amer. Math. Soc. Math. Surveys **14** (Amer. Math. Soc., Providence, R. I., 1977).

[28]J. J. Duistermaat, Fourier integral operators, Courant Institute Lecture Notes (New York University, New York, 1977).

[29]J. J. Duistermaat, "Oscillatory integrals, Lagrange immersions and unfolding of singularities," Comm. Pure Appl. Math. **27**, 207–81 (1974).

[30]R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin, New York, 1978).

[31]K. Yosida, *Functional Analysis*, Grund. Math. Wiss. **123** (Springer, Berlin, 1974). In Ref. 32 below, Feynman path integrals are used to compute all the terms of the semiclassical expansion for a particle in R^n acted on by a class of potentials in $\mathfrak{F}(\mathbf{R}^n)$. In Refs. 33, 34, and 35 formal use of phase space path integrals is made to obtain formulas for all the terms of the semiclassical expansion in the case of a particle on a Riemannian manifold and for more general systems.

[32]S. Albeverio and R. Hoegh-Krohn, "Oscillatory integrals

and the method of stationary phase in infinitely many dimensions with applications to the classical limit of quantum mechanics. I." Invent. Math. **40**, 59–106 (1977).

[33]F. Langouche, D. Roekaerts, and E. Tirapegui, "Higher order corrections through functional integrals to the WKB expansion in curved spaces," Lett. Nuovo Cimento **25**, 307–13 (1979).

[34]F. Langouche, D. Roekaerts, and E. Tirapegui, "WKB expansions with higher order corrections: the general case," Phys. Lett. A **72**, 413–16 (1979).

[35]F. Langouche, D. Roekaerts and E. Tirapegui, "WKB-type expansions for Langevin equations." Preprint, Inst. Phys. Théorique, Univ. de Louvain, Belgium (1979).

[36]K. D. Elworthy and A. Truman, "The classical limit of quantum mechanics in a curved space background," in Proceedings of the International Conference on Functional Integration and its Applications, Louvain, November 1979 (Plenum, New York) (to appear).

On rotations in a pseudo-Euclidean space and proper Lorentz transformations

K. N. Srinivasa Rao, D. Saroja,[a] and A. V. Gopala Rao
Department of Physics, University of Mysore, Manasagangotri, Mysore 570 006, India

(Received 7 May 1980; accepted for publication 22 August 1980)

It is shown that in a general pseudo-Euclidean space E^p_n, 2-flats (planes) passing through the origin of the coordinate system may be classified into six invariant types and explicit formulas for "planar rotations" in these flats are obtained. In the physically important case of the Minkowski World E^3_4, planar rotations are characterized as *rotationlike*, *boostlike* and *singular* transformations and an invariant classification of proper Lorentz transformations into these types is given. It is shown that a general nonsingular proper Lorentz transformation may be resolved as a *commuting* product of two transformations one of which is rotationlike and the other boostlike while a singular transformation may be written as a product of two rotationlike transformations, each with a rotation angle π. Such a rotationlike transformation with angle π is called "exceptional" following Weyl's terminology for similar transformations of SO(3). In all cases, explicit formulas for the angles and planes of rotations in terms of the elements of a given Lorentz matrix are obtained and the procedure yields in a natural manner an explicit formula for the image of L in the $D^{10}(D^{01})$ representation of SO(3,1) which in turn leads to two more classification schemes in terms of the character χ of L in the $D^{10}(D^{01})$ and the $D^{\frac{1}{2}0}(D^{0\frac{1}{2}})$ representations.

PACS numbers: 03.30. + p

1. INTRODUCTION

We consider here a general pseudo-Euclidean space E^p_n spanned by the *orthonormal* basis $\mathbf{I}_k$ with

$$\begin{aligned} \tilde{\mathbf{I}}_k &= (0,0,...0,1,0,...,0); \quad p \geqslant k \geqslant 1, \\ \tilde{\mathbf{I}}_k &= (0,0,...0,i,0,...0); \quad p+1 \leqslant k \leqslant n, \end{aligned} \tag{1.1}$$

where only the k th-component of $\mathbf{I}_k$ is nonzero and is either 1 or i as indicated and the tilde denotes matrix transposition. An arbitrary vector of E^p_n is given by

$$\mathbf{X} = \sum_k x_k \mathbf{I}_k, \tag{1.2}$$

where x_k are all real. The *scalar product* of any two vectors $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$\tilde{\mathbf{X}}\mathbf{Y} = \tilde{\mathbf{Y}}\mathbf{X} = \sum_{i,j} g_{ij} x_i y_j, \tag{1.3}$$

where the diagonal matrix

$$(g_{ij}) \equiv (\tilde{\mathbf{I}}_i \mathbf{I}_j) = \mathrm{diag}(1,1,...1,-1,-1,...-1), \tag{1.4}$$

with p "plus-ones" and $(n-p)$ "minus-ones," is the metric induced on E^p_n by the scalar product given in Eq. (1.3). $\mathbf{X}$ and $\mathbf{Y}$ are said to be *orthogonal* if $\tilde{\mathbf{X}}\mathbf{Y} = 0$. Following a familiar nomenclature used in relativity theory, we define a vector $\mathbf{X}$ to be *timelike*, *null* or *spacelike* according as

$$\mathbf{X}^2 \equiv \tilde{\mathbf{X}}\mathbf{X} \gtreqless 0. \tag{1.5}$$

A non-null vector $\mathbf{X}$ is said to be a *unit* vector if $\mathbf{X}^2 = \pm 1$. Evidently the signature of g_{ij} is $p-(n-p) = (2p-n)$. Here we may observe that the space E^{n-p}_n is completely equivalent to E^p_n except that E^{n-p}_n has a metric with the opposite signature $(n-2p)$ and the vector nomenclatures "timelike" and "spacelike" get interchanged.

In what follows, we adopt the following notation. As seen above, boldface capitals such as $\mathbf{I}_k$, $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{P}$, $\mathbf{Q}$ etc., denote n-dimensional column-vectors of E^p_n and $\tilde{\mathbf{I}}_k$ etc., the corresponding row vectors. Ordinary lightface capitals such as L,S,A,I,S_r,S_b etc., denote $n \times n$ matrices. E is the $n \times n$ unit matrix. Elements of a matrix A are denoted by A_{ij}. In particular, these symbols denote, respectively, 4-vectors and 4×4 matrices in the case of the Minkowski world E^3_4. Lower case boldface letters such as $\mathbf{e}$, $\mathbf{h}$, $\mathbf{m}$, $\mathbf{n}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and the two boldface script letters $\mathscr{E}$ and $\mathscr{H}$ denote 3-vectors.

2. TWO-FLATS PASSING THROUGH THE ORIGIN AND THEIR CLASSIFICATION

Any two-dimensional subspace of E^p_n defined by the parametric equations[1]

$$\mathbf{R} = \eta\mathbf{P} + \mu\mathbf{Q} + \mathbf{C}, \tag{2.1}$$

may be called a *2-flat* of E^p_n. Here $\mathbf{R}$, as usual, is the *radius vector* of a general point on the 2-flat defined by the fixed, linearly independent, vectors $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{C}$ of E^p_n and (η,μ) are two real parameters taken from the range $-\infty < \eta,\mu < \infty$. A 2-flat passing through the coordinate origin is given by Eq. (2.1) with $\mathbf{C} = 0$, and such a 2-flat is completely determined by the linearly independent vector pair $(\mathbf{P},\mathbf{Q})$ which however is not unique. Any other vector pair $(\mathbf{X},\mathbf{Y})$ related to $(\mathbf{P},\mathbf{Q})$ by

$$\mathbf{X} = a\mathbf{P} + b\,\mathbf{Q}, \quad \mathbf{Y} = c\mathbf{P} + d\,\mathbf{Q}, \tag{2.2}$$

where a,b,c and d are any four real numbers satisfying

$$ad - bc \neq 0, \tag{2.3}$$

[a] Formerly C.S.I.R. Senior Research Fellow.

serves equally well to define a 2-flat passing through the origin. This situation permits us to choose $(\mathbf{X},\mathbf{Y})$ to be an orthogonal pair of vectors. To see this, let us suppose that $(\mathbf{P},\mathbf{Q})$ is nonorthogonal. Then Schmidt's orthogonalization procedure adapted to E_n^p yields the following prescriptions for orthogonalization:

(i) If $\mathbf{P}$ and $\mathbf{Q}$ are both null, then $\mathbf{X} = \mathbf{P} + \mathbf{Q}$ and $\mathbf{Y} = \mathbf{P} - \mathbf{Q}$ are orthogonal and as $\mathbf{X}^2 = -\mathbf{Y}^2 = 2\tilde{\mathbf{P}}\mathbf{Q} \neq 0$, $\mathbf{X}$ is timelike if $\mathbf{Y}$ is spacelike and vice versa.

(ii) If at least one of the two vectors $(\mathbf{P},\mathbf{Q})$, say $\mathbf{P}$, is nonnull, then $\mathbf{X} = \mathbf{P}$ and $\mathbf{Y} = \mathbf{Q} - (\tilde{\mathbf{P}}\mathbf{Q}/\mathbf{P}^2)\mathbf{P}$ are orthogonal. Further $\mathbf{Y}^2 = \mathbf{Q}^2 - (\tilde{\mathbf{P}}\mathbf{Q})^2/\mathbf{P}^2$, and this shows that when $\mathbf{Q}$ is null, $\mathbf{X}$ is timelike if $\mathbf{Y}$ is spacelike and vice versa. However, when $\mathbf{Q}$ is not null, $\mathbf{Y}$ may be timelike, spacelike or null depending on $(\mathbf{P},\mathbf{Q})$.

Thus, the generating vector pair $(\mathbf{X},\mathbf{Y})$, of a 2-flat through the origin, can always be chosen to be an orthogonal pair and we shall assume that it is so in the rest of this paper. In terms of an orthogonal pair $(\mathbf{X},\mathbf{Y})$, the equation to a 2-flat through the origin becomes

$$\mathbf{R} = \eta\mathbf{X} + \mu\mathbf{Y}; \quad \tilde{\mathbf{X}}\mathbf{Y} = 0. \tag{2.4}$$

It is easy to see that a general E_n^p admits orthogonal vector pairs of the following six types: spacelike-spacelike (ss), spacelike-null (sn), spacelike-timelike (st), null-null (nn) null-timelike (nt) and timelike-timelike (tt). For example in E_4^2, we have $[(0,0,i,0), (0,0,0,i)]$, $[(0,0,i,0), (1,0,0,i)]$, $[(0,0,i,0), (1,0,0,0)]$, $[(0,1,i,0), (1,0,0,i)]$, $[(1,0,0,i), (0,1,0,0)]$, $[(1,0,0,0), (0,1,0,0)]$ which are orthogonal vector pairs belonging to the types tt, nt, st, nn, sn, and ss respectively. However, we may observe that all these six types of orthogonal vector pairs exist only when g_{ij} has at least two positive terms and two negative terms, i.e., when $n \geqslant 2 + p \geqslant 4$. In particular, in E_n^{n-1}; $n \geqslant 3$, of which the Minkowski world E_4^3 is a special case, only the three types ss, sn and st, of orthogonal vector pairs, are possible. This result can be seen easily as follows: Since $p = n - 1$, there is only one timelike member in every orthonormal basis of the type described in Eq. (1.1) and it is evident that given any timelike vector $\mathbf{T}$, one can always choose an orthonormal basis in which the components of $\mathbf{T}$ are given by $\tilde{\mathbf{T}} = (0,0,...0,it)$. This form of $\mathbf{T}$ immediately shows that orthogonal vector pairs of the type tt and nt are impossible in E_n^{n-1}. Further, if $\mathbf{X}$ and $\mathbf{Y}$ are a pair of orthogonal vectors of which, say, $\mathbf{X}$ is null, then in a suitable orthonormal basis in which $\mathbf{X}$ has components given by $\tilde{\mathbf{X}} = (0,0,...0,x,ix)$ and $\mathbf{Y}$ has components given by $\tilde{\mathbf{Y}} = (y_1,y_2,...y_{n-1},iy_n)$, $\tilde{\mathbf{X}}\mathbf{Y} = 0$ implies $y_{n-1} = y_n$ and hence $\tilde{\mathbf{Y}}\mathbf{Y} = y_1^2 + y_2^2 + \cdots + y_{n-2}^2 \geqslant 0$. Therefore $\mathbf{Y}$ can only be spacelike, or if it is null, it is a constant multiple of $\mathbf{X}$ and this proves that orthogonal vector pairs of the type nn (and nt) are impossible in E_n^{n-1}. Thus, we see that in E_n^{n-1}, only three types of orthogonal vector pairs namely ss, st and sn, are possible.

We now prove that a 2-flat defined by Eq. (2.4) admits only one type of an orthogonal vector pair $(\mathbf{X},\mathbf{Y})$. Let $(\mathbf{X}',\mathbf{Y}')$ be a new pair of orthogonal vectors in the 2-flat defined by Eq. (2.4). Then they are related to $(\mathbf{X},\mathbf{Y})$ by

$$\mathbf{X}' = \eta\mathbf{X} + \mu\mathbf{Y}, \quad \mathbf{Y}' = \eta'\mathbf{X} + \mu'\mathbf{Y}, \quad \eta\mu' - \eta'\mu \neq 0. \tag{2.5}$$

Evidently we have

$$(\mathbf{X}')^2 = \eta^2\mathbf{X}^2 + \mu^2\mathbf{Y}^2, \quad (\mathbf{Y}')^2 = (\eta')^2\mathbf{X}^2 + (\mu')^2\mathbf{Y}^2, \tag{2.6}$$

and

$$\tilde{\mathbf{X}}'\mathbf{Y}' = \eta\eta'\mathbf{X}^2 + \mu\mu'\mathbf{Y}^2 = 0. \tag{2.7}$$

From these formulas it is evident that when $(\mathbf{X},\mathbf{Y})$ belongs to the types ss, nn or tt, $(\mathbf{X}',\mathbf{Y}')$ is also of the types ss, nn or tt respectively. When $(\mathbf{X},\mathbf{Y})$ is of the type st, in the nontrivial case in which η,μ,η' and μ' are all nonzero, Eqs. (2.6) and (2.7) yield

$$(\mathbf{X}')^2 = (\eta/\mu')(\eta\mu' - \eta'\mu)\mathbf{X}^2,$$
$$(\mathbf{Y}')^2 = -(\eta'/\mu)(\eta\mu' - \eta'\mu)\mathbf{X}^2, \tag{2.8}$$

which clearly show that $\mathbf{X}'$ and $\mathbf{Y}'$ are both non-null. Further, from Eq. (2.7) we also have

$$(\eta\eta'/\mu\mu') = (-\mathbf{Y}^2/\mathbf{X}^2) > 0, \tag{2.9}$$

as $(\mathbf{X},\mathbf{Y})$ is an st-pair of orthogonal vectors. Therefore (η/μ') and (η'/μ) are of the same sign and from Eq. (2.8) it now follows that $(\mathbf{X}',\mathbf{Y}')$ is also an orthogonal vector pair of the type st. In the two trivial cases in which either $\eta = \mu' = 0$ or $\eta' = \mu = 0$, this result is evident. Lastly, when $(\mathbf{X},\mathbf{Y})$ is of the types nt or sn, with, say, $\mathbf{X}$ as the null vector, we have from Eqs. (2.5)–(2.7),

$$(\mathbf{X}')^2 = \mu^2\mathbf{Y}^2, \quad (\mathbf{Y}')^2 = (\mu')^2\mathbf{Y}^2, \quad \mu\mu' = 0. \tag{2.10}$$

Both the cases $\mu = 0, \mu' \neq 0$ and $\mu \neq 0, \mu' = 0$ evidently lead again to an orthogonal vector pair $(\mathbf{X}',\mathbf{Y}')$ in which one of the vectors is null (and is a constant multiple of the original null vectors $\mathbf{X}$) and the other non-null vector has the same norm as that of the original non-null vector $\mathbf{Y}$.

Thus we see that every 2-flat defined by Eq. (2.4) admits precisely one type of an orthogonal vector pair only and hence can be characterized by the type of orthogonal vector pair it admits. This leads to a classification of these 2-flats into six types, corresponding to the six types of orthogonal vector pairs discussed above. These 2-flats may thus be designated as ss-2-flats, st-2-flats, etc.

It is also interesting to note the relation between these 2-flats and the *null-cone*

$$\tilde{\mathbf{R}}\mathbf{R} = 0, \tag{2.11}$$

passing through the origin. Evidently, the points at which the 2-flat of Eq. (2.4) intersects this null-cone are given by

$$\mathbf{R} = \eta\mathbf{X} + \mu\mathbf{Y}, \tag{2.12}$$

where the (η,μ) satisfy

$$\eta^2\mathbf{X}^2 + \mu^2\mathbf{Y}^2 = 0. \tag{2.13}$$

Thus we observe that tt and ss 2-flats intersect the null-cone at only one point, namely the origin with $\eta = \mu = 0$. sn or nt 2-flats touch the null-cone along a line and with $\mathbf{X}$ as the null vector in the orthogonal pair $(\mathbf{X},\mathbf{Y})$ we find the equation to this line of tangency to be

$$\mu = 0. \tag{2.14}$$

A st-2-flat cuts the null-cone along two lines given by

$$\eta = \pm(\mathbf{Y}^2/\mathbf{X}^2)^{1/2}\mu,$$

and and nn-2-flat lies entirely on the null-cone. These results may be compared with the corresponding results for the case of E^3_4 given in Synge.[1]

3. PLANAR ROTATIONS IN E^p_n

We know[2] that the linear homogeneous transformations $A = (A_{ij})$, in E^p_n, which leave the quadratic form

$$\tilde{\mathbf{X}}\mathbf{X} = x_1^2 + x_2^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_n^2, \tag{3.1}$$

invariant, form the pseudo-orthogonal group G^p_n. The transformation matrix A is evidently orthogonal, i.e.,

$$\tilde{A}A = E, \tag{3.2}$$

where E is the unit $n \times n$ matrix. Moreover, since the n-vectors $A\,\mathbf{X}$ and $\mathbf{X}$ must have the same structure, given in Eq. (1.2), the elements A_{ij} of A must satisfy certain *reality conditions*. On block-dividing A into the form

$$A = \begin{pmatrix} \mathscr{B} & i\mathscr{C} \\ i\mathscr{D} & \mathscr{F} \end{pmatrix}, \tag{3.3}$$

these reality conditions imply that $\mathscr{B}, \mathscr{C}, \mathscr{D}, \mathscr{F}$ are all real matrices of orders $p \times p$, $p \times (n-p)$, $(n-p) \times p$ and $(n-p) \times (n-p)$ respectively. The set of all such A with $\det(A) = +1$, forms a subgroup G'^p_n of G^p_n.

We now consider certain abelian subgroups of G'^p_n which may be interpreted as groups of *planar rotations*. Let $(\mathbf{X},\mathbf{Y})$ be a linearly independent orthogonal pair of vectors of E^p_n. Further, let us assume that $\mathbf{X}$ and $\mathbf{Y}$, whenever non-null, have been normalized to ± 1, so that in general the norms $\mathbf{X}^2$ and $\mathbf{Y}^2$ have only the values $\pm 1,0$, depending upon the nature of $\mathbf{X}$ and $\mathbf{Y}$. Then the skew-symmetric matrix

$$S = \mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}}; \quad \tilde{\mathbf{X}}\mathbf{Y} = 0, \tag{3.4}$$

is evidently nonzero, and the matrix

$$A = \exp(S\theta), \tag{3.5}$$

where θ is a scalar parameter, satisfies the conditions given in Eqs. (3.2) and (3.3). Moreover

$$\det(A) = \det[\exp(S\theta)] = +1, \tag{3.6}$$

as the trace of S is zero. Thus, the set of all such A defined by Eqs. (3.4)–(3.6), forms an abelian subgroup of G'^p_n. Obviously S is the corresponding infinitesimal transformation. The orthogonal vector pair $(\mathbf{X},\mathbf{Y})$ defining S also defines a 2-flat, of E^p_n, passing through the origin. If $\eta\mathbf{X} + \mu\mathbf{Y}$ is an arbitrary vector of this 2-flat (plane), then

$$S(\eta\mathbf{X} + \mu\mathbf{Y}) = \mu\mathbf{Y}^2\mathbf{X} - \eta\mathbf{X}^2\mathbf{Y}, \tag{3.7}$$

is again a vector of the same plane. If $\mathbf{Z}$, on the other hand, is a vector orthogonal to this plane, then we have

$$S\mathbf{Z} = 0; \quad (\tilde{\mathbf{X}}\mathbf{Z} = \tilde{\mathbf{Y}}\mathbf{Z} = 0). \tag{3.8}$$

Thus $A = \exp(S\theta)$ tranforms the plane [defined by $(\mathbf{X},\mathbf{Y})$] into itself and leaves invariant any vector orthogonal to it. Moreover, since A preserves the norm of a vector and $\det(A) = +1$, it may be regarded as a *planar rotation* (*2-flat rotation*).

We now evaluate the planar rotation matrices for each of the six types of 2-flats discussed in Sec. 2. For this, we note that

$$S^2 = (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})(\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}}) = -\mathbf{Y}^2\mathbf{X}\tilde{\mathbf{X}} - \mathbf{X}^2\mathbf{Y}\tilde{\mathbf{Y}}, \tag{3.9}$$

$$S^3 = -(\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})(\mathbf{Y}^2\mathbf{X}\tilde{\mathbf{X}} + \mathbf{X}^2\mathbf{Y}\tilde{\mathbf{Y}}) = -\mathbf{X}^2\mathbf{Y}^2S. \tag{3.10}$$

Using these in Eq. (3.5), we obtain six particular forms of $A = \exp(S\theta)$, corresponding to the six types of planes as follows.

In an nn-2-flat, with $\mathbf{X}^2 = \mathbf{Y}^2 = 0$, we have

$$A = E + S\theta = E + (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})\theta. \tag{3.11}$$

It is interesting to note that this transformation, though not an identity transformation, leaves all vectors in the nn-2-flat unaltered. In an nt-2-flat, with $\mathbf{X}^2 = 0$ and $\mathbf{Y}^2 = -1$, we have

$$\begin{aligned} A &= E + S\theta + \tfrac{1}{2}S^2\theta^2 \\ &= E + (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})\theta + \tfrac{1}{2}\mathbf{X}\tilde{\mathbf{X}}\theta^2. \end{aligned} \tag{3.12}$$

In an sn-2-flat, with $\mathbf{X}^2 = 0$ and $\mathbf{Y}^2 = 1$, we have

$$\begin{aligned} A &= E + S\theta + \tfrac{1}{2}S^2\theta^2 \\ &= E + (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})\theta - \tfrac{1}{2}\mathbf{X}\tilde{\mathbf{X}}\theta^2. \end{aligned} \tag{3.13}$$

In a ss-2flat, with $\mathbf{X}^2 = \mathbf{Y}^2 = 1$, we have

$$\begin{aligned} A &= E + S\sin\theta + S^2(1 - \cos\theta) \\ &= E + (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})\sin\theta - (\mathbf{X}\tilde{\mathbf{X}} + \mathbf{Y}\tilde{\mathbf{Y}})(1 - \cos\theta). \end{aligned} \tag{3.14}$$

In a st-2-flat, with $\mathbf{X}^2 = 1$ and $\mathbf{Y}^2 = -1$, we have

$$\begin{aligned} A &= E + S\sinh\theta + S^2(\cosh\theta - 1) \\ &= E + (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})\sinh\theta + (\mathbf{X}\tilde{\mathbf{X}} - \mathbf{Y}\tilde{\mathbf{Y}})(\cosh\theta - 1). \end{aligned} \tag{3.15}$$

In a tt-2-flat, with $\mathbf{X}^2 = \mathbf{Y}^2 = -1$, we have

$$\begin{aligned} A &= E + S\sin\theta + S^2(1 - \cos\theta) \\ &= E + (\mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}})\sin\theta + (\mathbf{X}\tilde{\mathbf{X}} + \mathbf{Y}\tilde{\mathbf{Y}})(1 - \cos\theta). \end{aligned} \tag{3.16}$$

We may now define the *angle of rotation* φ as the angle between an arbitrary non-null vector lying in the plane and its image under the rotation. To determine the angle between two non-null vectors, we use the definition given in Petrov[3] and set

$$\cos\varphi = \pm \tilde{\mathbf{P}}\mathbf{Q}/(|\mathbf{P}^2||\mathbf{Q}^2|)^{1/2}, \tag{3.17}$$

where φ is angle between the n-vectors $\mathbf{P}$ and $\mathbf{Q}$ and the plus sign is taken when both $\mathbf{P}^2$ and $\mathbf{Q}^2$ are positive and the minus sign when both of them are negative. Obviously, this definition breaks down when a null vector is involved. If we now set $\mathbf{P} = \eta\mathbf{X} + \mu\mathbf{Y}$ and $\mathbf{Q} = A\,\mathbf{P}$ in Eq. (3.17), we get

$$\cos\varphi = \pm \tilde{\mathbf{P}}A\,\mathbf{P}/|\mathbf{P}^2|, \tag{3.18}$$

where we must choose the plus sign if $\mathbf{P}$ is spacelike and the minus sign if $\mathbf{P}$ is timelike. An angle of rotation φ is evidently defined by this formula for planar rotations in all 2-flats except the nn-2-flat. Using formulas (3.12)–(3.16) we find that $\varphi = 0$ for all *rotations* in the sn and nt 2-flats, $\varphi = \theta$ for rotations in the ss and tt 2-flats and $\varphi = i\theta$, a pure imaginary angle, for rotations in the st 2-flats. Lastly we note that the formulas in Eqs. (3.11)–(3.16) yield the following known special cases.

(i) Let $\mathbf{X} = \mathbf{I}_k$ and $\mathbf{Y} = \mathbf{I}_l$ be the unit vectors defining the $k - l$ coordinate 2-flat of E^p_n. Then we obtain from the formulas (3.14)–(3.16), the following nonzero components of $A = (A_{ij})$:

$$\left.\begin{array}{c} A_{kk} = A_{ll} = \cos\varphi, \quad A_{kl} = -A_{lk} = \sin\varphi, \\ \text{all other } A_{mm} = 1, \end{array}\right\} \tag{3.19}$$

where $\varphi = \theta$, $\varphi = -\theta$, or $\varphi = i\theta$ according as the coordinate 2-flat considered is ss, tt or st. This special form justifies the terminology rotations for the planar transformation A.

(ii) The rotation matrix for a rotation through a (real) angle θ about the axis $\hat{\mathbf{a}} = (a_1,a_2,a_3)$ in the Euclidean space $E_3 \equiv E_3^3$, in the more familiar form as given in Jeffreys and Jeffreys,[4] follows from Eq. (3.14) on writing $\mathbf{X}\times\mathbf{Y} = \hat{\mathbf{a}}$ and we get

$$A = (A_{ij}) = \begin{pmatrix} a_1^2(1-\cos\theta)+\cos\theta & a_1a_2(1-\cos\theta)+a_3\sin\theta & a_1a_3(1-\cos\theta)-a_2\sin\theta \\ a_2a_1(1-\cos\theta)-a_3\sin\theta & a_2^2(1-\cos\theta)+\cos\theta & a_2a_3(1-\cos\theta)+a_1\sin\theta \\ a_3a_1(1-\cos\theta)+a_2\sin\theta & a_3a_2(1-\cos\theta)-a_1\sin\theta & a_3^2(1-\cos\theta)+\cos\theta \end{pmatrix}. \tag{3.20}$$

This gives the well-known relation

$$A_{11}+A_{22}+A_{33} = 1+2\cos\theta, \tag{3.21}$$

for the determination of the angle of rotation for a given A, and a somewhat more specific form

$$A_{ij} - A_{ji} = 2a_k\sin\theta, \quad (i,j,k = 1,2,3 \text{ cyclic}), \tag{3.22}$$

to the relations, as given in Hamermesh[2] or Wigner,[5] for the determination of the axis of rotation. Further, the form of A given by Eq. (3.20) [and also Eq. (3.14) directly] shows that A will be symmetric for $\theta = \pi$ and assumes the form

$$\begin{pmatrix} 2a_1^2-1 & 2a_1a_2 & 2a_1a_3 \\ 2a_2a_1 & 2a_2^2-1 & 2a_2a_3 \\ 2a_3a_1 & 2a_3a_2 & 2a_3^2-1 \end{pmatrix}. \tag{3.23}$$

Then Eq. (3.22) becomes a trivial identity leaving the axis $\hat{\mathbf{a}}$ undetermined. One has now to solve for the a_k from the elements of some row or column of the matrix (3.23), as for example from

$$A_{11} = 2a_1^2 - 1, \quad A_{12} = 2a_1a_2, \quad A_{13} = 2a_1a_3. \tag{3.24}$$

In this case, $\det(E + A) = 0$ and A would be what Weyl[6] calls an *exceptional* matrix since such matrices do not fit directly into Cayley's parametrization of the rotation group SO(3) and some effort is necessary, as Weyl puts it, to "render these exceptions ineffective."

We note that in the exceptional case, $\hat{\mathbf{a}}$ as given by Eq. (3.24) is determined only up to an ambiguity in sign and this reflects the fact that in the topological representation of the rotation group by a sphere of radius π, diametrically opposite points are to be identified. As an example of an exceptional rotation matrix, we have

$$A = \frac{1}{3}\begin{pmatrix} 0 & \sqrt{6} & \sqrt{3} \\ \sqrt{6} & -1 & \sqrt{2} \\ \sqrt{3} & \sqrt{2} & -2 \end{pmatrix}, \tag{3.25}$$

which evidently has $\theta = \pi$ and

$$\hat{\mathbf{a}} = \pm\left(1/\sqrt{2}, 1/\sqrt{3}, 1/\sqrt{6}\right). \tag{3.26}$$

4. PROPER LORENTZ TRANSFORMATIONS IN E_4^3

As an interesting application of the results of the previous section we devote the rest of this paper to a complete discussion of the physically important case of proper Lorentz transformations in the Minkowski world E_4^3. We show that any general *nonsingular* (or *non-null*) Lorentz transformation may be factored into a *commuting* product of two *planar* transformations one of which is *equivalent*, by a proper Lorentz trasformation, to a *pure rotation* and the other to a *pure boost*. We accordingly call these factor transformations *rotationlike* and *boostlike*. We also note that this resolution is a *Lorentz invariant* one and is to be contrasted with the other known result (see for example Anderson[7]) that a Lorentz transformation may be expressed as a (noncommuting) product of a pure-rotation and a pure-boost. A *singular* (or *null*) Lorentz transformation, on the other hand, is shown to be factorizable into a product of two rotationlike transformations, each with a rotation angle π. These factor transformations are *symmetric* in Minkowski coordinates with $x_4 = ict$, and show certain special features like the symmetric rotations of Sec. 3 and are therefore termed *exceptional* following Weyl. These do not appear to have been noticed in the literature. Our analysis also yields the necessary and sufficient conditions for a Lorentz transformation to be planar and leads to an invariant classification of Lorentz transformations of all types. We note here that a different scheme of classification and a prescriptive procedure for determining the angles and planes of rotation based on the antisymmetric part of a Lorentz transformation has been given by Bazanski[8] who also gives a formula which is essentially the same as our Eq. (4.39). Since, however, a pure boost is symmetric in real coordinates and an exceptional transformation in Minkowski coordinates, one or the other has its antisymmetric part identically zero and is naturally excluded in his classification. Our procedure, on the other hand, covers all cases and yields *explicit formulas* for the angles and planes of rotation in terms of the elements of a given Lorentz matrix L. Moreover, based as it is on group-theoretical considerations, as we shall see in the next section, our method yields an explicit formula for the three-dimensional complex orthogonal representation of the Lorentz group SO(3,1) which in turn leads to two other classifications schemes in terms of the characters $\chi(L)$ in the D^{10} and $D^{\frac{1}{2}0}$ representations.

It may be observed that there is a close analogy between the methods adopted here and those of electromagnetic theory because the algebra of the infinitesimal Lorentz transformations is the same as that of the electromagnetic field tensor. In the case of the nonexceptional (null as well as non-null) transformations, the analogy is with the reduction of an electromagnetic field at an event to its canonical form whereas in the exceptional case it is with the problem of the extraction of an extremal root of the electromagnetic energy tensor as in the RMW theory.[9]

A proper Lorentz transformation in E_4^3 is represented by the 4×4 matrix $L=(L_{ij}), i,j=1,2,3,4$, with $L_{\alpha4}$ and $L_{4\alpha}$ for $\alpha=1,2,3$ purely imaginary while all other elements are real and

$$\tilde{L}L=L\tilde{L}=E, \quad \det L=+1, \tag{4.1}$$

where E is the 4×4 unit matrix. Note that we are following the conventions described in Sec. 1 and for E_4^3 this simply means that we are employing the conventional Minkowski coordinates with $x_4=ict$. Since the proper Lorentz group is a Lie group, every proper Lorentz transformation L may be written as

$$L=\exp(I), \tag{4.2}$$

where the infinitesimal transformation I has the form[1]

$$I=\begin{pmatrix} 0 & h_3 & -h_2 & ie_1 \\ -h_3 & 0 & h_1 & ie_2 \\ h_2 & -h_1 & 0 & ie_3 \\ -ie_1 & -ie_2 & -ie_3 & 0 \end{pmatrix}. \tag{4.3}$$

We observe that I has precisely the same structure as the electromagnetic field tensor with 3-vector fields $\mathbf{e}$ and $\mathbf{h}$. In this case we may look upon the Lie parameters e_1, e_2, e_3 and h_1, h_2, h_3 as the components of two 3-vectors $\mathbf{e}$ and $\mathbf{h}$ which we shall call the *parameter vectors* of L. The eigenvalues of I are thus[1] $\pm i\theta_r$, $\pm\theta_b$, where

$$\theta_r^2=\tfrac{1}{2}(\mathbf{h}^2-\mathbf{e}^2)+[\tfrac{1}{4}(\mathbf{h}^2-\mathbf{e}^2)^2+(\mathbf{h}\cdot\mathbf{e})^2]^{1/2}, \tag{4.4}$$

$$\theta_b^2=-\tfrac{1}{2}(\mathbf{h}^2-\mathbf{e}^2)+[\tfrac{1}{4}(\mathbf{h}^2-\mathbf{e}^2)^2+(\mathbf{h}\cdot\mathbf{e})^2]^{1/2}, \tag{4.5}$$

and those of L are $\exp(\pm i\theta_r)$, $\exp(\pm\theta_b)$. A Lorentz transformation for which $\mathbf{e}=\mathbf{h}=0$ is evidently *trivial* (the identity transformation). There are two types of *nontrivial* Lorentz transformations. If the two invariants $\mathbf{h}\cdot\mathbf{e}$ and $(\mathbf{h}^2-\mathbf{e}^2)$ are both zero, then the eigenvalues of I are all zero and hence those of L are all equal to $+1$. Such a Lorentz transformation is called *singular* (*null*) by Synge[1] and its I corresponds to a null electromagnetic field. A Lorentz transformation is *nonsingular* (*non-null*) if at least one of $(\mathbf{h}^2-\mathbf{e}^2)$ and $\mathbf{h}\cdot\mathbf{e}$ is nonzero. The "singularity" of L arises here from the fact that $\det I=0$ although $\det L$ itself is $+1$.

We now invoke a basic result of electromagnetic theory (see for example Synge[1] or Landau and Lifshitz[10]) that there exist Lorentz frames in which a non-null electromagnetic field has its electric and magnetic vectors parallel. Adapted to our case, this means that if at least one of $(\mathbf{h}^2-\mathbf{e}^2)$ and $\mathbf{h}\cdot\mathbf{e}$ is different from zero, there exists a Lorentz transformation, T, $\tilde{T}T=E$ such that

$$I'=TI\tilde{T}=\begin{bmatrix} 0 & \theta_r & 0 & 0 \\ -\theta_r & 0 & 0 & 0 \\ 0 & 0 & 0 & i\theta_b \\ 0 & 0 & -i\theta_b & 0 \end{bmatrix}, \tag{4.6}$$

where the common direction of the parameter vectors $\mathbf{e}'$ and $\mathbf{h}'$ of I' has been chosen to be the z axis (of the new frame) rather than the x axis as chosen by Synge[1] in his discussion of the "geometry" of the electromagnetic field. This shows, in complete analogy with the four-dimensional rotation matrix,[11,12] that every nonsingular Lorentz transformation may be brought to the canonical block-diagonal form (4-*screw*)

$$L'=TL\tilde{T}=T(\exp I)\tilde{T}=\exp(TI\tilde{T})=\exp(I')$$

$$=\begin{bmatrix} \cos\theta_r & \sin\theta_r & 0 & 0 \\ -\sin\theta_r & \cos\theta_r & 0 & 0 \\ 0 & 0 & \cosh\theta_b & i\sinh\theta_b \\ 0 & 0 & -i\sinh\theta_b & \cosh\theta_b \end{bmatrix}, \tag{4.7}$$

by a proper Lorentz transformation. We have thus given a direct proof of Synge's theorem[13] that *every nonsingular proper Lorentz transformation is equivalent to a 4-screw.* Writing $L'=R'B'=B'R'$, where

$$R'=\begin{bmatrix} \cos\theta_r & \sin\theta_r & 0 & 0 \\ -\sin\theta_r & \cos\theta_r & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.8}$$

and

$$B'=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cosh\theta_b & i\sinh\theta_b \\ 0 & 0 & -i\sinh\theta_b & \cosh\theta_b \end{bmatrix}, \tag{4.9}$$

we see that R' is a *pure rotation* and B' is a *pure boost.* Evidently R' is a *planar* Lorentz transformation in the ss-2-flat defined by the 4-vectors

$$\tilde{\mathbf{X}}'=(1,0,0,0), \quad \tilde{\mathbf{Y}}'=(0,1,0,0), \tag{4.10}$$

and the corresponding angle of rotation is θ_r. Similarly, B' is a planar transformation in the st 2-flat defined by the 4-vectors

$$\tilde{\mathbf{Z}}'=(0,0,1,0), \quad \tilde{\mathbf{W}}'=(0,0,0,i), \tag{4.11}$$

and has the angle of rotation $i\theta_b$. We thus have, in terms of the matrices of the original basis,

$$L=RB=BR, \quad R=\tilde{T}R'T, \quad B=\tilde{T}B'T, \tag{4.12}$$

where R is equivalent to a pure rotation and B to a pure boost by the Lorentz transformation T and have, respectively, the same invariant angles θ_r and $i\theta_b$. We say that R is *rotationlike* and B is *boostlike.* Moreover, R' and B' and consequently R and B are planar transformations since the latter are obtained from the former by a mere change of basis.

We observe from Eq. (4.6) that in the canonical basis (primed letters denote quantities in the canonical basis),

$$I'=\theta_r S_r'+\theta_b S_b', \tag{4.13}$$

where

$$S_r'\equiv\mathbf{X}'\tilde{\mathbf{Y}}'-\mathbf{Y}'\tilde{\mathbf{X}}'=\begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{4.14}$$

and

$$S_b'\equiv\mathbf{Z}'\tilde{\mathbf{W}}'-\mathbf{W}'\tilde{\mathbf{Z}}'=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \end{bmatrix}, \tag{4.15}$$

with $(\mathbf{X}',\mathbf{Y}',\mathbf{Z}',\mathbf{W}')$ defined in Eqs. (4.10) and (4.11), are the infinitesimal transformations in the $\mathbf{X}'-\mathbf{Y}'$ and $\mathbf{Z}'-\mathbf{W}'$

planes, respectively. Thus we get

$$I = \theta_r S_r + \theta_b S_b; \quad S_r = \tilde{T} S_r' T, \quad S_b = \tilde{T} S_b' T, \quad (4.16)$$

for the infinitesimal transformations in the original basis. Here the suffixes r and b in Eqs. (4.13)–(4.16) are mere labels indicating rotation and boost respectively and *no* summation is implied by these repeated indices. We follow the same convention throughout the paper. From Eq. (4.16) it follows that

$$S_r = \mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}}, \text{ and } S_b = \mathbf{Z}\tilde{\mathbf{W}} - \mathbf{W}\tilde{\mathbf{Z}}, \quad (4.17)$$

where (**X**,**Y**,**Z**,**W**) are the inverse images of the orthonormal tetrad of vectors (**X**′,**Y**′,**Z**′,**W**′) under the transformation T, i.e., $\mathbf{X} = \tilde{T}\mathbf{X}'$, $\mathbf{Y} = \tilde{T}\mathbf{Y}'$ etc. Thus, in the original basis, a nonsingular Lorentz transformation L is characterized by the two invariant angles θ_r and $i\theta_b$ and the orthonormal tetrad (**X**,**Y**,**Z**,**W**) giving the two blades (planes) of the transformation. We now proceed to determine these quantities in terms of the elements, L_{ij} of the given Lorentz transformation L.

We know that the characteristic equation[14] of any $n \times n$ matrix A is given by

$$\sum_{r=0}^{n} (-1)^r \lambda^{n-r} p_r = 0, \quad (4.18)$$

where p_r is the sum of all rth order principal minors in the determinant of A. Thus, we have for L,

$$\lambda^4 - \chi\lambda^3 + \xi\lambda^2 - \chi\lambda + 1 = 0, \quad (4.19)$$

where

$$\chi = L_{11} + L_{22} + L_{33} + L_{44} = \mathrm{spur} L, \quad (4.20)$$

and ξ is the sum of all principal minors of the second order. In Eq. (4.19) the coefficient of $(-\lambda)$ is also χ because each element of a proper orthogonal matrix A ($A\tilde{A} = E$) is equal to its cofactor. More generally, *any minor of an orthogonal matrix is equal to its algebraic complement* by Jacobi's theorem and hence it is sufficient to compute only three distinct second order minors of L to obtain ξ. Since the roots of Eq. (4.19) are already known to be $\exp(\pm i\theta_r)$ and $\exp(\pm\theta_b)$, we have from Vieta's formulas (see for example, Kurosh[15])

$$2\cos\theta_r + 2\cosh\theta_b = \chi, \quad (4.21)$$

$$2 + 4\cos\theta_r \cosh\theta_b = \xi, \quad (4.22)$$

and hence

$$\cosh\theta_b = \tfrac{1}{4}(\chi + \sigma), \quad (4.23)$$

$$\cos\theta_r = \tfrac{1}{4}(\chi - \sigma), \quad (4.24)$$

where

$$\sigma \equiv +(\chi^2 - 4\xi + 8)^{1/2}. \quad (4.25)$$

These equations determine θ_r and θ_b explicitly in terms of L_{ij}. Since θ_r is actually a rotation angle [see Eq. (4.8)], we take it to lie between 0 and π. Similarly we take θ_b to be positive. From Eq. (4.22) we have

$$1 + \tfrac{1}{2}\xi = 2 + 2\cos\theta_r \cosh\theta_b,$$

and hence it follows that

$$(\chi - 1 - \tfrac{1}{2}\xi) = 2(\cosh\theta_b - 1)(1 - \cos\theta_r) \geqslant 0,$$

where the equality holds if and only if at least one of θ_r or θ_b is zero. But when θ_r or θ_b is zero, the transformation is evidently planar and thus

$$\chi = 1 + \tfrac{1}{2}\xi, \quad (4.26)$$

is the necessary and sufficient condition for a proper Lorentz transformation to be planar. Further, if $\theta_r = 0$ and $\theta_b \neq 0$, we have $\chi > 4$ while if $\theta_r \neq 0$ and $\theta_b = 0$, $\chi < 4$. For $\theta_r = \theta_b = 0$, we have $\chi = 4$. We thus have the following invariant classification of proper Lorentz transformations: *A proper Lorentz transformation is nonplanar or planar according as* $\chi \gtrless 1 + \frac{1}{2}\xi$, *and a planar transformation is rotationlike, singular, or boostlike according as* $\chi \lesseqgtr 4$. In particular, a rotation-like transformation L with $\theta_r = \pi$ has some special features due to the fact that $\det(L + E) = 0$ for it and we call such a transformation *exceptional* following Weyl's characterization of similar transformations of the rotation group SO(3). *A nonplanar proper Lorentz transformation* (*with* $\chi > 1 + \frac{1}{2}\xi$) *can always be written as a commuting product of two planar transformations one of which is rotation-like and the other boost-like.*

Next, we consider the problem of expressing the two planes associated with L in terms of its elements. Let $S = \mathbf{U}\tilde{\mathbf{V}} - \mathbf{V}\tilde{\mathbf{U}}$ be the infinitesimal transformation in the 2-flat determined by the orthogonal pair (**U**,**V**). As before, let us assume that **U** and **V**, whenever non-null, have been normalized to ± 1. One of these vectors, say **U**, must necessarily be spacelike since a null or timelike vector cannot be orthogonal to another null or timelike vector. We now show that in the plane defined by the orthogonal pair (**U**,**V**), there always exists another orthogonal pair (**P**,**Q**) in which the spacelike vector **P** has its temporal component equal to zero. Consider first, the case in which both **U** and **V** are spacelike. Then (**P**,**Q**) given by

$$\mathbf{P} = \mathbf{U}\cos\varphi - \mathbf{V}\sin\varphi,$$

$$\mathbf{Q} = \mathbf{U}\sin\varphi + \mathbf{V}\cos\varphi,$$

are again orthogonal and yield the same S. If we now choose φ such that $\tan\varphi = (u_4/v_4)$, then p_4 would be zero (u_4, v_4 and p_4 are respectively the time-components of **U**,**V** and **P**). Next, if **V** is timelike, we consider the vector pair (**P**,**Q**) given by

$$\mathbf{P} = \mathbf{U}\cosh\varphi - \mathbf{V}\sinh\varphi,$$

$$\mathbf{Q} = -\mathbf{U}\sinh\varphi + \mathbf{V}\cosh\varphi.$$

This pair (**P**,**Q**) is also orthogonal with **P** spacelike and **Q** timelike and yields the same S. But p_4 would be zero only if we can choose a φ such that $\tanh\varphi = (u_4/v_4)$. Since, however, $|\tanh\varphi| \leqslant 1$, this choice is possible only if $|u_4/v_4| \leqslant 1$. This is certainly true, because

$$\begin{aligned} u_4^2 v_4^2 &= (u_1 v_1 + u_2 v_2 + u_3 v_3)^2 \\ &\leqslant (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) \\ &= (1 + u_4^2)(-1 + v_4^2) \\ &= -1 - u_4^2 + v_4^2 + u_4^2 v_4^2, \end{aligned}$$

so that we have $v_4^2 - u_4^2 \geqslant 1$. Finally if **V** is a null vector, we consider the new orthonormal pair (**P**,**Q**) defined by

$$\mathbf{P} = \mathbf{U} - k\mathbf{V},$$

$$\mathbf{Q} = \mathbf{V},$$

which again lead to the same S. Evidently we may choose k such that $p_4 = 0$. Thus we may assume, without loss of generality, that every infinitesimal Lorentz transformation of the form

$$S = \mathbf{U}\tilde{\mathbf{V}} - \mathbf{V}\tilde{\mathbf{U}}, \tag{4.27}$$

is such that the spacelike vector $\mathbf{U}$ has $u_4 = 0$. Since $\mathbf{U}$ is a unit vector, it follows that

$$\tilde{\mathbf{U}} = (\hat{\mathbf{u}},0), \tag{4.28}$$

where $\hat{\mathbf{u}}$ is a unit 3-vector. We have assumed that when $\mathbf{V}$ is a non-null vector, $\tilde{\mathbf{V}}\mathbf{V} = \pm 1$ and in the case of a null $\mathbf{V}$ it is convenient to assume that it is "normalized" in the sense

$$v_1^2 + v_2^2 + v_3^2 = v_4^2 = 1, \quad \tilde{\mathbf{V}}\mathbf{V} = 0. \tag{4.29}$$

Now we proceed to determine such a pair $(\mathbf{U},\mathbf{V})$ corresponding to a given planar infinitesimal transformation S whose parameter vectors satisfy the conditions

$$\mathbf{h}\cdot\mathbf{e} = 0, \quad \mathbf{h}^2 - \mathbf{e}^2 = +1, \quad -1, \text{ or } 0. \tag{4.30}$$

With $\tilde{\mathbf{U}} = (\hat{\mathbf{u}},0)$ as given by Eq. (4.28) and $\tilde{\mathbf{V}} = (\mathbf{v},iv_4)$, Eqs. (4.27) and (4.3) give

$$(\hat{\mathbf{u}}\times\mathbf{v}) = \mathbf{h}, \quad \hat{\mathbf{u}}v_4 = \mathbf{e}. \tag{4.31}$$

This shows that the given S must have $\mathbf{h}\cdot\mathbf{e} = 0$ in agreement with Eq. (4.30). Moreover, since $\mathbf{U}$ and $\mathbf{V}$ are orthogonal we have

$$\hat{\mathbf{u}}\cdot\mathbf{v} = 0 \tag{4.32}$$

and this, in conjunction with Eqs. (4.30) and (4.31), implies

$$\mathbf{h}\times\hat{\mathbf{e}} = \mathbf{v} \text{ and } v_4 = |\mathbf{e}|. \tag{4.33}$$

We thus have, for the orthonormal pair $(\mathbf{U},\mathbf{V})$,

$$\tilde{\mathbf{U}} = (\hat{\mathbf{e}},0), \quad \tilde{\mathbf{V}} = (\mathbf{h}\times\hat{\mathbf{e}},i|\mathbf{e}|). \tag{4.34}$$

Observe that

$$|\mathbf{h}\times\hat{\mathbf{e}}|^2 - |\mathbf{e}|^2 = |\mathbf{h}|^2|\hat{\mathbf{e}}|^2 - |\mathbf{e}|^2 = \mathbf{h}^2 - \mathbf{e}^2,$$

so that $\mathbf{V}$ is spacelike or timelike according as $(\mathbf{h}^2 - \mathbf{e}^2) = \pm 1$, respectively, and is null when $(\mathbf{h}^2 - \mathbf{e}^2) = 0$. In the latter case we may factor out the common magnitude $\theta = |\mathbf{e}| = |\mathbf{h}|$ from S and write it as $S = \theta S_1$ where the parameter vectors of S_1 are unit 3-vectors. We then have

$$S_1 = \mathbf{U}\tilde{\mathbf{V}} - \mathbf{V}\tilde{\mathbf{U}}, \quad \tilde{\mathbf{U}} = (\hat{\mathbf{e}},0), \quad \tilde{\mathbf{V}} = (\hat{\mathbf{h}}\times\hat{\mathbf{e}},i). \tag{4.35}$$

If the given S has $\mathbf{e} = 0$, then Eq. (4.34) fails to give $(\mathbf{U},\mathbf{V})$. S is then characterized only by $\mathbf{h}$ with $\mathbf{h}^2 = 1$ and we take $v_4 = 0$ in accordance with Eq. (4.31). The 3-vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{h}$ would then form an orthonormal triplet and we choose $(\mathbf{U},\mathbf{V})$ as

$$\left.\begin{aligned}\tilde{\mathbf{U}} &= (h_1^2 + h_2^2)^{-1/2}[h_2, -h_1,0,0],\\ \tilde{\mathbf{V}} &= (h_1^2 + h_2^2)^{-1/2}[-h_1h_3, -h_2h_3, h_1^2 + h_2^2,0]\end{aligned}\right\}. \tag{4.36}$$

We thus note that the Eqs. (4.34)–(4.36) give the "orthonormal" pair $(\mathbf{U},\mathbf{V})$ explicitly in terms of the parameter vectors of an infinitesimal transformation S which satisfies the conditions given in Eq. (4.30). The problem is completely solved on showing that we can always extract two planar transformations S_r and S_b, both of which satisfy the conditions given in Eq. (4.30), from a given Lorentz transformation L. Then the above formulas for $(\mathbf{U},\mathbf{V})$ determine the two 2-flats spanning the blades of L.

A. Nonexceptional, nonsingular transformations

It follows from Eqs. (4.14) and (4.15) that the parameter vectors of S_r' and S_b' satisfy

$$\mathbf{e}_r'\cdot\mathbf{h}_r' = 0, \quad \mathbf{e}_b'\cdot\mathbf{h}_b' = 0, \quad \mathbf{h}_r'^2 - \mathbf{e}_r'^2 = \mathbf{e}_b'^2 - \mathbf{h}_b'^2 = 1.$$

We also observe that

$$S_r'^3 = -S_r', \quad S_b'^3 = S_b', \quad S_r'S_b' = S_b'S_r' = 0.$$

These relations, being invariant, are also true in the original basis and we have

$$\mathbf{e}_r\cdot\mathbf{h}_r = \mathbf{e}_b\cdot\mathbf{h}_b = 0, \quad \mathbf{h}_r^2 - \mathbf{e}_r^2 = \mathbf{e}_b^2 - \mathbf{h}_b^2 = 1, \tag{4.37a}$$

$$S_r^3 = -S_r, \quad S_b^3 = S_b, \quad S_rS_b = S_bS_r = 0. \tag{4.37b}$$

Therefore

$$\begin{aligned}L &= \exp(I) = \exp(\theta_r S_r + \theta_b S_b)\\ &= \exp(\theta_r S_r)\cdot\exp(\theta_b S_b) = RB = BR.\end{aligned} \tag{4.38}$$

where

$$R = \exp(\theta_r S_r) = E + S_r\sin\theta_r + S_r^2(1 - \cos\theta_r)$$

and

$$B = \exp(\theta_b S_b) = E + S_b\sinh\theta_b + S_b^2(\cosh\theta_b - 1).$$

On multiplying these and using Eq. (4.37), we obtain

$$\begin{aligned}L = E &+ S_r\sin\theta_r + S_b\sinh\theta_b\\ &+ S_r^2(1 - \cos\theta_r) + S_b^2(\cosh\theta_b - 1),\end{aligned} \tag{4.39}$$

which is a polynomial of the second degree in the planar infinitesimal transformations S_r and S_b. We now proceed to determine S_r and S_b from the given L, or what is the same thing, we obtain explicit expressions for the parameter vectors $\mathbf{e}_r$, $\mathbf{h}_r$, $\mathbf{e}_b$ and $\mathbf{h}_b$ in terms of the elements L_{ij} of L. Formulas given in Eqs. (4.34)–(4.36) will then give us the planes of rotation.

Let

$$L_A = \begin{bmatrix} 0 & \mathscr{H}_3 & -\mathscr{H}_2 & i\mathscr{E}_1 \\ -\mathscr{H}_3 & 0 & \mathscr{H}_1 & i\mathscr{E}_2 \\ \mathscr{H}_2 & -\mathscr{H}_1 & 0 & i\mathscr{E}_3 \\ -i\mathscr{E}_1 & -i\mathscr{E}_2 & -i\mathscr{E}_3 & 0 \end{bmatrix}, \tag{4.40}$$

where $2\mathscr{H}_\gamma = L_{\alpha\beta} - L_{\beta\alpha}$, $(\alpha,\beta,\gamma = 1,2,3$ cyclic), and $2\mathscr{E}_\alpha = L_{\alpha 4} - L_{4\alpha}$, $(\alpha = 1,2,3)$, be the antisymmetric part of L. Equating the antisymmetric parts on both sides of Eq. (4.39) we have

$$S_r\sin\theta_r + S_b\sinh\theta_b = L_A,$$

i.e.,

$$\mathbf{e}_r\sin\theta_r + \mathbf{e}_b\sinh\theta_b = \mathscr{E}, \tag{4.41}$$

$$\mathbf{h}_r\sin\theta_r + \mathbf{h}_b\sinh\theta_b = \mathscr{H}. \tag{4.42}$$

The relation $S_rS_b = S_bS_r$ leads to the following relations between the parameter vectors:

$$\mathbf{e}_r\times\mathbf{e}_b = \mathbf{h}_r\times\mathbf{h}_b \equiv \mathbf{m}, \tag{4.43}$$

$$\mathbf{e}_r\times\mathbf{h}_b = \mathbf{e}_b\times\mathbf{h}_r \equiv \mathbf{n}. \tag{4.44}$$

Forming the several expressions for $\mathbf{m}^2$, $\mathbf{n}^2$ and $\mathbf{m}\cdot\mathbf{n}$, we obtain

$$\begin{aligned}\mathbf{m}^2 = \mathbf{e}_r^2\mathbf{e}_b^2 - (\mathbf{e}_r\cdot\mathbf{e}_b)^2 &= -(\mathbf{e}_r\cdot\mathbf{h}_b)(\mathbf{e}_b\cdot\mathbf{h}_r)\\ &= \mathbf{h}_r^2\mathbf{h}_b^2 - (\mathbf{h}_r\cdot\mathbf{h}_b)^2,\end{aligned} \tag{4.45}$$

$$\begin{aligned}\mathbf{n}^2 = (\mathbf{e}_r\cdot\mathbf{e}_b)(\mathbf{h}_r\cdot\mathbf{h}_b) &= \mathbf{e}_r^2\mathbf{h}_b^2 - (\mathbf{e}_r\cdot\mathbf{h}_b)^2\\ &= \mathbf{e}_b^2\mathbf{h}_r^2 - (\mathbf{e}_b\cdot\mathbf{h}_r)^2,\end{aligned} \tag{4.46}$$

$$\mathbf{m}\cdot\mathbf{n} = -(\mathbf{e}_b\cdot\mathbf{e}_r)(\mathbf{e}_r\cdot\mathbf{h}_b) = (\mathbf{e}_r\cdot\mathbf{e}_b)(\mathbf{e}_b\cdot\mathbf{h}_r), \tag{4.47}$$

$$= -(\mathbf{e}_r\cdot\mathbf{h}_b)(\mathbf{h}_b\cdot\mathbf{h}_r) = (\mathbf{e}_b\cdot\mathbf{h}_r)(\mathbf{h}_r\cdot\mathbf{h}_b). \tag{4.48}$$

If $(\mathbf{e}_r\cdot\mathbf{e}_b)\neq 0$, Eq. (4.47) gives $(\mathbf{e}_b\cdot\mathbf{h}_r) = -(\mathbf{e}_r\cdot\mathbf{h}_b)$ and substituting this in Eq. (4.46) we get $\mathbf{e}_r^2\mathbf{h}_b^2 = \mathbf{e}_b^2\mathbf{h}_r^2$. But this would give, on using Eqs. (4.37), $\mathbf{e}_r^2 = -\mathbf{e}_b^2$, which is impossible unless $\mathbf{e}_r = \mathbf{e}_b = 0$ leading again to a contradiction. Thus we must have

$$\mathbf{e}_r\cdot\mathbf{e}_b = 0. \tag{4.49}$$

Similarly we obtain

$$\mathbf{h}_r\cdot\mathbf{h}_b = 0. \tag{4.50}$$

Equations (4.49) and (4.46) now show that $\mathbf{n} = 0$, i.e., $\mathbf{e}_r$ is parallel to $\mathbf{h}_b$ and $\mathbf{e}_b$ is parallel to $\mathbf{h}_r$. But the two expressions for $\mathbf{m}$ show that while $\mathbf{e}_b$ is parallel to $\mathbf{h}_r$, $\mathbf{h}_b$ is antiparallel to $\mathbf{e}_r$. Further, Eq. (4.45) yields $\mathbf{e}_r^2\mathbf{e}_b^2 = \mathbf{h}_r^2\mathbf{h}_b^2$ and on using Eq. (4.37) we get $\mathbf{e}_b^2 = \mathbf{h}_r^2$ and $\mathbf{e}_r^2 = \mathbf{h}_b^2$ leading to

$$\mathbf{e}_b = \mathbf{h}_r, \quad \mathbf{h}_b = -\mathbf{e}_r. \tag{4.51}$$

These relations show that S_r and S_b are *duals* of each other. Substituting Eq. (4.51) in Eqs. (4.41) and (4.42) and solving, we finally obtain

$$\mathbf{h}_r = (\sin^2\theta_r + \sinh^2\theta_b)^{-1}(\mathscr{E}\sinh\theta_b + \mathscr{H}\sin\theta_r) = \mathbf{e}_b, \tag{4.52}$$

$$\mathbf{e}_r = (\sin^2\theta_r + \sinh^2\theta_b)^{-1}(\mathscr{E}\sin\theta_r - \mathscr{H}\sinh\theta_b) = -\mathbf{h}_b. \tag{4.53}$$

Since θ_r and θ_b are known from Eqs. (4.23)–(4.25), these equations give the parameter vectors directly in terms of the elements L_{ij} of L. In view of the properties given in Eq. (4.37), the formula given in Eq. (4.34) [or (4.36) if $\mathbf{e}_r$ or $\mathbf{e}_b$ is zero] now determines the orthonormal pair $(\mathbf{X}\equiv\mathbf{U},\mathbf{Y}\equiv\mathbf{V})$ yielding S_r as

$$S_r = \mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}}.$$

Similarly we obtain the orthonormal pair $(\mathbf{Z}\equiv\mathbf{U},\mathbf{W}\equiv\mathbf{V})$ yielding S_b as

$$S_b = \mathbf{Z}\tilde{\mathbf{W}} - \mathbf{W}\tilde{\mathbf{Z}}.$$

It also follows that *the Lorentz transformation T, which sends the given L into the canonical block-diagonal form given in Eq.* (4.7), *has for its rows, the row-vectors* $(\tilde{\mathbf{X}},\tilde{\mathbf{Y}},\tilde{\mathbf{Z}}, -i\tilde{\mathbf{W}})$.

Although θ_r and θ_b are given by Eqs. (4.23)–(4.25), we may obtain expressions directly for $\sin\theta_r$ and $\sinh\theta_b$ from Eqs. (4.52) and (4.53). On using the relations $\mathbf{h}_r\cdot\mathbf{e}_r = 0$ and $\mathbf{h}_r^2 - \mathbf{e}_r^2 = 1$, we can solve for $\sin^2\theta_r$ and $\sinh^2\theta_b$ and we obtain

$$\sin^2\theta_r = \tfrac{1}{2}(\mathscr{H}^2 - \mathscr{E}^2) + [\tfrac{1}{4}(\mathscr{H}^2 - \mathscr{E}^2)^2 + (\mathscr{H}\cdot\mathscr{E})^2]^{1/2}, \tag{4.54}$$

$$\sinh^2\theta_b = -\tfrac{1}{2}(\mathscr{H}^2 - \mathscr{E}^2) + [\tfrac{1}{4}(\mathscr{H}^2 - \mathscr{E}^2)^2 + (\mathscr{H}\cdot\mathscr{E})^2]^{1/2}. \tag{4.55}$$

We observe that the whole procedure breaks down in two particular cases. First, if the vectors $\mathscr{H}$ and $\mathscr{E}$ of L_A are such that $\mathscr{H}\cdot\mathscr{E} = 0$ and $\mathscr{H}^2 - \mathscr{E}^2 = 0$, then $\sin^2\theta_r = \sinh^2\theta_b = 0$ and Eq. (4.52) and (4.53) become meaningless. We shall see that this corresponds to the singular Lorentz transformation for which $\theta_r = \theta_b = 0$. The second case is that of the exceptional Lorentz transformation for which L_A is identically zero and the procedure evidently breaks down. It will be seen that $\theta_r = \pi$ and $\theta_b = 0$ in this case, and we shall consider each of these cases separately, later.

From Eqs. (4.54) and (4.55), we see that $\sin\theta_r = (\mathscr{H}^2 - \mathscr{E}^2)^{1/2}$ and $\theta_b = 0$ if $\mathscr{H}\cdot\mathscr{E} = 0$ and $\mathscr{H}^2 > \mathscr{E}^2$. The corresponding transformation is rotation-like and we obtain from Eqs. (4.52) and (4.53)

$$\mathbf{h}_r = (\mathscr{H}^2 - \mathscr{E}^2)^{-1/2}\mathscr{H} = \mathbf{e}_b, \tag{4.56}$$

$$\mathbf{e}_r = (\mathscr{H}^2 - \mathscr{E}^2)^{-1/2}\mathscr{E} = -\mathbf{h}_b. \tag{4.57}$$

If on the other hand, $\mathscr{H}\cdot\mathscr{E} = 0$ and $\mathscr{E}^2 > \mathscr{H}^2$, we have a boostlike transformation with $\theta_r = 0$, $\sinh\theta_b = (\mathscr{E}^2 - \mathscr{H}^2)^{1/2}$ and

$$\mathbf{e}_b = (\mathscr{E}^2 - \mathscr{H}^2)^{-1/2}\mathscr{E} = \mathbf{h}_r, \tag{4.58}$$

$$\mathbf{h}_b = (\mathscr{E}^2 - \mathscr{H}^2)^{-1/2}\mathscr{H} = -\mathbf{e}_r. \tag{4.59}$$

Equations (4.56) and (4.57) determine the blades in the former case wherein $\theta_b = 0$ in the st-blade. Similarly Eqs. (4.58) and (4.59) give the blades in the latter case with $\theta_r = 0$ in the ss-blade.

B. Singular Lorentz transformations

In this case, the parameter vectors of the infinitesimal transformation satisfy

$$\mathbf{h}^2 - \mathbf{e}^2 = 0, \quad \mathbf{h}\cdot\mathbf{e} = 0, \tag{4.60}$$

so that we have $\theta_r = \theta_b = 0$ and each eigenvalue of L is $+1$. The infinitesimal transformation I [see Eq. (4.3)] now corresponds to a null electromagnetic field and all that one can do is to carry out a rotation T of the spatial axes such that the vector $\mathbf{h}$, say, is along the z axis and $\mathbf{e}$ is along the x axis. Denoting the common magnitude of $\mathbf{e}$ and $\mathbf{h}$ by θ, we have the canonical form

$$I' = TI\tilde{T} = \theta\begin{bmatrix} 0 & 1 & 0 & i \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -i & 0 & 0 & 0 \end{bmatrix} \equiv \theta S', \tag{4.61}$$

where

$$S' = \mathbf{X}'\tilde{\mathbf{N}}' - \mathbf{N}'\tilde{\mathbf{X}}', \quad \tilde{\mathbf{X}}' = (1,0,0,0), \quad \tilde{\mathbf{N}}' = (0,1,0,i), \tag{4.62}$$

with $\mathbf{X}'$ evidently spacelike and $\mathbf{N}'$ null. We also have

$$(S')^3 = 0, \tag{4.63}$$

so that we have in the original basis [see Eq. (3.13)]

$$L = \exp(\theta S) = E + \theta S + \tfrac{1}{2}\theta^2 S^2, \tag{4.64}$$

as a planar transformation in the sn 2-flat determined by

$$\mathbf{X} = \tilde{T}\mathbf{X}', \mathbf{N} = \tilde{T}\mathbf{N}' \text{ and}$$

$$S = \mathbf{X}\tilde{\mathbf{N}} - \mathbf{N}\tilde{\mathbf{X}} = \tilde{T}S'T. \tag{4.65}$$

Since the eigenvalues of L are all $+1$, we also have $\chi = 1 + \frac{1}{2}\xi = 4$ as expected. On equating θS to the antisymmetric part L_A of L, we obtain, because of Eq. (4.64),

$$\theta\mathbf{e} = \mathscr{E}, \quad \theta\mathbf{h} = \mathscr{H}, \tag{4.66}$$

where $\mathbf{e}$ and $\mathbf{h}$ are the parameter vectors of S and $\mathscr{E}$ and $\mathscr{H}$ are the vectors occurring in L_A. Since $\mathbf{e}$ and $\mathbf{h}$ satisfy Eq. (4.60), we must have

$$\mathscr{E}^2 - \mathscr{H}^2 = 0, \quad \mathscr{E}\cdot\mathscr{H} = 0. \tag{4.67}$$

On the other hand, since T is merely a spatial rotation of the form

$$T = \begin{pmatrix} \mathscr{R} & 0 \\ 0 & 1 \end{pmatrix} \tag{4.68}$$

where $\mathscr{R}$ is a three-dimensional spatial rotation matrix, it transforms the 3-dimensional vector $\mathbf{e}$ occurring in S into $\mathbf{e}' = \mathscr{R}\mathbf{e}$ of S'. Since $\mathbf{e}'$ is a unit vector as is clear from Eq. (4.61), $\mathbf{e} = \hat{\mathbf{e}}$ and hence $\mathbf{h} = \hat{\mathbf{h}}$ are unit vectors. Therefore we have

$$\theta = |\mathscr{E}| = |\mathscr{H}|, \quad \hat{\mathbf{e}} = \hat{\mathscr{E}}, \quad \hat{\mathbf{h}} = \hat{\mathscr{H}}, \tag{4.69}$$

and the spacelike-null vector pair $(\mathbf{X},\mathbf{N})$ determining the plane of the singular transformation is given by, in view of Eq. (4.35),

$$\tilde{\mathbf{X}} = (\hat{\mathscr{E}},0), \quad \hat{\mathbf{N}} = (\hat{\mathscr{H}} \times \hat{\mathscr{E}}, i), \tag{4.70}$$

expressed in terms of the elements of L. It is clear that the Lorentz transformation (rotation) T sending I to the canonical form I' of Eq. (4.61) is

$$T = \begin{bmatrix} \hat{\mathscr{E}}_1 & \hat{\mathscr{E}}_2 & \hat{\mathscr{E}}_3 & 0 \\ (\hat{\mathscr{H}}\times\hat{\mathscr{E}})_1 & (\hat{\mathscr{H}}\times\hat{\mathscr{E}})_2 & (\hat{\mathscr{H}}\times\hat{\mathscr{E}})_3 & 0 \\ \hat{\mathscr{H}}_1 & \hat{\mathscr{H}}_2 & \hat{\mathscr{H}}_3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.71}$$

where $\hat{\mathscr{E}}_\alpha$ and $\hat{\mathscr{H}}_\alpha$ are the components of the corresponding unit 3-vectors.

C. Exceptional Lorentz transformations

We have called a rotationlike planar transformation with $\theta_r = \pi$ and $\theta_b = 0$, an *exceptional Lorentz transformation*, and observe that it is *symmetric* as Eq. (4.39) with $\theta_r = \pi$ and $\theta_b = 0$ reduces to

$$L = E + 2S_r^2, \tag{4.72}$$

where S_r^2 is symmetric by virtue of the antisymmetry of S_r. We note, however, that an exceptional transformation is symmetric only in Minkowski coordinates and in a real coordinate system it is represented by an asymmetric matrix. Conversely, we show that we must have $\theta_r = \pi$ and $\theta_b = 0$ for a symmetric Lorentz transformation. Taking the symmetric L to be

$$L = \begin{bmatrix} L_{11} & q_3 & q_2 & ip_1 \\ q_3 & L_{22} & q_1 & ip_2 \\ q_2 & q_1 & L_{33} & ip_3 \\ ip_1 & ip_2 & ip_3 & L_{44} \end{bmatrix}, \tag{4.73}$$

we have

$$\chi = L_{11} + L_{22} + L_{33} + L_{44},$$

and

$$\begin{aligned} \tfrac{1}{2}\xi &= (L_{11}L_{44} + p_1^2) + (L_{22}L_{44} + p_2^2) + (L_{33}L_{44} + p_3^2) \\ &= L_{44}(\chi - L_{44}) + (L_{44}^2 - 1) = \chi L_{44} - 1, \end{aligned}$$

so that

$$1 + \tfrac{1}{2}\xi = \chi L_{44}. \tag{4.74}$$

But since L is symmetric ($L = \tilde{L}$), we have $L\tilde{L} = L^2 = E$ so that the eigenvalues of L are equal to ± 1 only. Since $\det(L) = +1$, all the eigenvalues must either be equal to $+1$, or while two of them are equal to $+1$, the other two must be equal to -1. In the former case $\chi = 4$ and in the latter $\chi = 0$. If $\chi = 4$, then Eq. (4.74) shows that $\xi = 8L_{44} - 2$ and thus we have $\sigma^2 = \chi^2 - 4\xi + 8 = 32(1 - L_{44})$. But $L_{44} \geqslant 1$ for a proper transformation and hence we must have $L_{44} = 1$ and $\sigma = 0$ as σ^2 can never be negative. The vanishing of σ implies that $\theta_r = \theta_b = 0$ [see Eqs. (4.23) and (4.24)]. Since L is orthogonal, $L_{44} = 1$ implies that $p = 0$ and L is then a pure rotation with $\theta_r = 0$ which simply means that L is only the trivial identity transformation. On the other hand, if $\chi = 0$, we have $\xi = -2$ and $\sigma = 4$ leading to $\theta_r = \pi$ and $\theta_b = 0$. Thus we have proved that *any symmetric Lorentz transformation (in Minkowski coordinates with $x_4 = ict$) must necessarily be exceptional.*

We now proceed to determine the plane of rotation of an exceptional Lorentz transformation L. Since $L_A \equiv 0$ for a symmetric L, Eqs. (4.52) and (4.53) evidently fail to determine the parameter vectors $\mathbf{e}_r$ and $\mathbf{h}_r$. The problem now is to determine an antisymmetric S_r satisfying Eq. (4.72) from a symmetric L with $L^2 = E$ and it is exactly analogous to the problem of the extraction of an "extremal root" of the electromagnetic energy tensor.[9]

If an orthonormal pair of vectors $(\mathbf{X},\mathbf{Y})$ generate the plane of the transformation, which is an ss-2-flat, then

$$S_r = \mathbf{X}\tilde{\mathbf{Y}} - \mathbf{Y}\tilde{\mathbf{X}},$$

and

$$S_r^2 = -\mathbf{X}\tilde{\mathbf{X}} - \mathbf{Y}\tilde{\mathbf{Y}}. \tag{4.75}$$

Therefore, $S_r^2\mathbf{X} = -\mathbf{X}$ and $S_r^2\mathbf{Y} = -\mathbf{Y}$ and it follows from Eq. (4.72) that $\mathbf{X}$ and $\mathbf{Y}$ are also eigenvectors of L belonging to the eigenvalue -1. If the given L has an especially simple structure, then one may determine a pair of orthonormal eigenvectors of L belonging to the eigenvalue -1 without much effort. Although the problem may be regarded as solved in principle by this prescription, we adopt another procedure, in conformity with our purpose of determining these vectors explicitly in terms of the elements L_{ij} of L.

With $\mathbf{h}_r \equiv \mathbf{h} = (h_1,h_2,h_3)$ and $\mathbf{e}_r \equiv \mathbf{e} = (e_1,e_2,e_3)$ as the parameter vectors of S_r and $L = (L_{ij})$, Eq. (4.72) implies the following relations in which the Greek suffixes take the run of values 1,2,3, and the sequence α,β,γ is cyclic in 1,2,3,

$$L_{\alpha\alpha} = 1 + 2(e_\alpha^2 + h_\alpha^2 - \mathbf{h}^2), \tag{4.76}$$

$$L_{44} = 1 + 2\mathbf{e}^2, \tag{4.77}$$

$$q_\gamma = L_{\alpha\beta} = L_{\beta\alpha} = 2(h_\alpha h_\beta + e_\alpha e_\beta), \quad \alpha \neq \beta, \tag{4.78}$$

$$p_\alpha = L_{4\alpha} = L_{\alpha 4} = 2(\mathbf{e}\times\mathbf{h})_\alpha. \tag{4.79}$$

On using $\mathbf{h}^2 - \mathbf{e}^2 = 1$ and $\mathbf{h}\cdot\mathbf{e} = 0$, we obtain from Eqs. (4.76) and (4.77)

$$e_\alpha^2 + h_\alpha^2 = \tfrac{1}{2}(L_{\alpha\alpha} + L_{44}), \tag{4.80}$$

while from Eqs. (4.78) and (4.79), we obtain

$$4e_\alpha h_\alpha = -q_\beta p_\beta + q_\gamma p_\gamma. \tag{4.81}$$

Equations (4.80) and (4.81) now yield

$$2e_\alpha = \pm(f_\alpha \pm g_\alpha), \tag{4.82}$$

$$2h_\alpha = \pm(f_\alpha \mp g_\alpha), \tag{4.83}$$

where

$$f_\alpha = 2^{-1/2}(L_{44} + L_{\alpha\alpha} - q_\beta p_\beta + q_\gamma p_\gamma)^{1/2}, \tag{4.84}$$

$$g_\alpha = 2^{-1/2}(L_{44} + L_{\alpha\alpha} + q_\beta p_\beta - q_\gamma p_\gamma)^{1/2}, \tag{4.85}$$

giving **e** and **h** explicitly in terms of the elements of L. Since we obtain the same L on replacing S_r by $-S_r$ in Eq. (4.72), the vectors **h** and **e** are determined only up to an indeterminacy in sign and we may take $2h_\alpha = f_\alpha \mp g_\alpha$ in Eq. (4.83) and the solution of our problem will then be the h_α and e_α that are also consistent with Eq. (4.81) and the relation $\mathbf{h}^2 - \mathbf{e}^2 = 1$. The formulas of Eq. (4.34) now give the orthonormal spacelike vectors $(\mathbf{X},\mathbf{Y})$ determining the plane of rotation.

Lastly, we prove a result concerning a resolution of singular transformations. We saw that any nonplanar, nonsingular transformation may be written as a commuting product of two planar transformations one of which is rotationlike while the other is boostlike. We shall now obtain a somewhat similar result for singular transformations, namely, that *every singular transformation is the product of two exceptional transformations.*

From Eq. (4.61), we have, for a singular L,

$$TL\tilde{T} = L' = \exp(l')$$
$$= \begin{bmatrix} 1 & \theta & 0 & i\theta \\ -\theta & 1-\frac{1}{2}\theta^2 & 0 & -i\theta^2/2 \\ 0 & 0 & 1 & 0 \\ -i\theta & -i\theta^2/2 & 0 & 1+\frac{1}{2}\theta^2 \end{bmatrix}. \tag{4.86}$$

But we may express L' as

$$L' = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} -1 & -\theta & 0 & -i\theta \\ -\theta & 1-\frac{1}{2}\theta^2 & 0 & -i\theta^2/2 \\ 0 & 0 & -1 & 0 \\ -i\theta & -i\theta^2/2 & 0 & 1+\frac{1}{2}\theta^2 \end{bmatrix}$$
$$\equiv L'_1 L'_2, \tag{4.87}$$

and observe that both L'_1 and L'_2 are symmetric and therefore exceptional. We thus obtain the resolution

$$L = L_1 L_2, \tag{4.88}$$

where $L_1 = \tilde{T} L'_1 T$ and $L_2 = \tilde{T} L'_2 T$ are evidently exceptional proving the result stated above. Moreover, since the T as given in Eq. (4.71) has a particularly simple structure, we may even give L_1 and L_2 explicitly. It is easy to check that

$$\left.\begin{aligned} (L_1)_{\alpha\beta} &= (L_1)_{\beta\alpha} = \hat{p}_\alpha \hat{p}_\beta - \hat{\mathscr{E}}_\alpha \hat{\mathscr{E}}_\beta - \hat{\mathscr{H}}_\alpha \hat{\mathscr{H}}_\beta, \\ (L_1)_{\alpha 4} &= (L_1)_{4\alpha} = 0, \quad (L_1)_{44} = 1, \end{aligned}\right\} \tag{4.89}$$

$$\left.\begin{aligned} (L_2)_{\alpha\beta} &= (L_2)_{\beta\alpha} = \hat{p}_\alpha \hat{p}_\beta (1 - \tfrac{1}{2}\theta^2) - (\hat{p}_\alpha \hat{\mathscr{E}}_\beta + \hat{p}_\beta \hat{\mathscr{E}}_\alpha)\theta \\ &\quad - (\hat{\mathscr{E}}_\alpha \hat{\mathscr{E}}_\beta + \hat{\mathscr{H}}_\alpha \hat{\mathscr{H}}_\beta), \\ (L_2)_{4\alpha} &= (L_2)_{\alpha 4} = -i\hat{\mathscr{E}}_\alpha \theta - i\hat{p}_\alpha \theta^2/2, \quad (L_2)_{44} = 1 + \tfrac{1}{2}\theta^2, \end{aligned}\right\} \tag{4.90}$$

where $\hat{\mathbf{p}} = \hat{\mathscr{H}} \times \hat{\mathscr{E}}$, $\theta = |\mathscr{E}| = |\mathscr{H}|$ and $\hat{p}_\alpha$, $\hat{\mathscr{E}}_\alpha$ and $\hat{\mathscr{H}}_\alpha$ are the components of the respective unit vectors.

D. Examples

As an illustration of the foregoing discussion, we consider a few simple examples of Lorentz transformations.

Example (i): Consider first,

$$L = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & \gamma & 0 & i(\gamma^2-1)^{1/2} \\ 1 & 0 & 0 & 0 \\ 0 & -i(\gamma^2-1)^{1/2} & 0 & \gamma \end{bmatrix}, \tag{4.91}$$

where $\gamma > 1$. Clearly $\chi = 2\gamma$ and $\xi/2 = 1$ so that $\chi > 1 + \frac{1}{2}\xi$ and $\sigma = 2\gamma$, and the transformation is nonplanar with $\theta_r = \pi/2$ and $\cosh\theta_b = \gamma$. Forming the antisymmetric part of L, we get $\mathscr{H} = (0,1,0)$, $\mathscr{E} = (0,(\gamma^2-1)^{1/2},0)$ and hence from Eqs. (4.52) and (4.53), $\mathbf{h}_r = \mathbf{e}_b = (0,1,0)$. The formulas of Eq. (4.34) then give

$$\tilde{\mathbf{X}} = (1,0,0,0), \quad \tilde{\mathbf{Y}} = (0,0,1,0),$$
$$\tilde{\mathbf{Z}} = (0,1,0,0), \quad \tilde{\mathbf{W}} = (0,0,0,i),$$

where the first two vectors determine the plane of $\theta_r = \pi/2$ and the last two determine the plane of $\theta_b = \cosh^{-1}\gamma$, a fact which is at once evident from the structure of L.

Example (ii): As a second example, consider

$$L = \begin{bmatrix} 1+(\gamma-1)v_x^2/v^2 & (\gamma-1)v_x v_y/v^2 & (\gamma-1)v_x v_z/v^2 & i\gamma v_x/c \\ (\gamma-1)v_y v_x/v^2 & 1+(\gamma-1)v_y^2/v^2 & (\gamma-1)v_y v_z/v^2 & i\gamma v_y/c \\ (\gamma-1)v_z v_x/v^2 & (\gamma-1)v_z v_y/v^2 & 1+(\gamma-1)v_z^2/v^2 & i\gamma v_z/c \\ -i\gamma v_x/c & -i\gamma v_y/c & -i\gamma v_z/c & \gamma \end{bmatrix}, \tag{4.92}$$

where $\gamma \equiv (1 - v^2/c^2)^{-1/2} > 1$, representing a pure boost with velocity $\mathbf{v} = (v_x, v_y, v_z)$ (see Møller[16] or Synge[1]). Here $\chi = 2\gamma + 2$ and $\frac{1}{2}\xi = 2\gamma + 1$ showing, as expected, that the transformation is planar and it is *boostlike* since $\chi > 4$. Moreover, $\sigma^2 = \chi^2 - 4\xi + 8 = 4(\gamma-1)^2$ so that $\sigma = 2(\gamma - 1)$. Hence $\cos\theta_r = \frac{1}{4}(\chi - \sigma) = 1$, i.e., $\theta_r = 0$ and $\cosh\theta_b = \frac{1}{4}(\chi + \sigma) = \gamma$. Forming the antisymmetric part of L, we get $\mathscr{H} = 0, \mathscr{E} = \gamma\mathbf{v}/c$, which yield through Eqs. (4.52) and (4.53) $\mathbf{e}_b = \gamma(\gamma^2-1)^{-1/2}\mathbf{v}/c$, $\mathbf{h}_b = 0$. But $|\mathbf{e}_b| = 1$ as $\mathbf{v}^2 = c^2(\gamma^2-1)\gamma^{-2}$ and hence the 4-vectors defining the plane of the boost [see Eq. (4.34)] are $\tilde{\mathbf{Z}} = (v_x/v, v_y/v, v_z/v, 0)$ and $\tilde{\mathbf{W}} = (0,0,0,i)$.

Example (iii): Another interesting example is

$$L = \begin{bmatrix} 0 & \gamma & 0 & i(\gamma^2-1)^{1/2} \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -i(\gamma^2-1)^{1/2} & 0 & \gamma \end{bmatrix}, \tag{4.93}$$

where $\gamma > 1$, and clearly the transformation is planar as $\chi = \gamma + 1$ and $\frac{1}{2}\xi = \gamma$. However it is rotationlike, singular or boostlike according as $\gamma \lesseqgtr 3$. But $\sigma^2 = (\gamma - 3)^2$ so that we must take $\sigma = 3 - \gamma$ or $\sigma = \gamma - 3$ according as $\gamma \lesssim 3$. When $\gamma < 3$, we have $\sigma = 3 - \gamma$, $\cos\theta_r = \frac{1}{2}(\gamma - 1)$ and $\theta_b = 0$, for the resulting rotationlike transformation. When $\gamma > 3$, we have $\sigma = \gamma - 3$ and the resulting boostlike transformation has $\theta_r = 0$ and $\cosh\theta_b = \frac{1}{2}(\gamma - 1)$. When $\gamma = 3$, $\sigma = 0$, $\theta_r = \theta_b = 0$ and the transformation is evidently singular. The vectors of the skew-symmetric part of L are $\mathscr{E} = \frac{1}{2}(\gamma^2 - 1)^{1/2}(1,1,0)$ and $\mathscr{H} = (0,0,\frac{1}{2}(\gamma + 1))$. Equations (4.34) and (4.56)–(4.59) now yield the vectors defining the planes of the transformations to be

$$\tilde{\mathbf{U}} = 2^{-1/2}[1,1,0,0],$$

and

$$\tilde{\mathbf{V}} = \begin{cases} (2(3-\gamma))^{-1/2}[-(\gamma+1)^{1/2},(\gamma+1)^{1/2},0,i(4\gamma-4)^{1/2}]; & \gamma < 3, \\ 2^{-1/2}[-1,1,0,i2^{1/2}]; & \gamma = 3, \\ (2(\gamma-3))^{-1/2}[-(\gamma+1)^{1/2},(\gamma+1)^{1/2},0,i(4\gamma-4)^{1/2}]; & \gamma > 3. \end{cases}$$

We observe that the L in question is obtained by carrying out a rotation through $\pi/2$ in the x-y plane followed by a boost along the x axis with a velocity v corresponding to γ and it is interesting that the composite transformation L could be any of the three types depending on γ and stays rotationlike up to a velocity as high as $c(8/9)^{1/2}$.

Example (iv): We finally consider one simple example of an exceptional transformation. For the symmetric matrix

$$L = \begin{bmatrix} 1-2\gamma^2 & 0 & 0 & 2i\gamma(\gamma^2-1)^{1/2} \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 2i\gamma(\gamma^2-1)^{1/2} & 0 & 0 & 2\gamma^2-1 \end{bmatrix}, \tag{4.94}$$

we evidently have $\chi = 0$, $\frac{1}{2}\xi = -1$ so that $\chi = 1 + \frac{1}{2}\xi = 0$ and hence we have $\theta_r = \pi$ and $\theta_b = 0$ in this case and the orthonormal vectors defining the plane of the rotation are obtained most simply as the eigenvectors of L belonging to the eigenvalue -1. It is easy to verify that

$$\tilde{\mathbf{U}} = (\gamma,0,0,-i\gamma\beta), \quad \tilde{\mathbf{V}} = (0,1,0,0),$$

are the orthonormal vectors belonging to the eigenvalue -1 of L spanning the plane of rotation. We note that we would have obtained the same vectors by the second method wherein the parameter vectors which follow from Eqs. (4.82) and (4.83) are $\mathbf{h} = (0,0,\gamma)$ and $\mathbf{e} = (0,\gamma\beta,0)$.

5. CLASSIFICATION OF LORENTZ TRANSFORMATIONS IN THE D^{10} AND $D^{\frac{1}{2}0}$ REPRESENTATIONS

We now give two other classification schemes of Lorentz transformations in terms of the characters $\chi_A(L)$ and $\chi_C(L)$ of L, where $\chi_A(L)$ is the character of the element L of SO(3,1) in its three dimensional complex orthogonal representation D^{10} (or D^{01}) and $\chi_C(L)$ is its character in the two-dimensional complex unimodular representation $D^{\frac{1}{2}0}$ (or $D^{0\frac{1}{2}}$). We shall do this by explicitly constructing from a given L, a complex orthogonal matrix A which follows as an immediate consequence of the relations in Eqs. (4.16), (4.41), and (4.42). It is of interest to observe in this connection that Landau and Lifshitz[10] make use of the idea that a Lorentz transformation may be regarded as a rotation through a complex angle in three-dimensional space and give the transformation that corresponds to a velocity along the x axis. Our matrix A is just the appropriate generalization of their transformation to an arbitrary Lorentz transformation. Writing $\mathbf{h}_r = \mathbf{e}_b = \boldsymbol{\alpha}$; $\mathbf{e}_r = -\mathbf{h}_b = \boldsymbol{\beta}$ and denoting $\mathbf{h} + i\mathbf{e}$ by $\mathbf{f}$, it follows from Eqs. (4.3) and (4.16), that

$$\mathbf{f} = \mathbf{h} + i\mathbf{e} = (\theta_r + i\theta_b)(\boldsymbol{\alpha} + i\boldsymbol{\beta}). \tag{5.1}$$

We claim that the skew-symmetric matrix

$$\mathscr{I} = \begin{bmatrix} 0 & f_3 & -f_2 \\ -f_3 & 0 & f_1 \\ f_2 & -f_1 & 0 \end{bmatrix}, \tag{5.2}$$

is the infinitesimal transformation in the D^{10} representation of SO(3,1). To see this, we observe that, if $I_{\alpha\beta}$, $I_{\gamma 4}$ ($\alpha,\beta,\gamma = 1,2,3$ cyclic) are the infinitesimal transformations of the self-representation of SO(3,1) in the coordinate planes, $I_{\alpha\beta}$, $-iI_{\gamma 4}$ would be the corresponding ones for the rotation group SO(4) which is a direct product of two three-dimensional rotation groups whose infinitesimal transformations are $J_\gamma = \frac{1}{2}(I_{\alpha\beta} - iI_{\gamma 4})$ and $K_\gamma = \frac{1}{2}(I_{\alpha\beta} + iI_{\gamma 4})$ so that we have the well-known relations

$$D^{jj'}(I_{\alpha\beta}) = D^{j}(J_\gamma)\times D^{j'}(1) + D^{j}(1)\times D^{j'}(K_\gamma), \quad D^{jj'}(-iI_{\gamma 4}) = D^{j}(J_\gamma)\times D^{j'}(1) - D^{j}(1)\times D^{j'}(K_\gamma). \tag{5.3}$$

Noting that $D^0(J) = D^0(K) = 0$ and $D^0(1) = 1$, we get

$$D^{10}(I_{\alpha\beta}) = D^{10}(-iI_{\gamma 4}) = D^{1}(J_\gamma), \tag{5.4}$$

and taking J_γ in the self-representation of the rotation group SO(3) we see that Eq. (5.4) yields for the representation matrix of I, exactly the same matrix $\mathscr{I}$ of Eq. (5.2). We thus have $D^{10}(I) = \mathscr{I}$, and obtain

$$A = D^{10}(L) = \exp\mathscr{I} = \exp(\theta S_3) = E_3 + \sin\theta\, S_3 + (1-\cos\theta)S_3^2, \tag{5.5}$$

TABLE I. Classification schemes of proper Lorentz transformations.

L / Representation	General nonplanar (Screw-like)	Planar transformations: Boostlike	Singular (Null)	Rotationlike
$D^{1/2\,1/2}$:SO(3,1)	$\frac{1}{2}\xi + 1 < \chi$	$\frac{1}{2}\xi + 1 = \chi > 4$	$\frac{1}{2}\xi + 1 = \chi = 4$	$\frac{1}{2}\xi + 1 = \chi < 4; \chi = 0$ Exceptional
D^{10}:SO(3,C)	χ_A complex or $\chi_A < -1$	$\chi_A > 3$	$\chi_A = 3$	$\chi_A < 3; \chi_A = -1$ Exceptional
$D^{\frac{1}{2}0}$:SL(2,C)	χ_C complex	$\chi_C > 2$ or $\chi_C < -2$	$\chi_C = \pm 2$	$-2 < \chi_C < 2; \chi_C = 0$ Exceptional

where $\theta = \theta_r + i\theta_b$ and S_3 is the three-dimensional skew-symmetric matrix constructed from $\mathbf{a} = \boldsymbol{\alpha} + i\boldsymbol{\beta}$. We notice that A of Eq. (5.5) has determinant $+1$ and has exactly the same structure as the A of Eq. (3.20) with the only difference that $\mathbf{a}$ and θ are here complex so that A is complex orthogonal and unimodular. Since $\boldsymbol{\alpha}^2 - \boldsymbol{\beta}^2 = 1$ and $\boldsymbol{\alpha}\cdot\boldsymbol{\beta} = 0$, $\mathbf{a}$ is a unit vector and is an eigenvector of A of Eq. (5.5) belonging to the eigenvalue $+1$ and may thus be regarded as the *complex axis of rotation*. We have from Eqs. (4.41) and (4.42)

$$\hat{\mathbf{a}} = \boldsymbol{\alpha} + i\boldsymbol{\beta} = (\sin\theta_r + i\sinh\theta_b)^{-1}(\mathscr{H} + i\mathscr{E}) = (\mathscr{H}^2 - \mathscr{E}^2 + 2i\mathscr{H}\cdot\mathscr{E})^{-1/2}(\mathscr{H} + i\mathscr{E}), \tag{5.6}$$

and $\theta = \cos^{-1}((\chi - \sigma)/4) + i\cosh^{-1}((\chi + \sigma)/2)$ from Eqs. (4.23) and (4.24) so that all the elements of $A = D^{10}(L)$ are expressed explicitly in terms of the given Lorentz transformation L. We note that, since the proper Lorentz group SO(3,1) and the complex orthogonal group SO(3,C) are both six-parameter groups and $I \longleftrightarrow \mathscr{I}$ is a one-one mapping, there exists exactly one complex orthogonal unimodular A corresponding to any given L, and conversely.

For a pure rotation, $\theta_b = 0$, $\mathscr{E} = 0$, $\sin\theta_r = |\mathscr{H}|$ so that we have $\hat{\mathbf{a}} = \hat{\mathscr{H}}$ and we recover the formula (3.20). For a pure boost as given by Eq. (4.92), $\theta_r = 0$, $\mathscr{H} = 0$, $\sinh\theta_b = |\mathscr{E}|$ and we have

$$\hat{\mathbf{a}} = \hat{\mathscr{E}} \equiv (v_x, v_y, v_z)v^{-1},$$

giving

$$D^{10}(L) = \begin{bmatrix} 1 + (\gamma - 1)(1 - v_x^2/v^2) & -(v_x v_y/v^2)(\gamma - 1) + i\gamma v_z/c & -(v_x v_z/v^2)(\gamma - 1) - i\gamma v_y/c \\ -(v_y v_x/v^2)(\gamma - 1) - i\gamma v_z/c & 1 + (\gamma - 1)(1 - v_y^2/v^2) & -(v_y v_z/v^2)(\gamma - 1) + i\gamma v_x/c \\ -(v_z v_x/v^2)(\gamma - 1) + i\gamma v_y/c & -(v_z v_y/v^2)(\gamma - 1) - i\gamma v_x/c & 1 + (\gamma - 1)(1 - v_z^2/v^2) \end{bmatrix}. \tag{5.7}$$

For a velocity along the x axis $v_x = v$, $v_y = v_z = 0$ and Eq. (5.7) reduces to the transformation as given by Landau and Lifshitz.[10]

If L is exceptional, we have $\mathscr{H} = \mathscr{E} = 0$; $\theta_b = 0$, $\theta_r = \pi$ and as expected Eq. (5.6) fails to determine $\hat{\mathbf{a}}$. The vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and hence $\hat{\mathbf{a}}$ must then be determined from Eqs. (4.82)–(4.85) and we have $\pi\mathbf{a} = \mathbf{f}$ and obtain with the corresponding S_3

$$D^{10}(L) = E_3 + 2S_3^2, \tag{5.8}$$

which follows from Eq. (5.5) with $\theta = \pi$. We observe that it has the same structure as Eq. (3.23) but with complex $\hat{\mathbf{a}}$.

When L is singular, we have $\theta_r = \theta_b = 0$; $|\mathscr{H}| = |\mathscr{E}|$ and $\mathbf{a} = \mathscr{H} + i\mathscr{E}$ is now a null vector. With $\theta = |\mathscr{H}| = |\mathscr{E}|$, we obtain with the appropriate S_3,

$$\begin{aligned} D^{10}(L) &= E_3 + S_3\theta + \tfrac{1}{2}S_3^2\theta^2 \\ &= \begin{bmatrix} 1 + \frac{1}{2}a_1^2\theta^2 & \frac{1}{2}a_1a_2\theta^2 + a_3\theta & \frac{1}{2}a_1a_3\theta^2 - a_2\theta \\ \frac{1}{2}a_2a_1\theta^2 - a_3\theta & 1 + \frac{1}{2}a_2^2\theta^2 & \frac{1}{2}a_2a_3\theta^2 + a_1\theta \\ \frac{1}{2}a_3a_1\theta^2 + a_2\theta & \frac{1}{2}a_3a_2\theta^2 - a_1\theta & 1 + \frac{1}{2}a_3^2\theta^2 \end{bmatrix}. \end{aligned} \tag{5.9}$$

This completes the explicit construction of the matrices of the D^{10} representation of the proper Lorentz group SO(3,1). The complex conjugate D^{01} representation is evidently realized by taking $\theta = \theta_r - i\theta_b$ and $\hat{\mathbf{a}} = \boldsymbol{\alpha} - i\boldsymbol{\beta}$. It is now easy to introduce a classification of proper Lorentz transformations based on the D^{10} or D^{01} representations. We have seen already that any complex orthogonal unimodular A belonging to SO(3,C) corresponds exactly to one proper Lorentz transformation L. Thus the trace χ_A of A would be the character of L in the $D^{10}(D^{01})$ representation of SO(3,1) and from the structure of the matrices A as given by the formulas (3.20) and (3.23) with complex elements and the formulas (5.7) and (5.9), we obtain immediately that *the Lorentz transformation L that corresponds to A is* (i) *a general nonplanar transformation if χ_A is complex or real and < -1,* (ii) *is planar if χ_A is real and $\geqslant -1$, and is rotationlike, singular, or boostlike according as $\chi_A \lesseqgtr 3$, and is exceptional if $\chi_A = -1$.*

The classification according to the $D^{\frac{1}{2}0}$ or $D^{0\frac{1}{2}}$ representations is also straightforward. We know that the two-dimensional complex unimodular group SL(2,C) which is the same as the $D^{\frac{1}{2}0}$ or $D^{0\frac{1}{2}}$ representation provides a double-valued representation of SO(3,1) and we have by the Clebsch–Gordon theorem,

$$D^{\frac{1}{2}0} \times D^{\frac{1}{2}0} = D^{10} \dotplus D^{00}. \tag{5.10}$$

If, therefore, χ_C is the trace of any complex unimodular matrix C belonging to SL(2,C), we have

$$\chi_C^2 = \chi_A + 1, \qquad (5.11)$$

and since there is exactly one L corresponding to $\pm C$, we arrive at the classification: *The Lorentz transformation L that corresponds to a given C is* (i) *a general nonplanar transformation if χ_C is complex*; (ii) *is planar if χ_C is real and is rotationlike, singular, or boostlike according as $|\chi_C| \lesseqgtr 2$, and is exceptional if $\chi_C = 0$.* We collect the three classification schemes in Table I.

Note added in proof: It is shown by Synge [J. L. Synge, Comm. Dublin Inst. for Adv. Stud. Ser. A **21**, 22 (1972)] that with each proper Lorentz transformation, one can associate a pair of *complex unit quaternions* $\pm q \equiv \pm (a_0 + a_\mu e_\mu)$, where the $e_\mu (\mu = 1,2,3)$ satisfy $e_\nu^2 = -1, e_\mu e_\nu = -e_\nu e_\mu = e_\rho (\mu,\nu,\rho = 1,2,3$ cyclic). Since the quaternion units e_μ have an irreducible representation $e_\mu \to -i\sigma_\mu$ in terms of the Pauli matrices σ_μ, we obtain $q \to a_0 E - i a_\mu \sigma_\mu = C$, an element of SL(2,$C$), and we recover the $D^{\frac{1}{2}0}$ representation of SO (3,1). Thus $\chi_C = 2a_0$ and we arrive at the classification: *A proper Lorentz transformation L which corresponds to a given q is* (i) *a general nonplanar one if a_0 is complex* (ii) *is planar if a_0 is real and is boostlike, singular, or rotationlike according as $|a_0| \gtreqless 1$ and is exceptional if $a_0 = 0$.* Table I would therefore be augmented by the following row, with the same column headings

Quaternion	a_0 complex	$a_0 > 1$ or $a_0 < -1$	$a_0 = \pm 1$	$-1 < a_0 < 1$ $a_0 = 0$ exceptional

It is of interest to observe that the four classification schemes neatly taper off according to the characterizing integers 4,3,2,1, which are the dimensions of the corresponding representations on formally regarding the quaternion representation as one-dimensional.

[1]J. L. Synge, *Relativity: The Special Theory* (North-Holland, Amsterdam, 1972), 2nd edition, pp. 60–1, 86–8, 330–5.
[2]M. Hamermesh, *Group Theory* (Addison–Wesley, 1964), p. 478.
[3]A. Z. Petrov, *Einstein Spaces* (Pergamon, Oxford, 1969), p. 10.
[4]H. Jeffreys and S. B. Jeffreys, *Methods of Mathematical Physics* (Cambridge U. P., 1956), p. 122.
[5]E. P. Wigner, *Group Theory* (Academic, New York, 1962), p. 149.
[6]H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, N. J. 1939), p. 56.
[7]J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967), p. 147.
[8]S. L. Bazanski, J. Math. Phys. **6**, 1201 (1965).
[9]C. W. Misner and J. A. Wheeler, Ann. Phys. (N.Y.) **2**, 525 (1957).
[10]L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1971), 3rd edition, pp. 64–5.
[11]H. Boerner, *Representations of Groups* (North Holland, Amsterdam, 1963), pp. 13–4.
[12]I. M. Gel'fand, *Lectures on Linear Algebra* (Interscience, New York, 1967), p. 124.
[13]J. L. Synge, *Relativity: The Special Theory* (North Holland, Amsterdam, 1958), 1st edition, pp. 86–94.
[14]F. Iyres, Jr., *Theory and Problems of Matrices* (McGraw–Hill, New Delhi, 1973), p. 151.
[15]A. Kurosh, *Higher Algebra* (Mir Publishers, Moscow, 1972), p. 154.
[16]C. Møller, *The Theory of Relativity* (Oxford University, London, 1972), p. 120.

Extension of inverse scattering method to nonlinear evolution equation in nonuniform medium

M. R. Gupta
Centre of Advanced Study in Applied Mathematics, Calcutta University, Calcutta–700009, India
Jayanti Ray
Department of Applied Mathematics, Calcutta University, Calcutta–700009, India

(Received 12 October 1979; accepted for publication 16 April 1980)

By allowing the entire spectrum of certain linear eigenvalue problems to evolve with time a general type of nonlinear evolution equation in nonuniform medium which is exactly integrable by the inverse scattering method has been derived. The derivative nonlinear Schrödinger equation or the nonlinear Schrödinger equation with linear or parabolic density profiles are special cases of this generalized form.

PACS numbers: 03.40.Kf, 02.30.Jr

I. INTRODUCTION

The method of inverse scattering has of late become a standard tool for solving initial value problems of nonlinear partial differential equations associated with the evolution of nonlinear waves. In this connection Ablowitz[1] *et al.* have shown that a general form of the nonlinear evolution equation whose exact solution can be determined by this technique from the inverse problem of the Zakharov–Shabat[2] type eigenvalue equation

$$(\partial_x - M)v = \begin{pmatrix} \partial_x & -q \\ -r & \partial_x \end{pmatrix} v = -i\rho \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} v = -i\rho\sigma_3 v \quad (1)$$

is

$$(\sigma_3 \partial_t + 2\Omega(L_A)) \begin{pmatrix} r \\ q \end{pmatrix} = 0, \quad (2)$$

where

$$L_A = \frac{1}{2i} \begin{pmatrix} \partial_x - 2r \int_{-\infty}^{x} dy\, q & 2r \int_{-\infty}^{x} dy\, r \\ -2q \int_{-\infty}^{x} dy\, q & -\partial_x + 2q \int_{-\infty}^{x} dy\, r \end{pmatrix} \quad (3)$$

and $\Omega(\rho)$ is an entire function of ρ. Recently Kaup and Newell[3] have developed an elegant approach through application of which they have derived the general class of exactly integrable nonlinear evolution equations associated with Eq. (1). It is not imperative in their analysis that the eigenvalue remains time invariant. In fact they considered the case where the bound state eigenvalues are assumed to depend on time in a prescribed manner. However in such general situations the evolution equations are usually nonlocal and further cannot be given explicitly in terms of the potentials r and q alone. The present analysis pertains to the case where the entire spectrum evolves with time in accordance with

$$\frac{d\rho}{dt} = f(\rho;t). \quad (4)$$

$f(\rho,t)$ is an entire function ρ with arbitrary functions of time t occurring as coefficients of the different powers of ρ. Equation (4) is solvable and ρ may therefore be expressed in terms of t and the initial ρ_0. It is shown that the nonlinear evolution equation which can be exactly solved in this case is

$$[\sigma_3 \partial_t - 2if(L_A,t)x + 2\Omega(L_A)] \begin{pmatrix} r \\ q \end{pmatrix} = 0. \quad (5)$$

In obtaining (5) we have followed the AKNS[1] method. Alternatively one could derive the time evolution equation of the scattering data in a manner as shown by Kaup[4] and, by employing the closure properties of the complete set of eigenfunctions of L_A and its adjoint[5], arrive at a more general form of evolution equation. The relationship between the two would be similar to that between the one determined by Kaup and Newell[3] and Eqs. (2) and (3). However our main purpose as exhibited by Eq. (5) is to show that the extension of the time dependence to the entire spectrum renders the initial value problem of nolinear wave propagation in certain types of nonuniform dispersive medium exactly integrable. The evolution equation though explicit in r and q may become nonlocal depending on the choice of $f(\rho;t)$. The nonlinear Schrödinger equation with linear[6] or parabolic density profiles[7] are special cases of the last equations. Further under restrictions similar to those on $\Omega(\rho)$[1,3] it is possible to extend $f(\rho;t)$ to suitable class of rational functions.

With appropriate modification it is possible to obtain the general class of exactly integrable nonlinear evolution equation associated with the Newell–Kaup eigenvalue problem

$$(\partial_x - \rho M)v = -i\rho^2\sigma_3 v,$$

the entire spectrum ρ being assumed to be time dependent. As shown by Newell and Kaup[8] the solution of the derivative nonlinear Schrödinger equation is related to the inverse scattering problem of this eigenvalue equation.

II. EVOLUTION EQUATION CORRESPONDING TO ZAKHAROV–SHABAT PROBLEM

We assume that the time dependence of v is given by

$$v_t = Nv = \begin{pmatrix} A & B \\ C & -A \end{pmatrix} v. \quad (6)$$

The eigenvalue ρ being dependent on time cross differentiation of (1) and (6) leads to

$$M_t - N_x + [M - i\rho\sigma_3, N] - i\rho_t\sigma_3 = 0. \quad (7a)$$

In explicit form

 0022-2488/81/102180-04$01.00

$$A_x = qC - rB - i\rho_t\,,$$
$$B_x + 2i\rho B = q_t - 2Aq\,,$$
$$C_x - 2i\rho C = r_t + 2Ar\,. \tag{7b}$$

The first equation of the set (7b) shows that for q and $r \to 0$ as $|x| \to \infty$ the asymptotic behavior of A is given by

$$A(x,t;\rho(t)) \underset{x\to-\infty}{\longrightarrow} A_0^{(-)} - i\rho_t x$$
$$\underset{x\to+\infty}{\longrightarrow} A_0^{(+)} - i\rho_t x\,. \tag{8}$$

We now proceed to a formal solution of (7) and the time evolution equation of the scattering data for the inverse problem of (1). A sufficient condition for the existence of the solutions subject to (7) could yield the general form for the nonlinear evolution equation for the potential q and r. This in essence is the procedure adopted by Ablowitz *et al.*[1]

Let $\phi,\bar\phi$ and $\psi,\bar\psi$ denote the pairs of linearly independent solutions of (1) with boundary conditions

$$\phi \underset{x\to-\infty}{\longrightarrow} \begin{pmatrix}1\\0\end{pmatrix} e^{-i\rho x};\quad \bar\phi \underset{x\to-\infty}{\longrightarrow} \begin{pmatrix}0\\-1\end{pmatrix} e^{i\rho x},$$
$$\psi \underset{x\to+\infty}{\longrightarrow} \begin{pmatrix}0\\1\end{pmatrix} e^{i\rho x};\quad \bar\psi \underset{x\to+\infty}{\longrightarrow} \begin{pmatrix}1\\0\end{pmatrix} e^{-i\rho x}. \tag{9}$$

The time development of the scattering data defined through

$$\phi = a\bar\psi + b\psi;\quad \bar\phi = \bar b\bar\psi - \bar a\psi;\quad a\bar a + b\bar b = 1 \tag{10}$$

are given by

$$a_t = (A_0^{(+)} - A_0^{(-)})a + B^{(+)}b;$$
$$b_t = C^{(+)}a - (A_0^{(+)} + A_0^{(-)})b\,,$$
$$\bar a_t = -(A_0^{(+)} - A_0^{(-)})\bar a - C^{(+)}\bar b; \tag{11}$$
$$\bar b_t = -B^{(+)}\bar a + (A_0^{(+)} + A_0^{(-)})\bar b\,,$$
$$B^{(+)} = \lim_{x\to+\infty} Be^{2i\rho x};\quad C^{(+)} = \lim_{x\to+\infty} Ce^{-2i\rho x}.$$

These follow easily on noting that the extra contributions proportional to $\pm i\rho_t x$ which arise from the time derivative of the asymptotic expressions for ϕ and $\bar\phi$ are exactly compensated by virtue of (8).

Writing $\Phi = \begin{pmatrix}\phi_1 & \bar\phi_1\\ \phi_2 & \bar\phi_2\end{pmatrix}$ a formal solution for (7a) and (7b) is now expressed in the form

$$(\Phi^{-1}N\Phi)_{x=x'} - (\Phi^{-1}N\Phi)_{x=x''}$$
$$= -i\rho_t[x\Phi^{-1}\sigma_3\Phi]_{x''}^{x'} + \int_{x''}^{x'} dx\,\Phi^{-1}M_t\Phi$$
$$+ i\rho_t\int_{x''}^{x'} dx\,x\partial_x(\Phi^{-1}\sigma_3\Phi)\,. \tag{12}$$

As $x'\to+\infty$ and $x''\to-\infty$ the integrated part on the right-hand side of (12) exactly cancels the corresponding unbounded terms on the left-hand side arising from $A^{(\pm)} = \lim_{x\to\pm\infty} A(x,t;\rho)$. The last integral in (12) which completely takes into account the effect of the time dependence of ρ may be written explicitly in the form ($x'\to+\infty$, $x''\to-\infty$),

$$-2i\rho_t\int_{-\infty}^{+\infty} dx\,x\begin{pmatrix} q\phi_2\bar\phi_2 + r\phi_1\bar\phi_1 & q\bar\phi_2^2 + r\bar\phi_1^2\\ -q\phi_2^2 - r\phi_1^2 & -q\phi_2\bar\phi_2 - r\phi_1\bar\phi_1\end{pmatrix}. \tag{13}$$

In deriving (13) we have used the eigenvalue equation (1) for ϕ and $\bar\phi$; it is also assumed that the behavior of q and r at $|x|\to\infty$ are such as to guarantee the convergence of the integrals.

Substituting (13) into (12) with $x'\to+\infty$ and $x''\to-\infty$ we arrive at the expressions for $A_0^{(+)}, B^{(+)}, C^{(+)}$ in terms of $A_0^{(-)}$ and the integrals of the form

$$I(\phi,\bar\phi) = \int_{-\infty}^{+\infty} dx\,(-q_t\phi_2\bar\phi_2 + r_t\phi_1\bar\phi_1)\,,$$
$$J(\phi,\bar\phi) = \int_{-\infty}^{+\infty} dx\,x(-q\phi_2\bar\phi_2 + r\phi_1\bar\phi_1)\,, \tag{14}$$

and finally with the help of (10) and (11) obtain

$$(\bar b\,|a)_t = \frac{K(\psi,\psi)}{a\bar b}\cdot(\bar b\,|a)\,, \tag{15}$$
$$(b\,|\bar a)_t = \frac{K(\bar\psi,\bar\psi)}{\bar a b}\cdot(b\,|\bar a)\,, \tag{16}$$
$$K(\psi,\psi) = I(\psi,\psi) - 2i\rho_t J(\psi,\psi)\,. \tag{17}$$

$K(\psi,\psi)$ is the extension of $I(\psi,\psi)$[1] to the case $\rho_t \neq 0$. As in the analogous case of time independent ρ we take

$$K(\psi,\psi) = 2\Omega(\rho)a\bar b = -2\Omega(\rho)\int_{-\infty}^{+\infty} dx\,(q\psi_2^2 + r\psi_1^2)\,,$$
$$K(\bar\psi,\bar\psi) = -2\Omega(\rho)\,\bar a b = 2\Omega(\rho)\int_{-\infty}^{+\infty} dx\,(q\bar\psi_2^2 + r\bar\psi_1^2)\,, \tag{18}$$

where $\Omega(\rho)$ is an arbitrary entire function of ρ. The time development of the scattering data are thus determined by the set of linearized equations (15) and (16). We now assume that $\rho_t = f(\rho,t)$ is also an entire function of ρ, the coefficients of the different powers of ρ being functions of t. Since $\Psi = \binom{\psi_1^2}{\psi_2^2}$ satisfies the equation $L\Psi = \rho\Psi$ where the operator L is identical with that in the analogous case for $\rho_t = 0$, we have

$$\Omega(\rho)\Psi = \Omega(L)\Psi;\quad f(\rho,t)\Psi = f(L,t)\Psi\,. \tag{19}$$

A sufficient criterion for the validity of (17) then leads to the nonlinear evolution equation for the potentials q and r,

$$[\sigma_3\partial_t - 2if(L_A,t)x + 2\Omega(L_A)]\begin{pmatrix}r\\q\end{pmatrix} = 0\,.$$

The operator L_A denotes the adjoint of L and is given by (3). Corresponding to the case where $\Omega(\rho), f(\rho)$ are the ratios of two entire functions we have the following generalizations:

$$[\Omega_2(L_A)f_2(L_A)\sigma_3\partial_t - 2i\Omega_2(L_A)f_1(L_A)x + 2\Omega_1(L_A)f_2(L_A)]$$
$$\times\begin{pmatrix}r\\q\end{pmatrix} = 0\,, \tag{20}$$

where

$$\Omega(\rho) = \Omega_1(\rho)|\Omega_2(\rho);\quad f(\rho) = f_1(\rho)|f_2(\rho)\,.$$

If $A_0^{(+)} = A_0^{(-)}$ and $B^{(+)} = C^{(+)} = 0$, Eqs. (11) have the solutions

$$a(\rho(t),t) = a(\rho_0,0)\,,$$

$$b(\rho(t),t) = b(\rho_0,0)\exp\left(-2\int_0^t \Omega(\rho(t'))\,dt'\right),$$
$$\rho_0 = \rho(t=0). \tag{21}$$

In the Marchenko equations determining q and r one has to make use of the above expressions for the time dependence of the scattering data. These last equations therefore enable us to obtain the solution of the nonlinear evolution equation (5) when the solutions of the corresponding homogeneous equations are known.

In particular with $f(\rho) = \epsilon$ and $\epsilon + \mu\rho^2$ and $\Omega(\rho) = i\rho^2$, Eq. (5) reduces to the nonlinear Schrödinger equation with linear and parabolic density profiles whose solutions have been obtained earlier.[6,7]

III. EVOLUTION EQUATION CORRESPONDING TO NEWELL-KAUP EIGENVALUE PROBLEM

If (6) determines the time evolution of the solution of the eigenvalue problem

$$(\partial_x - \rho M)v = -i\rho^2\sigma_3 v, \tag{22}$$

the Lax condition leads to

$$\rho M_t + \rho_t M - N_x + [\rho M - i\rho^2\sigma_3, N] - i(\rho^2)_t\sigma_3 = 0. \tag{23}$$

The asymptotic behavior appropriate to this case is simply obtained by replacing ρ and ρ_t by ρ^2 and $(\rho^2)_t$ in Eqs. (8) and (9). With these modifications (22) yields in the usual manner the formal solution for N in this case

$$[\Phi^{-1}N\Phi]_{x''}^{x'} = -i(\rho^2)_t[x\Phi^{-1}\sigma_3\Phi]_{x''}^{x'} + \int_{x''}^{x'} dx'\,[\rho\Phi^{-1}M_t\Phi + \rho_t\Phi^{-1}M\Phi + i(\rho^2)_t x\partial_x(\Phi^{-1}\sigma_3\Phi)]. \tag{24}$$

Here as well we find that the unbounded terms in (24) are compensated as $x' \to +\infty$ and $x'' \to -\infty$ leading to

$$(a\bar{a} - b\bar{b})A_0^{(+)} + \bar{a}bB^{(+)} + a\bar{b}C^{(+)} = A_0^{(-)} + \widetilde{K}(\phi,\bar{\phi}),$$
$$2abA_0^{(+)} + b^2B^{(+)} - a^2C^{(+)} = -\widetilde{K}(\phi,\phi),$$
$$2\bar{a}\bar{b}A_0^{(+)} - \bar{a}^2B^{(+)} + \bar{b}^2C^{(+)} = \widetilde{K}(\bar{\phi},\bar{\phi}), \tag{25}$$

where,

$$\widetilde{K}(\phi,\bar{\phi}) = \rho I(\phi,\bar{\phi}) - 4i\rho^2\rho_t J(\phi,\bar{\phi}) + \rho_t H(\phi,\bar{\phi}),$$
$$H(\phi,\bar{\phi}) = \int_{-\infty}^{+\infty} (-q\phi_2\bar{\phi}_2 + r\phi_1\bar{\phi}_1)\,dx. \tag{26}$$

The time development equation for the scattering data reduce to

$$(\bar{b}\,|a)_t = \widetilde{K}(\psi,\psi)|a^2; \quad (b\,|\bar{a})_t = \widetilde{K}(\psi,\psi)|\bar{a}^2. \tag{27}$$

If $q,r \to 0$ as $x \to \infty$, then it can be shown from (22) that $\Psi = \binom{\psi_1^2}{\psi_2^2}$ satisfies

$$\Lambda\Psi = \rho^2\Psi, \tag{28}$$

where

$$\Lambda = \frac{1}{2}\begin{pmatrix} i\partial_x + qr + q\int_x^\infty dy\, r_y & -q^2 - q\int_x^\infty dy\, q_y \\ -r^2 - r\int_x^\infty dy\, r_y & -i\partial_x + qr + r\int_x^\infty dy\, q_y \end{pmatrix} \tag{29}$$

Equation (27) suggests the *ansatz*

$$\widetilde{K}(\psi,\psi) = 2\Omega(\rho^2)a\bar{b} = -2\rho\Omega(\rho^2)\int_{-\infty}^{+\infty}(q\psi_2^2 + r\psi_1^2)\,dx, \tag{30}$$

where $\Omega(\rho^2)$ is an entire function of ρ^2. We next assume that the explicit time dependence of ρ is of the form

$$(\ln\rho^2)_t = f(\rho^2), \tag{31}$$

where $f(\rho^2)$ is also an entire function of ρ^2. Substituting for $\widetilde{K}$ from (26) into (30) we obtain after the usual transformation the nonlinear evolution equation for $\binom{r}{q}$ as a sufficient criterion for the integrability of this scattering data equations:

$$\sigma_3\binom{r}{q}_t + [\tfrac{1}{2}f(\Lambda_A)\sigma_3 - 2i\Lambda_A f(\Lambda_A)x + 2\Omega(L_A)]\binom{r}{q} = 0, \tag{32}$$

where Λ_A is the adjoint operator

$$\Lambda_A = \frac{1}{2}\begin{pmatrix} -i\partial_x + qr + r_x\int_{-\infty}^x dy\, q & -r^2 - r_x\int_{-\infty}^x dy\, r \\ -q^2 - q_x\int_{-\infty}^x dy\, q & i\partial_x + qr + q_x\int_{-\infty}^x dy\, r \end{pmatrix}. \tag{33}$$

With $\Omega(\rho^2) = -2i\rho^4$ and $f(\rho^2) = \alpha$, equation (32) reduces to

$$\begin{pmatrix} r \\ -q \end{pmatrix}_t + \tfrac{1}{2}\alpha\begin{pmatrix} r \\ -q \end{pmatrix} + \alpha\begin{pmatrix} (-xr)_x \\ (xq)_x \end{pmatrix} + \begin{pmatrix} ir_{xx} - (r^2q)_x \\ iq_{xx} + (q^2r)_x \end{pmatrix} = 0, \tag{34}$$

which describes approximately propagation of circularly polarized waves in a magnetoplasma in the presence of inhomogeneities. Clearly its exact solution can be obtained by the inverse scattering method.

ACKNOWLEDGMENT

The authors desire to express thanks to the referee for suggesting revisions. One of us (J. R.) also takes this opportunity to express thanks to the University Grants Commission, India for awarding a Research fellowship.

[1]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[2]V. E. Zakharov and A. B. Shabat, Zh. Eksp. Teor. Fiz. **61**, 1745 (1971); [Sov. Phys. JETP **34**, 908 (1972)].
[3]D. J. Kaup and A. C. Newell, Adv. in Math. **31**, 67 (1979).
[4]D. J. Kaup, SIAM J. Appl. Math. **31**, 121 (1976).
[5]D. J. Kaup, J. Math. Anal. Appl. **54**, 849 (1976).
[6]H. H. Chen and C. S. Liu, Phys. Rev. Lett. **37**, 693 (1976).
[7]M. R. Gupta, Phys. Lett. A **72**, 420 (1979).
[8]A. C. Newell and D. J. Kaup, J. Math. Phys. **19**, 798 (1978).

The behavior at infinity of isotropic vortices and monopoles [a)]

Bradley Plohr
The Rockefeller University, New York, New York 10021

(Received 6 April 1981; accepted for publication 18 May 1981)

We derive detailed asymptotic formulae for the behavior at infinity of isotropic vortex solutions of the abelian Higgs model and monopole solutions of the Yang–Mills Higgs model. In particular we find that the classical mass of the Higgs field is the smaller of m and twice the mass of the gauge field, where $(mc/\hbar)^2$ is the curvature of the Higgs self-interaction potential at the classical vacuum.

PACS numbers: 03.50.Kk, 11.10.Jj, 11.10.Lm

1. INTRODUCTION

In this paper we present an analysis of the asymptotic behavior at large distances of certain isotropic solutions of classical gauge field equations, namely the Nielsen–Olesen vortex solutions of the abelian Higgs model in two-dimensional Euclidean space, and the t' Hooft–Polyakov monopole solutions of the Yang–Mills Higgs model in three-dimensional Euclidean space. The behavior at infinity of these classical solutions field indicates that the Higgs mechanism of symmetry breaking is operative in the quantized versions of these models, and the exact asymptotics determines the classical approximation of the masses of the corresponding quantum particles. We find that the mass of the Higgs particle is $m_\phi = \min\{\hbar/c\cdot[V''(R_\infty)]^{1/2}, 2m_A\}$, where V is the Higgs self-interaction potential, R_∞ is the asymptotic value of the norm of the Higgs field (i.e., the classical vacuum), and m_A is the mass of the gauge field. A physical interpretation of this result is given in the Conclusion.

2. THE ABELIAN HIGGS MODEL

The abelian Higgs model describes a charged scalar Higgs field which is self-coupled via a potential V and which interacts with an abelian [i.e., U(1)] gauge field in two-dimensional Euclidean space. Thus the Higgs field ϕ is a complex-valued function on $\mathbb{R}^2$, and the gauge field may be written as iA, where A is a real-valued 1-form on $\mathbb{R}^2$. [We represent the Lie algebra of U(1) as $i\mathbb{R}$.] The potential V is assumed to be twice continuously differentiable, nonnegative, and symmetric about the origin; furthermore we assume that V has a zero at $R_\infty > 0$ such that $V''(R_\infty) > 0$ and $V(\hat{R}) > 0$ if $\hat{R} < R_\infty$. (See Fig. 1.)

Using units in which $R_\infty = 1$ and $e/\hbar c = 1$ (where the charge of the Higgs field is taken to be $-e < 0$), the Euclidean action for the abelian Higgs model is

$$\mathcal{A}(\phi,A) = \int_{\mathbb{R}^2} d^2x\,\{\tfrac{1}{2}\|F(A)\|^2 + \tfrac{1}{2}\|d_A\phi\|^2 + V(\|\phi\|)\}; \quad (2.1)$$

$F(A) = dA$ is the field strength of A and $d_A\phi = d\phi + iA\phi$ is the covariant derivative of ϕ. The critical points of $\mathcal{A}$ formally satisfy the equations

$$d_A^* d_A\phi + V'(\|\phi\|)\phi/\|\phi\| = 0, \quad (2.2a)$$

$$d^*dA + (1/2i)(\bar{\phi}d_A\phi - \overline{\phi d_A\phi}) = 0 \quad (2.2b)$$

(here * denotes the formal adjoint with respect to the appropriate inner product, and $\bar{}$ denotes complex conjunction). These are the abelian Higgs equations.

We wish to consider vortex solutions of Eqs. (2.2a) and (2.2b) characterized by having finite Euclidean action (2.1) and by exhibiting a classical version of the Higgs symmetry-breaking mechanism, viz.

$$\lim_{|x| \to \infty} \|\phi(x)\| = 1.$$

Following Nielsen and Olesen[1] we look for isotropic solutions of the form

$$\phi(x,y) = R(r)\exp(in\theta) \quad (2.3a)$$

and

$$A(x,y) = S(r)\,d\theta, \quad (2.3b)$$

where n is an integer, and r and θ are polar coordinates defined by $x = r\cos\theta$ and $y = r\sin\theta$. It can be shown[2] that for every integer n there exist Higgs and gauge fields ϕ and A of the form (2.3) which are twice continuously differentiable and satisfy (2.2) throughout $\mathbb{R}^2$, and for which the Euclidean action is finite and $\lim_{|x|\to\infty} \|\phi(x)\| = 1$. (The integer n is known as the vortex number.) In this paper we study the asymptotic behavior of such fields at large distances.

As one may easily verify, the real-valued functions R and S on $\mathbb{R}_+ = \{r\in\mathbb{R} \mid r > 0\}$ satisfy the coupled pair of nonlinear differential equations

$$-R''(r) - r^{-1}R'(r) + r^{-2}(S(r) + n)^2R(r) + V'(R(r)) = 0 \quad (2.4a)$$

and

$$-S''(r) + r^{-1}S'(r) + R^2(r)(S(r) + n) = 0. \quad (2.4b)$$

Furthermore,

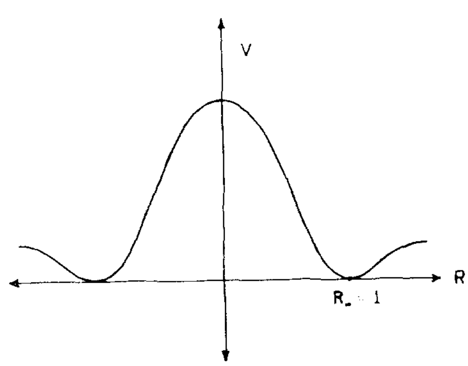


FIG. 1. A typical Higgs field self-interaction potential.

[a)] Supported in part by NSF-PHY-80-01979. This work represents part of the author's Ph.D. dissertation presented to the Department of Physics, Princeton University in August, 1980.

 0022-2488/81/102184-07$01.00

$$F(R,S) = \int_0^\infty r\,dr\,\{\tfrac{1}{2}r^{-2}[S'(r)]^2 + \tfrac{1}{2}[R'(r)]^2 + \tfrac{1}{2}r^{-2}R^2(r)(S(r)+n)^2 + V(R(r))\} \quad (2.5)$$

is finite since $F(R,S)$ coincides with $(2\pi)^{-1}\mathscr{A}(\phi,A)$; and we may assume that

$$\lim_{r\to\infty} R(r) = 1 \quad (2.6)$$

because $\lim_{|x|\to\infty}\|\phi(x)\| = 1$ and we can use a (global) gauge transformation to ensure that $R(r)$ is positive for large r. In Sec. 3 we will prove

Theorem 1: Suppose that R and S satisfy (2.4)–(2.6) let $m = [V''(1)]^{1/2}$. Then there exist constants α and β such that

$$S(r) = -n + \beta r K_1(r)\cdot(1 + o(\exp[-\min\{m,2\}r]))$$

and

$$S'(r) = -(S(r)+n)\cdot(1 - (2r)^{-1} + O(r^{-2}))$$

as $r\to\infty$, and such that:

(a) if $m<2$,

$$R(r) = 1 - \alpha K_0(mr)\cdot(1 + o(\exp(-mr)) - [m^2-4]^{-1}\beta^2[K_1(r)]^2(1+o(1))$$

and

$$R'(r) = -m(R(r)-1)\cdot(1 + (2mr)^{-1} + O(r^{-2}));$$

(b) if $m=2$,

$$R(r) = 1 - \tfrac{1}{2}\beta^2 r[K_1(r)]^2\cdot(1+o(1))$$

and

$$R'(r) = -2(R(r)-1)\cdot(1+o(1));$$

(c) if $m>2$,

$$R(r) = 1 - [m^2-4]^{-1}\beta^2[K_1(r)]^2\cdot(1+o(1))$$

and

$$R'(r) = -2(R(r)-1)\cdot(1+o(1))$$

as $r\to\infty$.

From this result follows the decay properties of gauge-invariant quantities of physical importance, such as $\|\phi\| = |R|$, $\bar\phi d_A\phi = R\cdot R'\,dr + iR^2\cdot(n+S)\,d\theta$, and $dA = r^{-1}S'\,dr\wedge r d\theta$. In particular, the mass of the gauge field is $m_A = \hbar/c\cdot e/\hbar c\cdot R_\infty$ and the mass of the Higgs boson is $m_\phi = \min\{\hbar/c\cdot[V''(R_\infty)]^{1/2}, 2m_A\}$.

3. PROOF OF THEOREM 1

In order to determine the asymptotics of our fields we will find it convenient to work with the shifted fields u and v defined by

$$u(r) = r^{1/2}(1 - R(r)) \quad (3.1a)$$

and

$$v(r) = r^{-1/2}(n + S(r)) \quad (3.1b)$$

instead of with R and S. Given that R and S satisfy Eqs. (2.4a) and (2.4b), u and v satisfy the differential equations

$$-u''(r) + [V''(1) - r^{-2}/4 - \hat V(r^{-1/2}u(r))r^{-1/2}u(r) + r^{-1}v^2(r)]u(r) = r^{-1/2}v^2(r) \quad (3.2a)$$

and

$$-v''(r) + [1 + \tfrac{3}{4}r^{-2} - 2r^{-1/2}u(r) + r^{-1}u^2(r)]v(r) = 0, \quad (3.2b)$$

where we have written $V'(1-\tilde R) = -V''(1)\tilde R + \hat V(\tilde R)\tilde R^2$. Note that $\hat V(\tilde R) = O(1)$ as $\tilde R\to 0$ by Taylor's theorem.

In the proof of Theorem 1 we will show that $u(r)$ and $v(r)$ vanish as $r\to\infty$. We will then be in a position to apply the following lemma, which shows that solutions of differential equations of the form of Eqs. (3.2a) and (3.2b) decay exponentially at infinity, assuming that they vanish at infinity. The proof of the lemma is based on an elementary form of the maximum principle.[3]

Lemma 3.1: Suppose w is a C^2 solution on $\mathbb{R}_+$ of the differential equation

$$-w''(r) + (\kappa^2 + f(r))w(r) = g(r),$$

where f and g are continuous functions on $\mathbb{R}_+$ *and* $\kappa>0$. If w and f vanish at infinity, and

$$g(r) = O(\exp(-\lambda r))$$

as $r\to\infty$ for some $\lambda>0$, then for every positive $\epsilon<1$

$$w(r) = O(\exp[-\min\{\kappa(1-\epsilon)^{1/2},\lambda\}r])$$

as $r\to\infty$.

Proof: Fix $0<\epsilon<1$. By hypothesis there is an $r_\epsilon\in\mathbb{R}_+$ and a constant γ_ϵ such that $\tfrac{1}{2}\epsilon\kappa^2 + f(r)\geqslant 0$ for all $r\geqslant r_\epsilon$ and such that $|g(r)|\leqslant\gamma_\epsilon\exp(-\lambda r)$ for all $r\geqslant r_\epsilon$. Define $\kappa_\epsilon = \kappa(1-\tfrac{1}{2}\epsilon)^{1/2}$, $f_\epsilon = \tfrac{1}{2}\epsilon\kappa^2 + f$, and $\mu_\epsilon = \min\{\kappa(1-\epsilon)^{1/2},\lambda\}$. Thus

$$-w''(r) + (\kappa_\epsilon^2 + f_\epsilon(r))w(r) = g(r),$$

with $f_\epsilon(r)\geqslant 0$ for all $r\geqslant r_\epsilon$ and $g(r)\leqslant\gamma_\epsilon\exp(-\mu_\epsilon r)$ for all $r\geqslant r_\epsilon$.

Define the function $\tilde w$ by

$$\tilde w(r) = \rho_\epsilon\exp(-\kappa_\epsilon r) + [\kappa_\epsilon^2 - \mu_\epsilon^2]^{-1}\gamma_\epsilon\exp(-\mu_\epsilon r) - w(r),$$

where $\rho_\epsilon\geqslant 0$ is chosen so that $\tilde w(r_\epsilon)\geqslant 0$. Then one easily verifies that

$$-\tilde w''(r) + [\kappa_\epsilon^2 + f_\epsilon(r)]\tilde w(r) = f_\epsilon(r)[\tilde w(r) + w(r)] + \gamma_\epsilon\exp(-\mu_\epsilon r) - g(r)\geqslant 0$$

for all $r\geqslant r_\epsilon$.

We will show that $\tilde w(r)\geqslant 0$ if $r\geqslant r_\epsilon$. Suppose to the contrary that $\tilde w(r_1)<0$ for some $r_1>r_\epsilon$; then because $\tilde w(r_\epsilon)\geqslant 0$ and $\tilde w(r)\to 0$ as $r\to\infty$ there exists an $r_{\min}>r_\epsilon$ at which $\tilde w$ attains a strictly negative minimum. But then

$$\tilde w''(r_{\min})\leqslant[\kappa^2 + f_\epsilon(r_{\min})]\tilde w(r_{\min})<0,$$

contradicting the fact that $\tilde w$ attains a minimum at $r_{\min}$; thus $\tilde w(r)\geqslant 0$ for all $r\geqslant r_\epsilon$. This inequality implies the bound

$$\tilde w(r)\leqslant\text{const}\ \exp(-\mu_\epsilon r)$$

for all $r\geqslant r_\epsilon$, since $\kappa_\epsilon\geqslant\mu_\epsilon$. Because we may just as well apply this argument to $-w$, we deduce that $w(r) = O(\exp(-\mu_\epsilon r))$ as $r\to\infty$.

By appealing to this lemma we will see that u and v satisfy equations which are modified Bessel equations, except for perturbations which are exponentially small at infinity. Therefore we will compare u and v to the modified Bessel functions that approximately solve Eqs. (3.2a) and (3.2b) to obtain more precise results about their asymptotic behavior. Let K_ν denote the usual[4] modified Bessel function of order ν that is subdominant at infinity, and define k_ν on $\mathbb{R}_+$ by

$$k_\nu(r) = (2r/\pi)^{1/2} K_\nu(r);$$

$w = k_\nu$ solves the differential equation

$$-w''(r) + (1 + [\nu^2 - \tfrac{1}{4}] r^{-2}) w(r) = 0.$$

The following properties[4] of k_ν are used below: if ν^2 is real, then k_ν is a real-valued function such that

$$k_\nu(r) = \exp(-r)\cdot(1 + O(r^{-1}))$$

and

$$k'_\nu(r) = -k_\nu(r)\cdot(1 + O(r^{-2}))$$

as $r \to \infty$.

Lemma 3.2 is a typical application of the variation-of-parameters technique from the theory of ordinary differential equations. It is a slightly refined form of the WKB approximation.

Lemma 3.2: Consider the differential equation

$$-w''(r) + (\kappa^2 + [\nu^2 - \tfrac{1}{4}] r^{-2} + h(r)) w(r) = 0,$$

where h is a continuous function on $\mathbb{R}_+$, ν^2 is real, and $\kappa > 0$. If $|h|$ is integrable at infinity then there exist C^2 solutions w^+ and w^- of this equation such that

$$w^\pm = \exp(\pm \kappa r)\cdot(1 + o(1))$$

and

$$[w^\pm]'(r) = \pm \kappa w^\pm(r)\cdot(1 + o(1))$$

as $r \to \infty$. If in addition $h(r) = O(\exp(-\mu r))$ as $r \to \infty$, where $\mu > 0$, then

$$w^-(r) = k_\nu(\kappa r)\cdot(1 + O(\exp(-\mu r))$$

and

$$[w^-]'(r) = -\kappa w^-(r)\cdot(1 + O(r^{-2})).$$

Proof: For notational convenience we will assume that $\kappa = 1$; the general case can be reduced to this case by rescaling the independent variable. We first seek a solution w^- of our equation which is of the form $w^- k_\nu \cdot z$, where $z(r) \to 1$ as $r \to \infty$. Motivated by standard tricks used to solve ordinary differential equations we proceed as follows. Suppose we can find a C^2 solution z of the Volterra integral equation

$$z(r) = 1 + \int_r^\infty dr'\, K(r,r') h(r') z(r'),$$

where

$$K(r,r') = [k_\nu(r')]^2 \int_r^{r'} dr''\, [k_\nu(r'')]^{-2}.$$

Then z satisfies the differential equation

$$-z''(r) - 2k'_\nu(r)[k_\nu(r)]^{-1} z'(r) + h(r) z(r) = 0,$$

so that the product $w^- = k_\nu \cdot z$ satisfies our equation. It is thus of interest to find z; we will construct a solution of the Volterra integral equation in the usual manner, namely by proving that its Neumann series converges.

Throughout the following fix $r_0 \in \mathbb{R}_+$ such that $k_\nu(r) > 0$ if $r \geqslant r_0$. It is easy to check that there is a constant ρ such that $0 \leqslant K(r,r') < \tfrac{1}{2}\rho$ and $0 \leqslant -(\partial K/\partial r)(r,r') \leqslant \rho$ for all $r' \geqslant r \geqslant r_0$. For later convenience let $H(r) = \tfrac{1}{2}\rho \int_r^\infty dr' |h(r')|$. Note that $H(r) = o(1)$ as $r \to \infty$ since $|h|$ is integrable at infinity; furthermore, $H(r) = o(\exp(-\mu r))$ as $r \to \infty$ in case $h(r) = o(\exp(-\mu r))$ as $r \to \infty$ by 1' Hôpital's rule.

Define $z_0 = 1$, and for nonnegative integers j define z_{j+1} inductively by

$$z_{j+1}(r) = \int_r^\infty dr' \quad K(r,r') h(r') z_j(r').$$

Then we find that $|z_j| \leqslant (j!)^{-1} H^j$ and $\tfrac{1}{2}|z'_j| \leqslant (j!)^{-1} H^j$ by a simple induction argument applied to the formulae for z_j and z'_j. From the first estimate we conclude that the series $\Sigma_{j=0}^\infty z_j$ converges uniformly on $]r_0, \infty[$ to a function z such that

$$|z(r) - 1| \leqslant \exp(H(r)) - 1 \leqslant \text{const}\, H(r)$$

for all $r > r_0$; the second estimate, along with the formula for z''_j, shows that z is C^2 on $]r_0, \infty[$ and that

$$\tfrac{1}{2}|z'(r)| \leqslant \exp(H(r)) - 1 \leqslant \text{const}\, H(r)$$

for all $r > r_0$. In addition, z is a solution of the Volterra integral equation, as follows by the Lebesgue dominated convergence theorem.

As a consequence $w^- = k_\nu \cdot z$ is a solution of our original equation with the properties that

$$\begin{aligned} w^-(r) &= k_\nu(r) z(r) = k_\nu(r)\cdot(1 + O(H(r))) \\ &= \exp(-r)\cdot(1 + o(1)) \end{aligned}$$

and

$$\begin{aligned} [w^-]'(r) &= k'_\nu(r) z(r) + k_\nu(r) z'(r) \\ &= -w^-(r)\cdot[1 + O(r^{-2}) + O(H(r))] \end{aligned}$$

as $r \to \infty$. We may construct a second solution w^+ by setting

$$w^+(r) = 2w^-(r) \int_{r_1}^r dr'\, [w^-(r')]^{-2}$$

for $r > r_1$ [where we choose $r_1 > r_0$ large enough that $w^-(r) \neq 0$ for all $r \geqslant r_1$]. That w^+, so defined, solves the same equation as does w^- is easily checked directly; the motivation for defining w^+ in this way comes from requiring that the Wronskian of w^+ and w^- be a constant, which is taken to be -2. The asymptotic behavior of w^+ may be determined as follows: by l'Hôpital's rule

$$\begin{aligned} \lim_{r\to\infty} w^+(r) w^-(r) &= \lim_{r\to\infty} \frac{2\int_{r_1}^r dr'\, [w^-(r')]^{-2}}{[w^-(r)]^{-2}} \\ &= \lim_{r\to\infty} \frac{2[w^-(r)]^{-2}}{-2[w^-]'(r)[w^-(r)]^{-3}} = 1, \end{aligned}$$

so that $w^+(r) = \exp(r)\cdot(1 + o(1))$ as $r \to \infty$; and

$$\begin{aligned} &\lim_{r\to\infty} [w^+]'(r)[w^+(r)]^{-1} \\ &= \lim_{r\to\infty} [w^-]'(r)[w^-(r)]^{-1} + 2[w^-(r) w^+(r)]^{-1} = 1, \end{aligned}$$

so that $[w^+]'(r) = w^+(r)\cdot(1 + o(1))$ as $r \to \infty$. This completes the proof.

We now are in a position to prove Theorem 1. Define u and v through Eqs. (3.1a) and (3.1b). First of all, there is a strictly positive constant such that

$$V(R(r)) \geqslant \text{const}\, (1 - R(r))^2$$

for sufficiently large r; this is because $V(\hat{R}) = V''(1) \times (1 - \hat{R})^2 + o(1 - \hat{R})^2$ as $\hat{R} \to 1$, where $V''(1) > 0$. Since $F(R,S)$ is finite we conclude that u is square-integrable at infinity. Since

$$u'(r) = -r^{1/2} R'(r) + \tfrac{1}{2} r^{-1} u(r)$$

the finiteness of $F(R,S)$ shows further that u' is square-integrable at infinity. Hence $(u^2)' = 2u\cdot u'$ is integrable at infinity, so that $\lim_{r\to\infty} u(r)$ exists; but this limit must vanish in order that u^2 be integrable at infinity. Since the third term

$$\int_0^\infty r\,dr\,\tfrac{1}{2}r^{-2}R^2(r)(S(r)+n)^2 = \int_o^\infty dr\,\tfrac{1}{2}(1-r^{-1/2}u(r))^2v^2(r)$$

in $F(R,S)$ is finite it follows that v is square-integrable at infinity. Since

$$v'(r) = r^{-1/2}S'(r) - \tfrac{1}{2}r^{-1}v(r),$$

the finiteness of $F(R,S)$ guarantees that v' is also square-integrable at infinity. As before we conclude that v vanishes at infinity.

An examination of Eqs. (3.2a) and (3.2b) shows that we may apply Lemma 3.1 to v: given any positive $\epsilon < 1$,

$$v(r) = O(\exp[-(1-\epsilon)^{1/2}r])$$

as $r\to\infty$. Using this bound we may apply Lemma 3.1 to u

$$u(r) = O(\exp[-\min\{m,2\}(1-\epsilon)^{1/2}r])$$

as $r\to\infty$.

Again referring to Eqs. (3.2a) and (3.2b) we see that Lemma 3.2 applies to the equation satisfied by v. Since v vanishes at infinity it must be proportional to the subdominant solution constructed in Lemma 3.2, so there exists a constant β' such that

$$v(r) = \beta' k_1(r)\cdot(1 + o(\exp[-\min\{m,2\}(1-\epsilon)^{1/2}r]))$$

and

$$v'(r) = v(r)\cdot(1 + O(r^{-2}))$$

as $r\to\infty$.

Consider now the equation satisfied by u, which we will write as

$$-u''(r) + [m^2 - \tfrac{1}{4}r^{-2} + h(r)]\,u(r) = g(r).$$

Let u_+ and u_- denote the solutions of the corresponding homogeneous equation that are constructed in Lemma 3.2. Using u_+ and u_- we may construct a particular solution u_p of the equation satisfied by u: let

$$u_p(r) = (2m)^{-1}\int_{r_1}^r dr'\,u_-(r)u_+(r')g(r') + (2m)^{-1}\int_r^\infty dr'\,u_+(r)u_-(r')g(r')$$

for $r \geqslant r_1$. Then

$$-u_p''(r) + (m^2 - \tfrac{1}{4}r^{-2} + h(r))u_p(r) = -(2m)^{-1}W(r)g(r),$$

where the Wronskian $W(r) = u_+(r)u'_-(r) - u_-(r)u'_+(r) = -2m(1+o(1)$ by Lemma 3.2; but since $W' = 0$ we must have the $W = -2m$, so that u_p satisfies the correct equation. Because $u - u_p$ satisfies the homogeneous equation there exist constants α_+ and α_- such that $u = \alpha_+u_+ + \alpha_-u_- + u_p$. We will determine the asymptotic behavior of u by examining u_p.

First of all,

$$\lim_{r\to\infty}\frac{(2m)^{-1}\int_\infty^r dr'\,u_-(r')g(r')}{g(r)[u_+(r)]^{-1}} = \lim_{r\to\infty}\frac{-2(m)^{-1}u_-(r)g(r)}{g(r)[u_+(r)]^{-1}\{(\ln g)'(r) - (\ln u_+)'(r)\}} = (2m)^{-1}\Big[m - \lim_{r\to\infty}(\ln g)'(r)\Big]^{-1}$$

by l'Hôpital's rule. But with $g(r) = r^{-1/2}v^2(r)$,

$$(\ln g)'(r) = -\tfrac{1}{2}r^{-1} + 2(\ln v)'(r) = -2(1 + \tfrac{1}{4}r^{-1} + O(r^{-2}))$$

as $r\to\infty$, so that the second term in u_p exhibits the behavior

$$(2m)^{-1}\int_r^\infty dr'\,u_+(r)u_-(r')g(r') = (2m)^{-1}(m+2)^{-1}g(r)\cdot(1+o(1))$$

as $r\to\infty$.

Suppose $m<2$. Then the first term in u_p may be written as

$$(2m)^{-1}\int_{r_1}^\infty dr'\,u_+(r')g(r')\cdot u_-(r) - (2m)^{-1}\int_r^\infty dr'\,u_-(r)u_+(r')g(r')$$

because $u_+(r)g(r)$ vanishes exponentially as $r\to\infty$. But

$$\lim_{r\to\infty}\frac{-(2m)^{-1}\int_r^\infty dr'\,u_+(r')g(r')}{g(r)[u_-(r)]^{-1}} = \lim_{r\to\infty}\frac{(2m)^{-1}u_+(r)g(r)}{g(r)[u_-(r)]^{-1}\{(\ln g)'(r) - (\ln u_-)'(r)\}} = (2m)^{-1}(m-2)^{-1}.$$

Thus there is a constant α'' such that

$$u_p(r) = \alpha''u_-(r) + [m^2-4]^{-1}g(r)\cdot(1+o(1))$$

as $r\to\infty$. Since u vanishes at infinity it follows that there is a constant α' such that

$$u(r) = \alpha' k_0(mr)\cdot(1 + o(\exp[-m(1-\epsilon)^{1/2}r])) + [m^2-4]^{-1}r^{-1/2}[\beta' k_1(r)]^2\cdot(1+o(1))$$

as $r\to\infty$. If on the other hand $m>2$, then because $g\cdot[u_-]^{-1}$ grows (exponentially) at infinity,

$$\lim_{r\to\infty}\frac{(2m)^{-1}\int_{r_1}^r dr'\,u_+(r')g(r')}{g(r)[u_-(r)]^{-1}} = \lim_{r\to\infty}\frac{(2m)^{-1}u_+(r)g(r)}{g(r)[u_-(r)]^{-1}\{(\ln g)'(r) - (\ln u_-)'(r)\}} = (2m)^{-1}(m-2)^{-1},$$

so that

$$u(r) = \alpha_- k_0(mr)\cdot(1 + o(\exp[-2(1-\epsilon)^{1/2}r])) + [m^2-4]^{-1}r^{-1/2}[\beta' k_1(r)]^2\cdot(1+o(1)).$$

Lastly, in case $m=2$, $rg(r)[u_-(r)]^{-1}$ grows (as $r^{1/2}$) *as* $r\to\infty$, so

$$\lim_{r\to\infty}\frac{(2m)^{-1}\int_{r_1}^r dr'\,u_+(r')g(r')}{rg(r)[u_-(r)]^{-1}} = \lim_{r\to\infty}\frac{(2m)^{-1}u_+(r)g(r)}{rg(r)[u_-(r)]^{-1}\{r^{-1} + (\ln g)'(r) - (\ln u_-)'(r)\}}$$

$$= (2m)^{-1}\lim_{r\to\infty}\left[r\{\tfrac{1}{2}r^{-1} + O(r^{-2})\}\right]^{-1} = m^{-1}$$

because $(\ln g)'(r) = -2 - \frac{1}{2}r^{-1} + O(r^{-2})$ as $r\to\infty$; thus

$$u(r) = \alpha_- k_0(mr)\cdot(1 + o(\exp[-m(1-\epsilon)^{1/2}r])) + m^{-1}r^{1/2}[\beta' k_1(r)]^2\cdot(1 + o(1))$$

as $r\to\infty$.

Recycling these formulae through Lemma 3.2 shows that the factors $(1-\epsilon)^{1/2}$ may be eliminated. The same analysis gives the asymptotic behavior of u' and v'; we omit the details. Finally, application of the definitions of u and v in terms of R and S and of k_ν in terms of k_ν finishes the proof.

Let us make some remarks about the constants α and β appearing in Theorem 1. First of all, β must be nonzero because v cannot vanish identically; the sign of β may be determined as follows. Note that

$$\frac{d}{dr}(v(r)v'(r)) = \left[(1 + r^{-1/2}u(r))^2 + \tfrac{3}{4}r^{-2}\right]v^2(r) + [v'(r)]^2$$

so that $v\cdot v'$ is strictly increasing; but $\lim_{r\to\infty} v(r)v'(r) = 0$, so we see that $v(r)v'(r) < 0$ for all r. Since $(\mathrm{sgn}\,n)\cdot v(r)$ is positive for small enough r, [because[2] $S(r)\to 0$ as $r\to 0$] we conclude that $(\mathrm{sgn}\,n)\cdot v(r)$ is positive for all r, whence $(\mathrm{sgn}\,n)\cdot\beta$ is positive. It seems that we cannot argue in this way to show that α is positive, but under the assumption that $|R(r)|\leqslant 1$ for large r it is clear that $\alpha > 0$.

Using a different approach we may derive inequalities which α and β satisfy in certain circumstances. Define the function i_ν by

$$i_\nu(r) = (2\pi r)^{1/2}I_\nu(r),$$

where I_ν is the usual[4] modified Bessel function of order ν that is subdominant at the origin; i_ν satisfies the same differential equation as does k_ν, and $i'_\nu k_\nu - i_\nu k'_\nu = 2$.

If we write the equation satisfied by v in the form

$$-v''(r) + \left[1 + \tfrac{3}{4}r^{-2}\right]v(r) = (1 - R^2(r))\cdot r^{-1/2}(n + S(r)) = f_v(r)$$

we find that $[i'_1 v - i_1 v']' = i_1 f_v$. By the proofs of Lemma 3.2 and Theorem 1,

$$i'_1(r)v(r) - i_1(r)v'(r) = 2\beta' + o(1) = (2\pi)^{1/2}\beta + o(1)$$

as $r\to\infty$. On the other hand, one may deduce[2] from the behavior of i_1 and v as $r\to 0$ that

$$i'_1(r)v(r) - i_1(r)v'(r) = (2\pi)^{1/2}n + o(1)$$

as $r\to 0$. Therefore

$$(2\pi)^{1/2}(\beta - n) = \int_0^\infty dr\, i_1(r)f_v(r).$$

Since $(\mathrm{sgn}\,n)\cdot f_v > 0$ if $|R|\leqslant 1$ (and $n\neq 0$) we find in particular that

$$(\mathrm{sgn}\,n)\cdot\beta > |n|$$

under the assumption that $|R|\leqslant 1$.

In a similar fashion we may write

$$\begin{aligned}-u''(r) + \left[m^2 - \tfrac{1}{4}r^{-2}\right]u(r) &= r^{1/2}[V'(R(r)) \\ &\quad + V''(1)(1 - R(r)) + r^{-2}(S(r) + n)^2 R(r)] \\ &= f_u(r)\end{aligned}$$

and show that[2]

$$(2\pi m)^{1/2}\alpha = 2m\alpha' = [mi'_0(mr)u(r) - i_0(mr)u'(r)]_0^\infty = \int_0^\infty dr\, i_0(mr)f_u(r),$$

so long as $m < 2$. Therefore

$$\alpha > 0$$

in case $R\geqslant 0$ and the potential V satisfies

$$V'(\hat{R}) + V''(1)(1 - \hat{R})\geqslant 0$$

for all $\hat{R}\geqslant 0$. For example, the quartic double-well potential V_m defined by

$$V_m(\hat{R}) = \tfrac{1}{8}m^2(1 - \hat{R}^2)^2$$

satisfies this condition. We note that for the vortex solutions that have been constructed[2] the field R satisfies $0\leqslant R\leqslant 1$.

4. THE YANG–MILLS HIGGS MODEL

The Yang–Mills Higgs model describes a scalar Higgs field which is self-coupled via a potential V and which interacts with a Yang–Mills [i.e., SU(2)] gauge field in three-dimensional Euclidean space. The Higgs field transforms according to some finite dimensional, real, symmetric representation $g\mapsto U(g)$: $W\to W$ of the Lie group SU(2) in the vector space W, and thus ϕ is a W-valued function on $\mathbb{R}^3$; the gauge field is a 1-form on $\mathbb{R}^3$ taking values in the Lie algebra su(2) of SU(2). The potential V is assumed to be twice continuously differentiable, nonnegative, and symmetric about the origin, and to have a zero at $\hat{R}_\infty > 0$ with $V''(R_\infty) > 0$ and $V(\hat{R}) > 0$ for $\hat{R} < R_\infty$. Using units in which $R_\infty = 1$ and $e/\hbar c = 1$ (where e is the coupling constant for the interaction between the Higgs and gauge fields) the Euclidean action for the Yang-Mills Higgs model is

$$\mathscr{A}(\phi,A) = \int_{\mathbb{R}^3} d^3x\, \{\tfrac{1}{2}\|F(A)\|^2 + \tfrac{1}{2}\|d_A\phi\|^2 + V(\|\phi\|)\}, \quad (4.1)$$

where $F(A) = dA + \frac{1}{2}[A,A]$ is the field strength (curvature) of A, and $d_A\phi = d\phi + U(A)\phi$ is the covariant derivative of ϕ. The criticial points of $\mathscr{A}$ formally satisfy the equations

$$d_A^* d_A\phi + V''(\|\phi\|)\phi/\|\phi\| = 0 \quad (4.2a)$$

and

$$d_A^* F(A) + J_A(\phi) = 0; \quad (4.2b)$$

here the Higgs current $J_A(\phi)$ is defined so that

$$(X|J_A(\phi))_{\mathrm{su}(2)} = (U(X)\phi|d_A\phi)_W$$

for all $X\in su(2)$. Monopoles are solutions of Eqs.(4.2a) and (4.2b) which have finite action (4.1) and exhibit symmetry breaking ($\|\phi(x)\|\to 1$ as $|x|\to\infty$).

One may construct[2,5–7] monopoles which are isotropic in the sense that

$$\phi(x) = R(|x|)\sum_m \tau^m Y^l_m\left(\frac{x}{|x|}\right) \quad (4.3a)$$

and

$$A(x) = S(|x|)\sum_{a,j,k}\frac{\sigma^a}{2i}\epsilon_{ajk}\frac{x^j dx^k}{|x|^2} \quad (4.3b)$$

for some integer l, where the σ^a, $a = 1, 2, 3$, are the usual

Pauli spin matrices, and the $\tau^m, m = -l,...,l$, transform according to the l th irreducible representation of SU(2) in such a way that

$$(1/i)L^a\phi + U(\sigma^a/2i)\phi = 0$$

(here $L^a = \sum_{j,k}\epsilon^{ajk}x_j(1/i)\,\partial/\partial x^k$ is the usual angular momentum operator). For example, we may take

$$\phi(x) = R(|x|)\sum_a \frac{\sigma^a}{2i}\frac{x^a}{|x|}$$

if ϕ transforms according to the $l = 1$ adjoint representation, as first shown by t'Hooft[8] and Polyakov.[9]

Assuming a suitable normalization for τ^m, the fields R and S may be shown to satisfy the differential equations

$$-R''(r) - 2r^{-1}R'(r) + l(l+1)r^{-2} \times (S(r)+1)^2R(r) + V'(R(r)) = 0 \qquad (4.4a)$$

and

$$-S''(r) + \tfrac{1}{2}l(l+1)R^2(r)(S(r)+1) + r^{-2}S(r)(S(r)+1)(S(r)+2) = 0, \qquad (4.4b)$$

to have

$$F(R,S) = \int_0^\infty r^2\,dr\,\{\tfrac{1}{2}[R'(r)]^2 + r^{-2}[S'(r)]^2 + \tfrac{1}{2}l(l+1)(S(r)+1)^2R^2(r) + \tfrac{1}{2}r^{-4}S^2(r)(S(r)+2)^2 + V(R(r))\}, \qquad (4.5)$$

finite, and to satisfy

$$\lim_{r\to\infty} R(r) = 1. \qquad (4.6)$$

Theorem 2: *Suppose that* R and S satisfy (4.4)–(4.6); let $m = [V''(1)]^{1/2}, m_l = [\tfrac{1}{2}l(l+1)]^{1/2}$, *and* $v_0 = i\,3^{1/2}/2$. Then there exist constants α and β such that

$$S(r) = -1 + \beta(m_lr)^{1/2}K_{v_0}(m_lr) \times (1 + o(\exp[-\min\{m,2m_l\}r]))$$

and

$$S'(r) = -m_l(S(r)+1)\cdot(1 + O(r^{-2}))$$

as $r\to\infty$, and such that:

(a) if $m < 2m_l$,

$$R(r) = 1 - \alpha(mr)^{-1}e^{-mr}(1 + o(\exp[-mr])) - 2m_l^4[m^2 - 4m_l^2]^{-1}\beta^2(m_lr)^{-1}[K_{v_0}(m_lr)]^2\cdot(1 + o(1))$$

and

$$R'(r) = -m(R(r)-1)\cdot(1 + (mr)^{-1} + O(r^{-2}));$$

(b) if $m = 2m_l$

$$R(r) = 1 - \tfrac{1}{2}m_l^2\beta^2\ln(m_lr)[K_{v_0}(m_lr)]^2\cdot(1 + o(1))$$

and

$$R'(r) = -m_l(R(r)-1)\cdot(1 + o(1));$$

(c) if $m > 2m_l$,

$$R(r) = 1 - 2m_l^4[m^2 - 4m_l^2]^{-1}\beta^2(m_lr)^{-1} \times [K_{v_0}(m_lr)]^2\cdot(1 + o(1))$$

and

$$R'(r) = -m_l(R(r)-1)\cdot(1 + o(1))$$

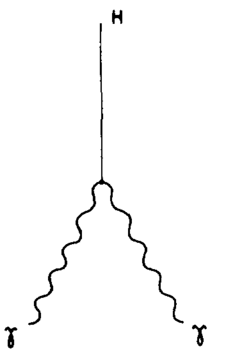


FIG. 2. The Feynman vertex coupling a Higgs particle to two massive photons.

as $r\to\infty$.

The proof[2] of Theorem 2 is exactly analogous to the proof of Theorem 1, so we spare the reader from the details. As far as the constants α and β are concerned, clearly $\beta \neq 0$, and $\beta > 0$ if $S(r) \geqslant -1$ for all large r; equally clearly $\alpha > 0$ if $|R(r)| \leqslant 1$ for all large r, and[2] one may show, in the manner of Sec. 3, that $\alpha > 0$ in case $R \geqslant 0$ and the potential satisfies the inequality (3.3).

The asymptotic behavior of physical fields such as $\|\phi\| = |R|, (\phi\,|d_A\phi)_W = R\cdot R'\,dr$, $\|\phi\|^2\|d_A\phi\|^2 - \|(\phi\,|d_A\phi)_W\|^2 = l(l+1)R^4(S+1)^2$, $\|F(A)\|^2 = 2r^{-2}(S')^2 + r^{-4}S^2(S+2)^2$, and (in case $l = 1$) $(\phi\,|F(A))_{\mathrm{su}(2)} = R\cdot S(S+2)d\theta\wedge\sin\theta\,d\phi$. In particular, the mass of the gauge particle $m_A = \hbar/c\cdot e/\hbar c\cdot R_\infty\cdot[\tfrac{1}{2}l(l+1)]^{1/2}$, while the mass of the Higgs boson is $m_\phi = \min\{\hbar/c\cdot[V''(R_\infty)]^{1/2}, 2m_A\}$.

5. CONCLUSION

Let us comment on the physics behind the result that the mass of the Higgs field cannot exceed twice the mass of the gauge field. In the case of the abelian Higgs model, it is seen from the proofs of Theorem 1 that this arises because the differential equation (3.2a) for the shifted field u has an inhomogeneous term $r^{-1/2}v^2(r)$ whose decay is twice that of the gauge field. In the context of the quantized version of this model, the term in the Euclidean action which gives rise to this inhomogeneous term in the field equations corresponds to the Fenynman vertex shown in Fig. 2 which describes the decay of a Higgs particle into two massive photons. Thus the peculiarity in the Higgs particle mass in the classical field theory reflects the existence of a decay mode $H\to 2\gamma$ in the quantum field theory. A similar interpretation is possible for the Yang-Mills Higgs model.

Finally, we note that Jaffe and Taubes[10] have studied nonisotropic vortices and monopoles with the quartic double-well potential $V_{mc/\hbar}$ and have established that $m_\phi \geqslant \min\{m,2m_A\}$; they conjecture that $m_\phi = \min\{m,2m_A\}$ in this more general setting.

ACKNOWLEDGMENTS

The author thanks Erhard Seiler for suggesting this problem and for his patient guidance during the course of this work, and Mark Ashbaugh for many interesting discussions.

[1]H. Nielsen and P. Olesen, "Vortex Models for Dual Strings,"Nucl. Phys. **B 61**, 361 (1973).
[2]B. Plohr, Princeton University Doctoral Dissertation, October 1980.
[3]M. Protter and H. Weinberg, *Maximum Principles in Differential Equations* (Prentice-Hall, Englewood Cliffs, N. J., 1967).
[4]Bateman Manuscript Project, *Higher Transcendental Functions*, Vol. II (McGraw-Hill, New York, 1953).
[5]Yu. Tyupkin, V. Fateev, and A. Svarts, "Particle-like Solutions of the Equations of Gauge Theories," Theoret. and Math. Phys. **26**, 270 (1977).
[6]J. Rawnsley, Dublin Institute for Advanced Studies Preprint DIAS-TP-77-09, 1977; "Spherically Symmetric Monopoles are Smooth," J. Phys. A **10**, L139 (1977).
[7]R. Weder, "Existence, Regularity, and Exponential Decay of Finite Energy Solutions to Gauge Fields," Communicaciones Técnicas **10**, 198 (1979).
[8]G. t'Hooft, "Magnetic Monopoles in Unified Gauge Theories,"Nucl. Phys. B **79**, 276 (1974).
[9]A. Polyakov, "Particle Spectrum in Quantum Field Theory,"ZhETF Pis. Red. **20**, 430 (1974) [JETP Lett. **20**, 194 (1974)].
[10]A. Jaffe and C. Taubes, *Vortices and Monopoles* (Birkhäuser, Boston, 1980).

Inverse scattering III. Three dimensions, continued

Roger G. Newton
Physics Department, Indiana University, Bloomington, Indiana 47405

(Received 12 February 1981; accepted for publication 10 April 1981)

This paper presents further progress in the solution of the three-dimensional inverse scattering problem for the Schrödinger equation. We prove that if the potential is in a specified class and produces no bound states, then the kernel of the generalized Marchenko equation defines a compact operator and the equation has a unique solution unless the operator has the eigenvalue 1. A partial characterization of scattering amplitudes associated with underlying local potentials without bound states is given and the potential is constructed without assuming its existence. An improved generalization of the Marchenko equation is presented for the case with bound states. The generalized Gel'fand–Levitan equation is critically reviewed.

PACS numbers: 03.65.Nk

I. INTRODUCTION

This paper continues the work of Ref. 1 (which will be referred to as II) on the inverse scattering problem for the Schrödinger equation in three dimensions, in which spherical symmetry of the potential is not assumed. Paper II introduced two methods for the reconstruction of such a potential: one was a generalization of the Marchenko equation, the other, of the Gel'fand–Levitan (GL) equation. While its results guaranteed the uniqueness of the reconstruction of an underlying potential, they did not include a detailed analysis of the nature of the kernels of the integral equations, nor a proof that the generalized Marchenko equation has a unique solution. Also, II contained no results on the construction problem if a scattering amplitude is given and the existence of an underlying potential is not known.

In Sec. 2 of this paper we analyze the kernel of the generalized Marchenko equation introduced in II in the absence of bound states. Under specified conditions on the potential, the kernel is found to define a compact self-adjoint operator $\mathscr{G}_x$ on $L^2(\mathbb{R}_+ \times S^2)$ whose square is in the Hilbert–Schmidt class. The central result of this section is Theorem 2.1 which states that unless $\mathscr{G}_x$ has the eigenvalue 1, the generalized Marchenko equation has a unique solution in $L^2(\mathbb{R}_+\times S^2)$. In fact, if $\mathscr{G}_x$ does not have the eigenvalue -1 either, the equation can be solved by iteration.

Section 3 is addressed to the construction problem if an underlying potential is not known to exist. It is restricted to the case without bound states (which is recognizable by means of the Levinson theorem). Theorem 3.1 provides a partial solution to the characterization problem.[2] On the assumption that the given scattering amplitude belongs to a specified class $\mathscr{A}$, a potential is found to exist and to lead to the originally given amplitude, provided that the solution of the generalized Marchenko equation is miraculous (which is a property that was defined in II). The integrability properties of the constructed potential have not yet been established.

In Sec. 4 we point out that the generalized Marchenko equation of II in the presence of bound states is not solvable by standard methods. (In order to show what the problem is we devote Appendix B to solving the "scalar" Hilbert problem, which can also be solved by quadrature, by a Marchenko-like equation.) We therefore introduce an alternative method that is based on removing the bound states, not by the Fredholm determinant of the Lippmann–Schwinger equation as in II, but by a procedure like that of Sec. 6 of II for the generalized Jost function. The effect is an equation to which the results of Sec. 2 are applicable; consequently the reconstruction with bound states works as well as without them. However, a transfer of the method of Sec. 3 to accomplish a partial characterization when there are bound states was found to lead to technical difficulties that have not yet been surmounted.

Section 5 contains a criticism of the generalized GL equation proposed in II. The crucial triangularity property of one of the kernels used, while probably correct, is found not to be firmly established. If it holds, all the known GL equations, including those for nonzero reference potentials and the nonlinear GL equation are readily generalized, and for central potentials one obtains the known radial GL equations upon expansion in terms of spherical harmonics. The publication of the details of these statements will be deferred until the basis on which they must rest is secure.

Appendices A and C provide detailed proofs of Lemmas 2.1 and 4.1.

2. PROPERTIES OF THE GENERALIZED MARCHENKO EQUATION

We shall assume here that there are no bound states. In the procedure of Sec. 4 of II, and in Theorem 4.1 of II, the function $R(k)$ may then be replaced by 1, and the generalized Marchenko equation may be written in the simplified form[3]

$$\eta_x(\alpha) = G_x(\alpha)\hat{1} + \int_0^\infty d\beta G_x(\alpha+\beta)\eta_x(\beta), \tag{2.1}$$

where

$$G_x(\alpha) = \frac{1}{2}\int_{-\infty}^{\infty} dk\, [\mathscr{S}_x^\dagger(k) - 1]Qe^{ik\alpha}, \tag{2.2}$$

$$\mathscr{S}_x(k;\theta,\theta') = S(k;\theta,\theta')\exp[ikx\cdot(\theta-\theta')], \tag{2.3}$$

or more explicitly

$$G_x(\alpha;\theta,\theta') = -i(2\pi)^{-2}\int_{-\infty}^{\infty} dk\, kA_x^*(k;-\theta',\theta)e^{ik\alpha}, \tag{2.2'}$$

in which

$$A_x(k;\theta,\theta') = A(k;\theta,\theta')\exp[ikx\cdot(\theta-\theta')] \tag{2.4}$$

is the scattering amplitude of a potential shifted by x, and $\hat{1}$ is the vector in $L^2(S^2)$ represented by $f(\theta) = 1$. (In contrast to II we now use a notation in which operators on $L^2(S^2)$ act to the right, and we no longer use subscripts to indicate the k dependence.) Since $\mathscr{S}_x$ depends on x, so do $G_x(\alpha)$ and $\eta_x(\alpha)$. Writing explicitly $\eta_x(\alpha,\theta)$, Theorem 4.1 of II stated that if V satisfies the hypotheses of Lemma 3.2 of II then it has the representation

$$V(x) = -2\theta\cdot\nabla\eta_x(0,\theta), \tag{2.5}$$

where $\eta_x(\alpha,\theta)$ is the only "miraculous" solution of (2.1) [i. e., for which the right-hand side of (2.5) is independent of θ] that satisfies the hypotheses of Lemma 3.2 of II.

If we wish to reconstruct an underlying potential $V(x)$ then it will be important to know if, in fact, Eq. (2.1) has a unique solution. Theorem 4.1 of II leaves open the possibility that (2.1) has more than one solution, but only one of them is *miraculous*. We therefore must investigate the properties of the kernel of the integral equation (2.1). In Appendix A we shall prove the following:

Lemma 2.1: Suppose that $V(x)$ satisfies the hypotheses of Lemma 2.1 of II, which are: $\exists a>0$ and C such that for all $y\in\mathbb{R}^3$,

$$\int d^3x\,|V(x)|^2 + \int d^3x\,|V(x)|\left(\frac{|x|+|y|+a}{|x-y|}\right)^2 < C, \tag{2.5a}$$

and $\exists x_0\in\mathbb{R}^3$ and a monotone function $M_1(t)\in L^1(\mathbb{R}_+)\cap L^2(\mathbb{R}_+)$ such that for all $x\in\mathbb{R}^3$,

$$|V(x+x_0)| \leqslant M_1(|x|); \tag{2.5b}$$

furthermore, suppose that $\exists x_0\in\mathbb{R}^3$, C, and ϵ, with $\frac{3}{4}<\epsilon<1$, and $M_2(t)$ such that for all $x\in\mathbb{R}^3$,

$$|\nabla V(x+x_0)| \leqslant M_2(|x|), \tag{2.5c}$$

and M_2 satisfies the inequalities

$$\int_0^1 dt\;t^{3/2}M_2(t) < \infty, \tag{2.5d}$$

$$F^2(s) \equiv \int_s^\infty dt\;tM_2(t) < Cs^{-2\epsilon}. \tag{2.5e}$$

Then the self-adjoint operator $\mathscr{G}_x$ on $L^2(\mathbb{R}_+\times S^2)$ whose kernel is $\mathscr{G}_x(\alpha,\beta;\theta,\theta') = G_x(\alpha+\beta,\theta,\theta')$, where $G_x(\alpha;\theta,\theta')$ is given by (2.2'), is compact and its square is Hilbert–Schmidt; $\exists C$ such that for all $x\in\mathbb{R}^3$ $\|\mathscr{G}_x^2\|_2 < C$.[4]

We note that if $\exists x_0\in\mathbb{R}^3$ and C, $\epsilon>0$, such that for all $x\in\mathbb{R}^3$,

$$|\nabla V(x+x_0)| < C|x|^{\epsilon-2}(1+|x|)^{-2\epsilon-2}, \tag{2.6}$$

then the hypotheses of this lemma are satisfied. The self-adjointness of $\mathscr{G}_x$ follows from the fact that A satisfies time-reversal invariance,[5]

$$A(-k,\theta,\theta') = A^*(k,\theta,\theta'), \tag{2.7}$$

and reciprocity,

$$A(k;\theta,\theta') = A(k;-\theta',-\theta). \tag{2.8}$$

Let us now write (2.1) in operator form,

$$\eta_x = \xi_x + \mathscr{G}_x\eta_x, \tag{2.1'}$$

where, explicitly,

$$\xi_x(\alpha,\theta) = \int d\theta'\;G_x(\alpha;\theta,\theta').$$

Iterating (2.1') once we have

$$\eta_x = \xi_x + \mathscr{G}_x\xi_x + \mathscr{G}_x^2\eta_x. \tag{2.1''}$$

It follows from Lemma 3.2 of II that $\xi_x\in L^2(\mathbb{R}_+\times S^2)$. [The hypotheses of this lemma are somewhat stronger than (2.5d), and (2.6) would not meet them. However, if

$$|\nabla V(x+x_0)| < C|x|^{\epsilon-1}(1+|x|)^{-2\epsilon-3}, \tag{2.6'}$$

then the hypotheses of Lemma 2.1 and of Lemma 3.2 of II are satisfied.] Since according to Lemma 2.1 $\mathscr{G}_x$ is bounded, $\mathscr{G}_x\xi_x\in L^2(\mathbb{R}_+\times S^2)$. Consequently, unless $\mathscr{G}_x$ has the eigenvalue 1, Eq. (2.1'') has a unique solution η_x in $L^2(\mathbb{R}_+\times S^2)$ that is constructable by the Fredholm method, and we may write it as

$$\eta_x = (\mathbb{1}-\mathscr{G}_x)^{-1}\xi_x.$$

Suppose that $\mathscr{G}_x$ has the eigenvalue ν, so that for $\alpha>0$,

$$\nu f(\alpha) = \int_0^\infty d\beta\;G_x(\alpha+\beta)f(\beta).$$

Define

$$p(k) = \int_0^\infty d\alpha\;e^{-ik\alpha}f(\alpha),$$

so that

$$\int_{-\infty}^\infty dk\,(p(-k),p(k)) = \int_{-\infty}^\infty dk\int_0^\infty d\alpha\int_0^\infty d\beta\;e^{-ik(\alpha+\beta)}(f(\beta),f(\alpha)) = 0,$$

where the inner product is in $L^2(S^2)$. Therefore we get

$$\begin{aligned}
0 &= \int_0^\infty d\alpha\,(\nu f(\alpha),\;\nu f(\alpha)) - \int_0^\infty d\alpha\int_0^\infty d\beta\,(\nu f(\alpha),\;G_x(\alpha+\beta)f(\beta))\\
&= \int_0^\infty dk\,(\nu p(k),\;\nu p(k)) - \int_{-\infty}^\infty dk\,(\nu p(k),\;[\mathscr{S}_x^\dagger(k)-\mathbb{1}]Qp(-k))\\
&= \int_0^\infty dk\,(\nu p(k),\;[\nu p(k)-\mathscr{S}_x^\dagger(k)Qp(-k)])\\
&= \int_0^\infty dk\,([\nu p(k)-\mathscr{S}_x^\dagger(k)Qp(-k)],\\
&\quad\times[\nu p(k)-\mathscr{S}_x^\dagger(k)Qp(-k)])\\
&\quad+\nu\int_{-\infty}^\infty dk\,([\mathscr{S}_x^\dagger(k)Qp(-k)-\nu p(k)],\;p(k))\\
&\quad+(\nu^2-1)\int_{-\infty}^\infty dk\,(p(k),\;p(k))\\
&\quad-\int_{-\infty}^\infty dk\,[(p(-k),\;Q\mathscr{S}_x\mathscr{S}_x^\dagger Qp(-k))\\
&\quad-(p(-k),\;p(-k))].
\end{aligned}$$

The second and fourth integrals vanish. It follows[6] that if

$\nu^2 > 1$ then $p(k) = 0$. Therefore no eigenvalue ν of $\mathscr{G}_x$ can be such that $\nu^2 > 1$.

Consequently, if neither 1 nor -1 is in the spectrum of $\mathscr{G}_x$ then the Neumann series for $(1 - \mathscr{G}_x)^{-1}$ converges. Thus we have proved

Theorem 2.1: If the potential satisfies the hypotheses of Lemma 3.2 of II and those of Lemma 2.1 [which are all satisfied if it obeys (2.6′)] then the generalized Marchenko equation (2.1) has a unique solution in $L^2(\mathbb{R}_+ \times S^2)$, unless its kernel $\mathscr{G}_x$ has the eigenvalue 1. If neither 1 nor -1 is in the spectrum of $\mathscr{G}_x$ then (2.1) can be solved by iteration.

It follows from this theorem that in the absence of bound states, and if $\mathscr{G}_x$ does not have the eigenvalue 1, the reconstruction of V via (2.1) and (2.5) works.

3. CONSTRUCTION AND CHARACTERIZATION

We shall no longer assume the existence of an underlying potential. Instead we start with a given amplitude function $A(k;\theta,\theta')$ with the following properties:

(i) $A(k;\theta,\theta')$ satisfies (2.7) and (2.8).

(ii) For all θ and θ', $A(k;\theta,\theta')$ is a continuous function of k for all $k\in\mathbb{R}$.

(iii) If $A(k)$ is the operator on $L^2(S^2)$ whose kernel is $A(k;\theta,\theta')$, and $S(k) = \mathbb{1} - (k/2\pi i)A(k)$, then for almost all k $S(k)$ is unitary and it satisfies the Levinson theorem for no bound states, namely, $\exists c$ such that

$$\log\ \det S(0) = \lim_{k\to\infty}(\log \det S(k) + ikc).$$

[We note that (ii) implies the existence and continuity of the Fredholm determinant $\det S(k)$ for all k.]

(iv) $kA\in SL^2(\mathbb{R})$, which means that $\exists C$ such that for all $f\in L^2(S^2)$, $\int dk\, k^2 \|A(k)f\|^2 < C\|f\|^2$. This implies that the Fourier transform of $kA(k)$ strongly converges in the mean, and we define $G_x(\alpha;\theta,\theta')$ by means of (2.2′) and (2.4). According to (2.7) the function $G_x(\alpha;\theta,\theta')$ is *real*. Let $G_x(\alpha)$ be the family of operators on $L^2(S^2)$ whose kernels are given by (2.2′) and let $\mathscr{G}_x$ and $\mathscr{G}'_x$ be the self-adjoint operators on $L^2(\mathbb{R}_+ \times S^2)$ whose kernels are given by

$$\mathscr{G}_x(\alpha,\beta;\theta,\theta') = G_x(\alpha+\beta;\theta,\theta'),$$

$$\mathscr{G}'_x(\alpha,\beta;\theta,\theta') = G'_x(\alpha+\beta;\theta,\theta') = -G_x(-\alpha-\beta;\theta,\theta').$$

[Their self-adjointness follows from (2.7) and (2.8).] We then assume

(v) $\mathscr{G}_x$ and $\mathscr{G}'_x$ are compact and $\mathscr{G}_x^2$ and $\mathscr{G}_x'^2$ are Hilbert–Schmidt.

(vi) Save for isolated values of x, neither $\mathscr{G}_x$ nor $\mathscr{G}'_x$ has the eigenvalue 1.

We shall refer to functions that satisfy assumptions (i) to (vi) as in class $\mathscr{A}$. Conditions (i) to (v) are met by potentials that satisfy the hypotheses of Theorem 2.1; condition (vi) may be assumed to hold for such potentials as well.

Since the given amplitude $A\in\mathscr{A}$, Eq. (2.1) has a unique solution $\eta_x\in L^2(\mathbb{R}_+\times S^2)$. We define

$$\gamma_x(k) = \hat{1} + \int_0^\infty d\alpha\, \eta_x(\alpha)e^{ik\alpha}. \tag{3.1}$$

[In the following we shall temporarily drop the subscript x on $\gamma_x(k)$. Its x dependence will not be used for the time being.] It follows from the properties of $\eta_x(\alpha)$ that $\|\gamma(k) - \hat{1}\|\in L^2(\mathbb{R})$ and that $\gamma(k)$ is the boundary value of an analytic function holomorphic in the upper half-plane such that there

$$\lim_{|k|\to\infty} \|\gamma(k) - \hat{1}\| = 0. \tag{3.2}$$

We shall refer to the class of functions with these properties of $\gamma(k) - \hat{1}$ as $\mathscr{H}^+$. It also follows from (2.1) that the function

$$f(k) = \gamma(-k) - \mathscr{S}^\dagger_x(k)Q\gamma(k) \tag{3.3}$$

is in $\mathscr{H}^+$. But because of (2.7), (2.8), and the unitarity of S, f satisfies the relation

$$f(-k) = Q\mathscr{S}_x(k)f(k). \tag{3.4}$$

The fact that $f\in\mathscr{H}^+$ and (3.4) are now utilized in the same manner that led to (2.1) to show that the Fourier transform of $f(k)$,

$$\sigma(\alpha) = \frac{1}{2\pi}\int_{-\infty}^\infty dk\, f(k)e^{-ik\alpha}, \tag{3.5}$$

[which converges in the mean because of (3.3) and the properties of S and γ] vanishes for $\alpha < 0$, and that for $\alpha > 0$ it must solve the homogeneous integral equation

$$\sigma(\alpha) = \int_0^\infty d\beta\, G'_x(\alpha+\beta)\sigma(\beta). \tag{3.6}$$

Since by assumption (vi) 1 is not in the spectrum of $\mathscr{G}'_x$, (3.6) has only the trivial solution $\sigma(\alpha) = 0$, which implies $f(k) = 0$ and hence by (3.3),

$$\gamma(-k) = \mathscr{S}^\dagger_x(k)Q\gamma(k). \tag{3.7}$$

In order to derive the Schrödinger equation we apply the operator $\Delta - 2\theta\cdot\nabla\partial/\partial\alpha$ to Eq. (2.1) and define

$$\Gamma_x(\alpha,\theta) = (\Delta - 2\theta\cdot\nabla\partial/\partial\alpha)\eta_x(\alpha,\theta). \tag{3.8}$$

After an integration by parts and some algebra one finds that $\Gamma_x(\alpha)$ satisfies the equation

$$\Gamma_x(\alpha) = \Gamma_x^{(0)}(\alpha) + \int_0^\infty d\beta\, G_x(\alpha+\beta)\Gamma_x(\beta), \tag{3.9}$$

where

$$\Gamma_x^{(0)}(\alpha,\theta) = -2\int d\theta'\, G_x(\alpha;\theta,\theta')\theta'\cdot\nabla\eta_x(0,\theta').$$

Consequently, if the solution $\eta_x(\alpha)$ is *miraculous*, i. e., if it is such that $\theta\cdot\nabla\eta_x(0,\theta)$ is independent of θ, then we may define

$$V(x) = -2\theta\cdot\nabla\eta_x(0,\theta) \tag{3.10}$$

and write

$$\Gamma_x^{(0)}(\alpha) = V(x)G_x(\alpha)\hat{1}.$$

Because (2.1) has a unique solution, comparison of (2.1) and (3.9) allows us to conclude that

$$\Gamma_x(\alpha) = V(x)\eta_x(\alpha).$$

In view of (3.8) this means that $\eta_x(\alpha,\theta)$ satisfies the partial differential equation

$$[\Delta - 2\theta\cdot\nabla\partial/\partial\alpha - V(x)]\eta_x(\alpha,\theta) = 0. \tag{3.11}$$

Use of (3.1) now shows that therefore $\gamma_x(k,\theta)$ satisfies the equation

$$[\Delta + 2ik\theta\cdot\nabla - V(x)]\gamma_x(k,\theta) = 0.$$

Finally we define

$$\psi(k,\theta,x) = \gamma_x(k,\theta)\exp(ik\theta\cdot x) \tag{3.12}$$

and we find that ψ satisfies the Schrödinger equation,

$$(\Delta + k^2)\psi = V\psi, \tag{3.13}$$

with the potential V defined by (3.10).

As indicated in II it is sometimes more convenient to rewrite Eq. (2.1) in the form

$$\check{\eta}_x(\alpha,\theta) = \int d\theta'\,\check{g}(\alpha + x\cdot\theta';\theta,\theta') + \int d\theta' \int_{x\cdot\theta'}^{\infty} d\beta\,\check{g}(\alpha+\beta;\theta,\theta')\check{\eta}_x(\beta,\theta'), \tag{3.14}$$

where

$$\check{\eta}_x(\alpha,\theta) = \eta_x(\alpha - \theta\cdot x,\theta) \tag{3.15}$$

and

$$\check{g}(\alpha;\theta,\theta') = -i(2\pi)^{-2}\int_{-\infty}^{\infty} dk\,kA^*(k;\theta,-\theta')e^{ik\alpha}. \tag{3.16}$$

One then finds that if $\check{\eta}_x(\alpha,\theta)$ satisfies (3.14) then

$$\lim_{|x|\to\infty} |x|\,\check{\eta}_x(\alpha + |x|,\theta) = 2\pi\int_\alpha^\infty d\beta\,\check{g}(\beta;-\hat{x},\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,A(k;\hat{x},\theta)e^{-ik\alpha}. \tag{3.17}$$

Use of (3.12) and (3.1) leads to

$$\psi(k,\theta,x) = e^{ik\theta\cdot x} + \int_{\theta\cdot x}^{\infty} d\alpha\,\check{\eta}_x(\alpha,\theta)e^{ik\alpha}, \tag{3.18}$$

and hence from (3.17),

$$\lim_{|x|\to\infty} |x|\,[\psi(k,\theta,x) - e^{ik\theta\cdot x}]e^{-ik|x|} = A(k;\hat{x},\theta). \tag{3.19}$$

We therefore have proved

Theorem 3.1 (Partial Characterization): Suppose that $A(k;\theta,\theta')\in\mathscr{A}$. Then Eq. (2.1), with G_x constructed from A, has a unique solution η_x in $L^2(\mathbb{R}\times S^2)$. Suppose further that this solution $\eta_x(\alpha,\theta)$ is *miraculous*, i. e., $\theta\cdot\nabla\eta_x(0,\theta)$ is independent of θ. Then the function (3.12), given in terms of the Fourier transform (3.1) of $\eta_x(\alpha,\theta)$, satisfies the Schrödinger equation (3.13) with the potential V given by (3.10). Furthermore, the outgoing and incoming wavesolutions are connected by (3.7), in which the S matrix is given by $S = \mathbb{1} - (k/2\pi i)A$, and the asymptotic form of the wavefunction is given by (3.19).

This theorem implies that the essential *characterization* of a scattering amplitude associated with a local potential is the *miracle*. However, it gives only a partial characterization because it does not state the properties of the associated potential, i. e., the class to which it belongs. It should also be noted that class $\mathscr{A}$ [property (iii)] implies that there are no bound states. At this time a construction procedure, in contrast to a reconstruction procedure, for potentials that cause bound states is unknown.

4. THE GENERALIZED MARCHENKO EQUATION WITH BOUND STATES

In II we removed the bound-state poles from $\psi(k,\theta,x)$ by multiplying it by the Fredholm determinant $D(k)$ of the Lippmann–Schwinger equation. This gave rise to the presence of the factor $R(k) = D(-k)/D(k)$ in the kernel of the generalized Marchenko equation, so that $\mathscr{S}_x^\dagger(k)$ in (2.2) is replaced by $\mathscr{S}_x^\dagger(k)R(k) = (\mathscr{S}_x^\dagger - \mathbb{1})R + (R-1)\mathbb{1}$. The operator $\mathscr{G}_x$ on $L(\mathbb{R}_+\times S^2)$ is thereby replaced by a sum of two terms

$$\mathscr{G}_x\bar{\mathscr{G}}_x = \bar{\mathscr{G}}_x^{(1)} + \bar{\mathscr{G}}^{(2)},$$

the first of which is defined by a kernel just like (2.2′), except that A_x^* is replaced by A_x^*R; and the second is the tensor product of the unit operator on $L^2(S^2)$ and the operator $\mathscr{g}$ on $L^2(\mathbb{R}_+)$ whose kernel is

$$g(\alpha,\beta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,[R(k) - 1]e^{ik(\alpha+\beta)},$$

$$\bar{\mathscr{G}}^{(2)} = \mathscr{g}\otimes\mathbb{1}.$$

It is not difficult to generalize Lemma 2.1 for $\bar{\mathscr{G}}_x^{(1)}$, i. e., to prove the compactness of $\bar{\mathscr{G}}_x^{(1)}$, as well as to prove that $\mathscr{g}$ is compact. One then constructs $(\mathbb{1} - \bar{\mathscr{G}}_x)^{-1}$ as follows:

$$(\mathbb{1} - \bar{\mathscr{G}}_x)^{-1} = [\mathbb{1} - (\mathbb{1} - \bar{\mathscr{G}}^{(2)})^{-1}\bar{\mathscr{G}}_x^{(1)}]^{-1}(\mathbb{1} - \bar{\mathscr{G}}^{(2)})^{-1},$$

in which

$$(\mathbb{1} - \bar{\mathscr{G}}^{(2)})^{-1} = (\mathbb{1} - \mathscr{g})^{-1}\otimes\mathbb{1}.$$

This procedure works, provided that neither $\mathscr{g}$ nor $(\mathbb{1} - \mathscr{g})^{-1}\bar{\mathscr{G}}_x^{(1)}$ have the eigenvalue 1. Unfortunately, as we show in Appendix B, when there are bound states $\mathscr{g}$ always has 1 as an eigenvalue. Therefore, though Theorem 4.1, of II is correct, in the presence of bound states it is not very useful.

We therefore give here another method of dealing with bound states in the generalized Marchenko equation is analogous to the method used in II for removing the bound states from S before finding the Jost function. The matter is slightly different here because of the need to remove them not only from S but from $\mathscr{S}_x$ for all x.

Let $Y_{\kappa_m}^b(\theta)$, $b = 1,\ldots,N_m$, be characters of the N_m-fold degenerate eigenvalue $-\kappa_m^2$ as defined in Sec. 5 of II. As shown there they can, in principle, be obtained from the forward scattering amplitude and its angle derivatives. Let $\mathscr{H}_m^x$ be the N_m-dimensional subspace of $L^2(S^2)$ spanned by the functions $Y_{\kappa_m}^b(-\theta)\exp(\kappa_m\theta\cdot x)$, $b = 1,\ldots,N_m$, with fixed m. (If there are l bound states then there are l such subspaces $\mathscr{H}_m^x$, $m = 1,\ldots,l$, but they are not necessarily disjoint.) The functions $Y_{\kappa_m}^b(\theta)\exp(\kappa_m\theta\cdot x)$ are the characters of the potential translated by x, and they can be chosen to be *real*. Let $B_m = B_m^2 = B_m^\dagger$, $m = 1,\ldots,l$, be the set of projections on $C_m^{-1}\mathscr{H}_m^x$, where

$$C_m = \left(\mathbb{1} + B_1\frac{2\kappa_1}{\kappa_m - \kappa_1}\right)\cdots\left(\mathbb{1} + B_{m-1}\frac{2\kappa_{m-1}}{\kappa_m - \kappa_{m-1}}\right). \tag{4.1}$$

The operators B_m depend on x, and the requirement of self-adjointness makes this dependence complicated.[7] We then define the family of operators

$$\Pi(k) = \left(\mathbb{1} + B_1 \frac{2i\kappa_1}{k - i\kappa_1}\right) \cdots \left(\mathbb{1} + B_l \frac{2i\kappa_l}{k - i\kappa_l}\right) \quad (4.2)$$

which, for real k, are *unitary* and obey the relation

$$\Pi(-k) = \Pi^*(k).$$

(By Π^* we mean the operator whose kernel is the complex conjugate of that of Π.)

The reduced functions $\gamma^{\rm red}$ are now defined by

$$\gamma^{\rm red}(k) = \Pi^{-1}(k)\gamma(k). \quad (4.3)$$

The residue of $\gamma(k)$ at $k = i\kappa_m$ being an operator whose range is $\mathscr{H}_m^x$ [see (5.8) of II], and

$$\Pi^{-1}[i\kappa_m] = \cdots(\mathbb{1} - B_m)C_m^{-1},$$

the definition of B^m assures that $\Pi^{-1}(i\kappa_m)$ annihilates the residue of $\gamma(k)$ at $i\kappa^m$. Therefore $\gamma_{\rm red}(k) \in \mathscr{H}^+$.

Corresponding to the definition (4.3) we define a reduced S matrix

$$\mathscr{S}_x^{\rm red}(k) = Q[\Pi(k)]^{-1}Q\mathscr{S}_x(k)\Pi(-k), \quad (4.4)$$

so that (3.7) leads to the equation

$$\gamma_{\rm red}(-k) = (\mathscr{S}_x^{\rm red}(k))^\dagger Q\gamma^{\rm red}(k).$$

Furthermore, $\mathscr{S}_x^{\rm red}$ is unitary; it satisfies $\mathscr{S}_x^{\rm red}(-k) = \mathscr{S}_x^{\rm red}(k)^*$, and reciprocity, $\tilde{\mathscr{S}}_x^{\rm red} = Q\mathscr{S}_x^{\rm red}Q$. Since

$$\det\Pi(k) = \prod_1^l \left(\frac{k + i\kappa_m}{k - i\kappa_m}\right)^{N_m},$$

we have

$$\det\mathscr{S}_x^{\rm red} = \prod_1^l \left(\frac{k + i\kappa_m}{k - i\kappa_m}\right)^{2N_m} \det S. \quad (4.5)$$

As a result, if S satisfies the Levinson theorem appropriate to l eigenvalues with a total of ΣN_m eigenfunctions, then $\mathscr{S}_x^{\rm red}$ satisfies the same for no eigenvalues, i. e., condition (iii) of Sec. 3.

It is now straightforward to prove Lemma 4.1: Define

$$G_x^{\rm red}(\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, [\mathscr{S}_x^{\rm red}(k) - \mathbb{1}]^\dagger Q e^{ik\alpha},$$

and let $\mathscr{G}_x^{\rm red}$ be the self-adjoint operator on $L^2(\mathbb{R}_+ \times S^2)$ whose kernel is $\mathscr{G}_x^{\rm red}(\alpha,\beta;\theta,\theta') = G_x^{\rm red}(\alpha + \beta,\theta,\theta')$. Then on the same hypotheses as in Lemma 2.1, $\mathscr{G}_x^{\rm red}$ is compact, its square is Hilbert-Schmidt, and $\exists C$ such that for all $x \|\mathscr{G}_x^2\|_2 < C$.

The proof will be given in Appendix C.

As a result of this lemma Theorem 2.1 is applicable to the generalized Marchenko equation for $\alpha > 0$,

$$\eta_x^{\rm red}(\alpha) = G_x^{\rm red}(\alpha)\hat{1} + \int_0^\infty d\beta\, G_x^{\rm red}(\alpha + \beta)\eta_x^{\rm red}(\beta). \quad (4.6)$$

Then by (4.3)

$$\eta_x(\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ik\alpha}[\gamma(k) - \hat{1}]$$

$$= \eta_x^{\rm red}(\alpha) + \Omega_x(\alpha)\hat{1} + \int_0^\infty d\beta\, \Omega_x(\alpha - \beta)\eta_x^{\rm red}(\beta),$$

where

$$\Omega_x(\alpha) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, e^{-ik\alpha}[\Pi(k) - \mathbb{1}].$$

The explicit form (4.2) of $\Pi(k)$ allows us to conclude that $\Omega_x(\alpha) = 0$ for $\alpha > 0$, and for $\alpha < 0$,

$$\Omega_x(\alpha) = -2\sum_m \kappa_m e^{\kappa_m\alpha} C_m B_m D_m, \quad (4.7)$$

where C_m is given by (4.1) and

$$D_m = \left(\mathbb{1} + B_{m+1}\frac{2\kappa_{m+1}}{\kappa_m - \kappa_{m+1}}\right) \cdots \left(\mathbb{1} + B_l \frac{2\kappa_l}{\kappa_m - \kappa_l}\right).$$

Consequently for $\alpha > 0$,

$$\eta_x(\alpha) = \eta_x^{\rm red}(\alpha) + \int_\alpha^\infty d\beta\, \Omega_x(\alpha - \beta)\eta_x^{\rm red}(\beta), \quad (4.8)$$

while for $\alpha < 0$,

$$\eta_x(\alpha) = \Omega_x(\alpha)\hat{1} + \int_0^\infty d\beta\, \Omega_x(\alpha - \beta)\eta_x^{\rm red}(\beta). \quad (4.8')$$

Now if we insert the Fourier transform (3.1) into the Schrödinger equation $(\Delta + 2ik\theta\cdot\nabla)\gamma_x(k,\theta) = V\gamma_x(k,\theta)$, we find that for $\alpha > 0$ and for $\alpha < 0$, $\eta_x(\theta,\alpha)$ must be a solution of the differential equation

$$\left[\Delta - 2\frac{\partial}{\partial\alpha}\theta\cdot\nabla - V(x)\right]\eta_x(\theta,\alpha) = 0,$$

and it must satisfy the boundary condition across its discontinuity at $\alpha = 0$,

$$V(x) = -2\theta\cdot\nabla[\eta_x(\theta,0+) - \eta_x(\theta,0-)]. \quad (4.9)$$

Insertion of (4.8) and (4.8') in this equation leads to

$$V(x) = -2\theta\cdot\nabla[\eta_x^{\rm red}(\theta,0+) - \Omega_x(\theta)], \quad (4.10)$$

where by (4.7),

$$\Omega_x(\theta) = -2\sum_m \kappa_m \int d\theta'[C_m B_m D_m](\theta,\theta'). \quad (4.7')$$

The reconstruction is now accomplished by solving (4.6) and using (4.10).

On the other hand, if the existence of an underlying potential is not known then a construction procedure analogous to that of Sec. 3 is enormously complicated by the fact that Π depends on x, and a method such as using (3.10) does not appear to be feasible. As a result there is at the present time no known construction or characterization in the presence of bound states.

5. CRITICAL REMARKS ON THE GENERALIZED GL EQUATION

The method of Sec. 8 of II, based on Sec. 7, is flawed. The first criticism is that the function $\phi(k,\theta,x)$ defined by (7.1) of II is not known to be of exponential order $|\theta\cdot x|$ as a function of k. Eq. (7.1) leads only to the conclusion that ϕ is an entire analytic function of k of exponential order $|x|$. Therefore the Povsner–Levitan representation (7.3) has to be replaced by

$$\phi(k,\theta,x) = e^{ik\theta\cdot x} - \int_{-|x|}^{|x|} d\alpha\, q(x,\theta,\alpha)e^{ik\alpha}, \quad (5.1)$$

where $q(x,-\theta,-\alpha) = q(x,\theta,\alpha)$. Insertion of (5.1) in the Schrödinger equation leads to the conclusion that q must have a discontinuity at $\alpha = \theta\cdot x$, that for $\alpha < \theta\cdot x$ and for $\alpha > \theta\cdot x$ it must satisfy the partial differential equation

$$\left[\frac{\partial^2}{\partial\alpha^2} - \Delta + V(x)\right]q = 0, \quad (5.2)$$

and that it must satisfy the boundary conditions

$$q(x,\theta,|x|-)=0, \tag{5.3}$$

$$2\theta\cdot\nabla[q(x,\theta,\theta\cdot x+)-q(x,\theta,\theta\cdot x-)]=V(x). \tag{5.4}$$

We may also define the three-dimensional Fourier transform (7.8) of II,

$$\phi(k,\theta,x)=e^{ik\theta\cdot x}-\int d^3y\, e^{ik\theta\cdot y}\check{h}(x,y), \tag{5.5}$$

so that $q(x,\theta,\alpha)$ may be taken to be the Radon transform of $\check{h}(x,y)$,

$$q(x,\theta,\alpha)=\int d^3y\,\delta(\alpha-\theta\cdot y)\check{h}(x,y). \tag{5.6}$$

Insertion of (5.6) in (5.2)–(5.4) leads to the partial differential equation for $\check{h}$,

$$[\Delta_y-\Delta_x+V(x)]\check{h}(x,y)=V(x)\delta^3(x-y). \tag{5.7}$$

Introduction of the variables $u=\frac{1}{2}(x+y)$, $v=x-y$, leads to the equation

$$[V(u+\tfrac{1}{2}v)-2\nabla_u\cdot\nabla_v]\check{h}(u+\tfrac{1}{2}v,\ u-\tfrac{1}{2}v)=V(u)\delta^3(v) \tag{5.7'}$$

together with the requirement that

$$\check{h}(u+\tfrac{1}{2}v,\ u-\tfrac{1}{2}v)=0\quad\text{for }u\cdot v<0.$$

Integrating (5.7) over a sphere of radius ϵ in v space, we obtain the formula for V,

$$V(u)=-2\lim_{\epsilon\to 0}\epsilon^2\int_{\theta\cdot u>0}d\theta\,\theta\cdot\nabla_u\check{h}(u+\tfrac{1}{2}\epsilon\theta,\ u-\tfrac{1}{2}\epsilon\theta). \tag{5.8}$$

It is easily seen that (5.7) used in (5.5) leads to the Schrödinger equation for ϕ.

Alternatively we may require that, for $|y|<|x|$, $\check{h}$ satisfy the partial differential equation

$$[\Delta_y-\Delta_x+V(x)]\check{h}(x,y)=0. \tag{5.7''}$$

If we then insert the representation

$$\phi(k,\theta,x)=e^{ik\theta\cdot x}-\int_{|y|<|x|}d^3y\,e^{ik\theta\cdot y}\check{h}(x,y) \tag{5.5'}$$

in the Schrödinger equation we find that for $|x|=|y|$, $\check{h}$ must be related to the potential by the formula

$$V(x)\delta(\hat{x},\hat{y})=-2\hat{x}\cdot\nabla_x[\check{h}(|x|\hat{x},|x|\hat{y})|x|^2]. \tag{5.9}$$

The crucial question now is whether from the fact that the support of $q(x,\theta,\alpha)$ as a function of α lies in the interval $|\alpha|\leqslant|x|$, one may conclude that the support of $\check{h}(x,y)$ as a function of y is confined to the ball $|y|\leqslant|x|$. As q is discontinuous the best applicable theory of the Radon transform appears to be that in distribution spaces, and the desired result follows,[8] *provided* that q meets the (infinitely many) moment conditions that for all n and all $l>n$, $|m|\leqslant l$,

$$\int d\theta\,Y_l^m(\theta)\int_{-|x|}^{|x|}d\alpha\,\alpha^n q(x,\theta,\alpha)=0.$$

Whether, in fact, q satisfied these conditions is unknown; therefore the support question for $\check{h}(x,y)$ cannot be regarded as settled.

If the support of $\check{h}(x,y)$ is contained in $|y|\leqslant|x|$ then the generalized GL equation (8.4) of II is readily derived as in Sec. 8 of II. A generalized GL equation for $q(x,\theta,\alpha)$ then follows by a *partial* Radon transform. One can also readily derive further generalized GL equations for nonzero "reference potentials" as well as a generalized *nonlinear* GL equation. What is more, if the potential is central then expansion of $\check{h}$ on the basis of the spherical harmonics leads from (8.4) of II to the well-known radial GL equations. However, since the foundation of the GL method is the "triangularity" of $\check{h}(x,y)$, and this support problem is still unsolved, we shall postpone the publication of the details of these statements until it is.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant No. PHY 77-25337. I am indebted to Margaret Cheney for some useful discussions.

APPENDIX A

We shall prove here Lemma 2.1

As a preliminary we note that properties (2.5d) and (2.5e) imply that

$$\int_0^\infty dt\,t^2M_2(t)<\infty, \tag{A1}$$

$$\int_0^\infty dt\,t^{3/2}M_2(t)<\infty. \tag{A2}$$

As a further preliminary we shall prove that for each $0<\alpha\leqslant 1$ there exists a $C<\infty$ such that

$$I=\int d\theta''\,|\theta-\theta''|^{\alpha-2}|\theta'-\theta''|^{\alpha-2}<C\,|\theta-\theta'|^{2\alpha-2} \tag{A3}$$

for all $\theta,\theta'\in S^2$.

We use $\theta+\theta'$ as the z axis and $\theta-\theta'$ as the x axis in the θ'' integration, setting $(\theta+\theta')\cdot\theta''=|\theta+\theta'|u$, $(\theta-\theta')\cdot\theta''=|\theta-\theta'|(1-u^2)^{1/2}\cos\phi$, so that $(\theta'-\theta'')^2(\theta-\theta'')^2=(2-|\theta+\theta'|u)^2-|\theta-\theta'|(1-u^2)\cos^2\phi$, and hence

$$\mathrm{I}=\int_{-1}^1du\int_0^{2\pi}d\phi\,[4(1-\tfrac{1}{2}|\theta+\theta'|u)^2-|\theta-\theta'|(1-u^2)\cos^2\phi\,]^{1/2\alpha-1}.$$

It is easily seen that for $\alpha<1$ the integral

$$\int_0^\pi d\phi\,|a^2-\cos^2\phi\,|^{1/2\alpha-1}=2\int_0^1dt\,(1-t^2)^{-1/2}|a^2-t^2|^{1/2\alpha-1}$$

is $O(|a^2-1|^{1/2\alpha-1/2})$ near $a^2=1$. Since it is $O(|a^2-1|^{1/2\alpha-1})$ as $a^2\to\infty$, $\exists C$ such that for all a,

$$\int_0^\pi d\phi\,|a^2-\cos^2\phi\,|^{1/2\alpha-1}<C\frac{|a^2-1|^{1/2\alpha-1/2}}{(1+|a^2-1|)^{1/2}}.$$

It follows that if $\alpha^2\geqslant b^2$ then

$$\int_0^\pi d\phi\,(a^2-b^2\cos^2\phi\,)^{1/2\alpha-1}\leqslant C\,(a^2-b^2)^{1/2\alpha-1/2}|a|^{-1}.$$

In the present instance $a^2=(2-|\theta+\theta'|u)^2$, $a^2-b^2=(|\theta+\theta'|-2u)^2$. Therefore,

$$I \leqslant C \int_{-1}^{1} du \; |(|\theta + \theta'| - 2u)|^{\alpha - 1} |(2 - |\theta + \theta'|u)|^{-1}.$$

This integral converges for all θ and θ', except when $\theta = \theta'$, where one easily finds that it is $O(|\theta - \theta'|^{2\alpha - 2})$. Thus inequality (A3) follows for $\alpha < 1$. For $\alpha = 1$ the argument has to be modified and is left to the reader.

We next consider the following integrals:

$$I_n(k,\theta,\theta') = \int d^3x_0 \cdots d^3x_n \, f(x_0,\ldots,x_n) \times \exp\{ik\,[\theta\cdot x_0 - \theta'\cdot x_n + h\,(x_0,\ldots,x_n)]\},$$

where

$$f(x_0,\ldots,x_n) = \frac{V(x_0)\cdots V(x_n)}{|x_0 - x_1||x_1 - x_2|\cdots|x_{n-1} - x_n|},$$

$$h\,(x_0,\ldots x_n) = |x_0 - x_1| + |x_1 - x_2| + \cdots + |x_{n-1} - x_n|.$$

We have the Fourier transform

$$\hat{I}_n(\alpha,\theta,\theta') = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\; e^{ik\alpha} k I_n(k,\theta,\theta')$$

$$= i\int d^3x_0 \cdots d^3x_n\; \delta[\alpha + \theta\cdot x_0 - \theta'\cdot x_n + h\,(x_0,\ldots,x_n)]\theta\cdot\nabla_0 f,$$

and hence

$$\int_{-\infty}^{\infty} d\alpha\; |\hat{I}_n(\alpha,\theta,\theta')|^2 = \int d^3x_0 \cdots d^3x_n d^3x_0' \cdots d^3x_n' \times \delta[\theta\cdot(x_0 - x_0') - \theta'\cdot(x_n - x_n') + h\,(x_0,\ldots,x_n) - h\,(x_0',\ldots,x_n')] \times \theta\cdot\nabla_0 f(x_0,\ldots,x_n)\theta\cdot\nabla_0' f(x_0',\ldots,x_n').$$

The integral $\int d\alpha\; \alpha|\hat{I}_n|^2$ has a factor of $[\theta.x_0' - \theta'\cdot x_n' + h\,(x_0',\ldots,x_n')] < 2(|x_0'| + \cdots + |x_n'|)$ in its integrand. Therefore for $m = 0, 1,$

$$\int_0^{\infty} d\alpha\; \alpha^m \int d\theta\, d\theta'\; |\hat{I}_n(\alpha,\theta,\theta')|^2 < C\int d^3x_0 \cdots d^3x_n' \times |\nabla_0 f(x_0,\ldots,x_n)|\,|\nabla_0' f(x_0',\ldots,x_n')| \times \frac{(|x_0'| + \cdots + |x_n'|)^m}{|x_0 - x_0'|}.$$

If V satisfies (2.5a) then the integrals over $x_1,\ldots,x_n$ and x_1' $,\ldots,x_n'$ all converge and are uniformly bounded (with the $|x_0'|$ term in the numerator included). The subsequent integrals over x_0 and x_0' are all of the form $\int d^3x\, d^3y\, |V(x)V(y)|/|x - y|$, $\int d^3x\, d^3y |V(x)||\nabla V(y)|/|x - y|$, or $\int d^3x\, d^3y \times|\nabla V(x)||\nabla V(y)|/|x - y|$. The first converges by (2.5a); the second and third converge by (2.5a) and (A2). We therefore have for $m = 0$, 1 and all $n > 1$,

$$\int_0^{\infty} d\alpha\; \alpha^m \int d\theta\, d\theta'\; |\hat{I}_n(\alpha,\theta,\theta')|^2 < \infty. \qquad \text{(A4)}$$

We now write

$$-4\pi A\,(k;\theta,\theta') = \int d^3x\, d^3y\; V(x)[(\mathbb{1} - G_0V)^{-1}](x,y) \times \exp[ik\,(\theta'\cdot y - \theta\cdot x)] = \sum_0^m A^{(n)} + R^{(m)},$$

where G_0 is the Green's function $-e^{ik|x-y|}/4\pi|x - y|$, and

$$A^{(0)} = \int d^3x\; V(x)\exp[ikx\cdot(\theta' - \theta)],$$

$$A^{(n)} = \int d^3x\, d^3y\; V(x)(G_0V)^n(x,y)\exp[ik\,(\theta'\cdot y - \theta\cdot x)],$$

$$R^{(m)} = \int d^3x\, d^3y\; V(x)[(\mathbb{1} - G_0V)^{-1}(G_0V)^{m+1}](x,y) \times\exp[ik\,(\theta'\cdot y - \theta\cdot x)].$$

We also define

$$\hat{A}\,(\alpha;\theta,\theta') = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\; ke^{ik\alpha}A\,(k;\theta,\theta') = \sum_0^m \hat{A}^{(n)} + \hat{R}^{(m)}.$$

For $n \geqslant 1$, $A^{(n)}$ is of the form I_n and hence $\hat{A}^{(n)}$ is of the form $\hat{I}_n$. Therefore by (A4), for $m = 0$, 1 and $n \geqslant 1$,

$$\int_0^{\infty} d\alpha\; \alpha^m \mathrm{tr}[\hat{A}^{(n)}(\alpha)\hat{A}^{(n)\dagger}(\alpha)] < \infty \qquad \text{(A5)}$$

if $\hat{A}^{(n)}(\alpha)$ is the operator on $L^2(S^2)$ whose kernel is $\hat{A}^{(n)}(\alpha;\theta,\theta')$.

We next consider R_n:

$$R_n(k,\theta,\theta') = \int d^3x\, d^3y\; V^{1/2}(x)(\check{K}K^n)(x,y)\,|V(y)|^{1/2} \times\exp[ik\,(\theta'\cdot y - \theta\cdot x)],$$

where $V^{1/2} = |V|^{1/2}\mathrm{sgn}V$, $K = |V|^{1/2}G_0V^{1/2}$, and $\check{K} = |V|^{1/2}GV^{1/2} = (\mathbb{1} - K)^{-1}K$. Therefore

$$k^2\mathrm{tr}R_nR_n^\dagger \leqslant C\int \frac{d^3x\, d^3x'\, d^3y\, d^3y'}{|x - y'||y - y'|}\,|V(x)V(x')V(y)V(y')|^{1/2} \times|(\check{K}K^n)(x,y)|\,|\check{K}K^n(x',y')|.$$

Since $\check{K}$ is uniformly bounded it follows from the corollary to Lemma 2.1 of II that for $n \geqslant 2$,

$$\int_{-\infty}^{\infty} dk\; \|\check{K}K^n\|_2^2 < \infty.$$

Therefore by Schwarz's inequality and (2.5a), for $n \geqslant 2$,

$$\int_0^{\infty} d\alpha\; \mathrm{tr}\hat{R}_n\hat{R}_n^\dagger \leqslant \int_{-\infty}^{\infty} dk\; k^2\mathrm{tr}R_nR_n^\dagger < \infty. \qquad \text{(A6)}$$

We must also examine the derivative of R_n with respect to k. When the derivative acts on $\exp[ik\,(\theta'\cdot y - \theta\cdot x)]$ it simply brings in an additional factor of $|x|$ or $|y|$, which by (2.5a) still leads to a finite result. When it acts on $\check{K}K^n$ we use

$$\frac{\partial}{\partial k}\check{K} = (\mathbb{1} - K)^{-1}\frac{\partial K}{\partial k}(\mathbb{1} - K)^{-1}.$$

But since both $(\mathbb{1} - K)^{-1}$ and $\partial K/\partial k$ are uniformly bounded operator families, we have for $n \geqslant 3$,

$$\int_{-\infty}^{\infty} dk\; k^2\mathrm{tr}\frac{\partial R_n}{\partial k}\frac{\partial R_n^\dagger}{\partial k} < \infty,$$

and hence

$$\int_0^{\infty} d\alpha\; \alpha^2\mathrm{tr}\hat{R}_n\hat{R}_n^\dagger < \infty,$$

as well as by (A6),

$$\int_0^\infty d\alpha\, \alpha \operatorname{tr}\hat{R}_n \hat{R}_n^\dagger < \infty. \tag{A7}$$

Now the operator $\mathscr{G}_x$ on $L^2(\mathbb{R}_+ \times S^2)$ is such that by (2.2′),

$$\|\mathscr{G}_x\|_2^2 = \int_0^\infty d\alpha \int_0^\infty d\beta \int d\theta\, d\theta' \,|G_x(\alpha+\beta;\theta,\theta')|^2$$
$$= (2\pi)^{-2} \int_0^\infty d\alpha\, \alpha \operatorname{tr}\hat{A}_x(\alpha)\hat{A}_x^\dagger(\alpha)$$

by integration by parts. If we write $A = A^{(0)} + A'$ and correspondingly for $\mathscr{G}_x$, then (A5) and (A7) imply that

$$\|\mathscr{G}_x'\|_2 < C < \infty. \tag{A8}$$

What is more, since A_x is the scattering amplitude of a potential that is translated by x, and the constant C in (2.5a) and the functions M_1 and M_2 in (2.5b) and (2.5c) are unchanged by such a translation, the constant C in (A8) is independent of x.

We must now separately consider $A^{(0)}$.

$$G_x^{(0)}(\alpha;\theta,\theta')$$
$$= -i(2\pi)^{-2}\int_{-\infty}^\infty dk\, k e^{ik\alpha} A_x^{(0)}(k;\theta,\theta')$$
$$= \frac{1}{8\pi^2}\frac{\partial}{\partial\alpha}\int d^3y\, V(y+x)\delta[\alpha + (\theta'-\theta)\cdot y]$$
$$= \frac{1}{8\pi^2}|\theta-\theta'|^{-2}\int d^3y\, [(\theta'-\theta)\cdot\nabla_y V(y+x)]$$
$$\times \delta[\alpha + (\theta'-\theta)\cdot y].$$

Therefore by (2.5c),

$$|G_x^{(0)}(\alpha;\theta,\theta')|$$
$$\leqslant \frac{1}{8\pi^2(\theta-\theta')^2}\int d^2y_\perp M_2\left[\left(|y_\perp|^2 + \frac{\alpha'^2}{(\theta-\theta')^2}\right)^{1/2}\right]$$
$$= \frac{1}{8\pi(\theta-\theta')^2}\int_0^\infty dt^2 M_2\left[\left(t^2 + \frac{\alpha'^2}{(\theta-\theta')^2}\right)^{1/2}\right]$$
$$= \frac{1}{4\pi(\theta-\theta')^2}F^2\left(\frac{\alpha'}{|\theta-\theta'|}\right),$$

where $\alpha' = |\alpha - 2|x_0| - 2|x||$ and F is defined by (2.5e). Since F is monotone, by (2.5e),

$$|G_x^{(0)}(\alpha;\theta,\theta')|$$
$$\leqslant \frac{1}{4\pi}F(\tfrac{1}{2}\alpha')\left(\frac{\alpha'}{|\theta-\theta'|}\right)^\epsilon F\left(\frac{\alpha'}{|\theta-\theta'|}\right)\alpha'^{-\epsilon}|\theta-\theta'|^{\epsilon-2}$$
$$\leqslant CF(\tfrac{1}{2}\alpha')\alpha'^{-\epsilon}|\theta-\theta'|^{\epsilon-2}. \tag{A9}$$

Now write

$$B(\alpha,\beta;\theta,\theta') = \int_0^\infty d\gamma \int d\theta''\, G_x^{(0)}(\alpha+\gamma;\theta,\theta'')$$
$$G_x^{(0)}(\beta+\gamma;\theta'',\theta').$$

Then by (A3) and (A9)

$$|B(\alpha,\beta;\theta,\theta')| \leqslant CF^{1/2}(\tfrac{1}{2}\alpha)F^{1/2}(\tfrac{1}{2}\beta)\alpha^{-1/2\epsilon}\beta^{-1/2\epsilon}$$
$$\int_0^\infty d\gamma\gamma'^{-\epsilon}F(\tfrac{1}{2}\gamma')$$
$$\times\int d\theta''|\theta-\theta''|^{\epsilon-2}|\theta'-\theta''|^{\epsilon-2}$$
$$\leqslant C|\theta-\theta'|^{2\epsilon-2}\alpha^{-1/2\epsilon}\beta^{-1/2\epsilon}F^{1/2}(\tfrac{1}{2}\alpha)$$
$$F^{1/2}(\tfrac{1}{2}\beta)\int_0^\infty d\gamma\,\gamma'^{-\epsilon}F(\tfrac{1}{2}\gamma').$$

Therefore, by (2.5e)

$$\int_0^\infty d\alpha\, d\beta \int d\theta\, d\theta'\, |B(\alpha,\beta;\theta,\theta')|^2$$
$$< C\left[\int_0^\infty ds\, s^{-\epsilon}F(\tfrac{1}{2}s)\right]^4 \int d\theta\, d\theta'\, |\theta-\theta'|^{4\epsilon-4}.$$

The last integral is of the form $\int_{-1}^1 du\,(1-u)^{2\epsilon-2}$, which converges for $\epsilon > \frac{1}{2}$. The first integral converges for $\frac{1}{2} < \epsilon < 1$, by (2.5e). Thus if $\frac{3}{4} < \epsilon < 1$ both converge, and

$$\|\mathscr{G}_x^{(0)2}\|_2^2 = \int_0^\infty d\alpha \int_0^\infty d\beta\, \|B(\alpha,\beta)\|_2^2 \leqslant C < \infty \tag{A10}$$

for all x.

One also readily proves by means of (A9), (A3), and (2.5e) that $\mathscr{G}_x^{(0)}$ is a bounded (uniformly in x) operator. Now (A8) implies that $\|\mathscr{G}_x'^2\|_2 \leqslant \|\mathscr{G}_x'\|_2^2 \leqslant C < \infty$, and $\|\mathscr{G}_x^{(0)}\mathscr{G}_x'\|_2 \leqslant \|\mathscr{G}_x^{(0)}\|\|\mathscr{G}_x'\|_2 \leqslant C < \infty$. Therefore $\|\mathscr{G}_x^2\|_2 \leqslant C < \infty$ and $\mathscr{G}_x$ is *boundea*. Since it is self-adjoint if follows that it is *compact*.

APPENDIX B

It is instructive to solve the scalar Hilbert problem by the same method used to solve (3.7). We are to find a function $D(k)$ that is the boundary value of an analytic function holomorphic in[9] $\mathbb{C}^+$, and such that there

$$\lim_{|k|\to\infty} D(k) = 1, \tag{B1}$$

and such that

$$D(-k) = R(k)D(k), \tag{B2}$$

where $R(k)$ is continuous, $R(-k) = R^*(k) = 1/R(k)$ and $|R(k) - 1| \in L^2(-\infty,\infty)$. We shall call a solution of (B2) with the required analyticity a 1-solution if it satisfies (B1), and a 0-solution if $D+1$ satisfies (B1).

Defining

$$\eta(\alpha) = \frac{1}{2\pi}\int_{-\infty}^\infty dk\, e^{-ik\alpha}[D(k) - 1]$$
$$= \frac{1}{2\pi}\int_\infty dk\, e^{ik\alpha}[D(-k) - 1], \tag{B3}$$

we conclude that for $\alpha < 0$ we must have $\eta(\alpha) = 0$, and for $\alpha > 0$ the function $\eta(\alpha)$ must satisfy the Marchenko-like equation

$$\eta(\alpha) = g(\alpha) + \int_0^\infty d\beta\, g(\alpha+\beta)\eta(\beta), \tag{B4}$$

where

$$g(\alpha) = g^*(\alpha) = \frac{1}{2\pi}\int_{-\infty}^\infty dk\, [R(k) - 1]e^{ik\alpha}.$$

Does (B2) always have a 1-solution? If so, is it unique? Is a 1-

solution of (B2) always the one-sided inverse Fourier transform

$$D(k)=1+\int_0^\infty d\alpha\, e^{ik\alpha}\eta(\alpha) \tag{B5}$$

of the (unique?) solution of (B4)? If (B2) has no 1-solution, does (B4) have no solution?

In this case the problem can be solved by quadrature. If we define $\delta = i\frac{1}{2}\ln R$ so that $\delta\to 0$ as $k\to\infty$, then the explicit solution for $\mathrm{Im}k>0$ is well known to be

$$D(k)=\left[\prod_1^N\left(1-\frac{k_n^2}{k^2}\right)\right]\exp\left[\frac{1}{\pi}\int_{-\infty}^{\infty}dk'\,\frac{\delta(k')}{k-k'}\right], \tag{B6}$$

where the k_n are the zeros of D in $\mathbb{C}^+$, whose number N is fixed by the Levinson theorem,

$$\delta(0)=\pi N. \tag{B7}$$

N is called the *index*.

Case 1: Suppose that $N=0$. Then (B6) implies that (B2) has a unique 1-solution. Assume that (B4) has more than one solution. Then the homogeneous form of (B4) must have a nontrivial solution:

$$\zeta(\alpha)=\int_0^\infty d\beta\, g(\alpha+\beta)\zeta(\beta),\quad \alpha>0, \tag{B8}$$

and the self-adjoint operator $\mathscr{g}$ on $L^2(R_+)$ whose kernel is $g(\alpha,\beta)=g(\alpha+\beta)$ must have the eigenvalue 1. In order for (B4) nonetheless to have a solution it is necessary for ζ to be orthogonal to the inomogeneity $g(\alpha)$:

$$\int_0^\infty d\alpha\, g(\alpha)\zeta(\alpha)=0.$$

By (B8) this equation is identical to the requirement that $\zeta(0)=0$, or that its one-sided inverse Fourier transform

$$h(k)=\int_0^\infty d\alpha\, e^{ik\alpha}\zeta(\alpha) \tag{B9}$$

have the property

$$\int_{-\infty}^{\infty}dk\, h(k)=0 \tag{B10}$$

Now multiply (B8) by $\zeta(\alpha)$ and integrate:

$$\int_0^\infty d\alpha\,\zeta^2(\alpha)=\int_0^\infty d\alpha\int_0^\infty d\beta\,\zeta(\beta)g(\alpha+\beta)\zeta(\alpha).$$

In the Fourier-transform language this reads

$$\int_{-\infty}^{\infty}dk\,|h(k)|^2=\int_{-\infty}^{\infty}dk\,h^2(k)[R(k)-1]=\int_{-\infty}^{\infty}dk\,h^2R,$$

because of the analyticity of h in $\mathbb{C}^+$ and the fact that it vanishes as $|k|\to\infty$. Therefore $0=\int_{-\infty}^{\infty}dk\,h(h^*-Rh)$ $=\int dk\,(h^*-Rh)(h-R^*h^*)+[\int dk\,h(Rh-h^*)]^*$ and hence, $\int dk\,|h^*-Rh|^2=0$. Consequently, $h(-k)=R(k)h(k)$. Thus h is a 0-solution of (B2). But since for $N=0$ the 1-solution of (B2) is unique, it follows that $h=0$ and hence $\zeta=0$. Thus for $N=0$, 1 cannot be in the spectrum of the operator $\mathscr{g}$, i. e., $(\mathbb{1}-\mathscr{g})$ is invertable.

An argument similar to the one given above shows that $\mathscr{g}$ cannot have eigenvalues whose squares are larger than one. Therefore, unless $\mathscr{g}$ has the eigenvalue -1, the Neumann series for $(\mathbb{1}-\mathscr{g})^{-1}$ converges.

Case 2: Suppose now that $N>0$. Then we know that the 1-solution of (B2) is not unique. Let f_1 and f_2 be two solutions that differ by the location of one of their zeros:

$$f_1(k_0)=f_2(k_0')=0.$$

Then by (B6)

$$\frac{f_1}{f_2}=\frac{k^2-k^2}{k^2-k_0'^2}$$

and the difference

$$\Delta=f_1-f_2=f_2(k_0'^2-k_0^2)/(k^2-k_0'^2) \tag{B11}$$

is a 0-solution of (B2). Thus its Fourier transform must solve (B8), and (B10) follows from (B11) by Cauchy's theorem. Thus the existence of a nontrivial solution of (B8) still allows the existence of solutions of (B4) that lead to the 1-solutions of (B2). Thus if $N>0$ then $\mathscr{g}$ must have the eigenvalue 1, but (B4) is nevertheless solvable.

Case 3: If $N<0$ then there are no 1-solutions of (B2). All solutions of (B2) that are merorphic in $\mathbb{C}^+$ and tend to 1 as $|k|\to\infty$ must have $-N$ poles there. Suppose then that (B4) nevertheless has a solution. Then its one-sided inverse Fourier transform (B5) is such that

$$E(k)=D(-k)-R(k)D(k) \tag{B12}$$

is a 0-solution of the equation

$$E(-k)=R^*(k)E(k). \tag{B13}$$

If $E=0$ then D is a 1-solution of (B2), which cannot exist; hence E cannot vanish identically. Its Fourier transform must therefore be a nontrivial solution of the integral equation

$$\zeta(\alpha)=-\int_0^\infty d\beta\, g(-\alpha-\beta)\zeta(\beta) \tag{B14}$$

for $\alpha>0$. Thus the operator $\mathscr{g}'$ on $L^2(\mathbb{R}_+)$ whose kernel is $\mathscr{g}'(\alpha,\beta)=g(-\alpha-\beta)$ must have the eigenvalue -1.

We also note that if $N<0$ then the index of R^* is positive. Hence the problem

$$1/D(-k)=R^*(k)[1/D(k)] \tag{B15}$$

has a nonunique 1-solution and the operator $\mathscr{g}'$ must have the eigenvalue 1. We conclude that if $N<0$ then $\mathscr{g}'$ must have both 1 and -1 as eigenvalues.

We may use the same argument as for case 3, for case 2, where the index of R^* is negative. Thus when $N>0$, $\mathscr{g}$ must have -1 as an eigenvalue as well. When $N=0$ then the index of R^* is also zero, and $(\mathbb{1}-\mathscr{g}')$ must be invertible too.

Thus when $N=0$, 1 is not in the spectra of either $\mathscr{g}$ or $\mathscr{g}'$; when $N>0$, both 1 and -1 are in the spectrum of $\mathscr{g}$; when $N<0$, both 1 and -1 are in the spectrum of $\mathscr{g}'$.

It is noteworthy that the eigenvalues ± 1 of $\mathscr{g}$ or $\mathscr{g}'$ have such a remarkable stability under what may appear to be rather general kinds of perturbations of the function $R(k)$. If $|N|$ has given large integral value then it takes a "large change" in $\delta(k)$ to remove 1 or -1 from the spectrum of $\mathscr{g}$ or $\mathscr{g}'$. We note, on the other hand, that some large changes in $\delta(k)$ are small, in reasonable norms, for $R(k)=\exp[-2i\delta(k)]$. In other words, the mappings from the operators $\mathscr{g}$ and $\mathscr{g}'$ to the phase $\delta(k)$ are not continuous, and small changes in R that remove 1 or -1 from the spectrum of $\mathscr{g}$ or $\mathscr{g}'$ appear as large changes in δ.

APPENDIX C

Here we shall prove Lemma 4.1.

It follows from (4.2) that $\Pi(k)$ may be decomposed as

$$\Pi(k) = \mathbb{1} + k^{-1}\Pi_0 + \Pi'(k), \tag{C1}$$

where Π_0 is constant, $\|\Pi_0\|_2 < \infty$, and for all $k \in \mathbb{R}$, $\|\Pi'(k)\|_2 < \infty$ as well as

$$\|\Pi'(k)\|_2 = O(k^{-2}) \tag{C2}$$

as $k \to +\infty$. Furthermore, $\|d\Pi'(k)/dk\|_2 < \infty$ and

$$\|d\Pi'(k)/dk\|_2 = 0(k^{-3}), \tag{C3}$$

so that

$$\int_{-\infty}^{\infty} dk\,(k^2+1)\|\Pi'(k)\|_2^2 < \infty, \tag{C4}$$

$$\int_{-\infty}^{\infty} dk\,(k^2+1)\|d\Pi'(k)\,/\,dk\|_2^2 < \infty. \tag{C5}$$

What is more

$$\int_{-\infty}^{\infty} dk\,\|\Pi(k) - \mathbb{1}\|_2^2 < \infty \tag{C6}$$

and

$$\int_{-\infty}^{\infty} dk\,\|d\Pi(k)/dk\|_2^2 < \infty. \tag{C7}$$

Since $\mathscr{S}_x = \mathbb{1} + (ik/2\pi)A_x$, we may decompose

$$\mathscr{S}_x^{\mathrm{red}} - \mathbb{1} = (ik/2\pi)A_x + D + E + H, \tag{C8}$$

where[11]

$$D(k) = Q\tilde{\Pi}(-k)Q\Pi(-k) - \mathbb{1},$$

$$E(k) = (-i/2\pi)[Q\Pi_0 QA_x(k) + A_x(k)\Pi_0]$$

$$\begin{aligned} H(k) = (i/2\pi)[&Q\tilde{\Pi}'(-k)QA_x(k)(k\mathbb{1} - \Pi_0) \\ &+ Q(k\mathbb{1} - \tilde{\Pi}_0)QA_x(k)\Pi'(-k)] \\ &+ Q\tilde{\Pi}(-k)QA_x(k)\Pi'(-k). \end{aligned}$$

By (C6) and (C7)

$$\int_{-\infty}^{\infty} dk\,\|D(k)\|_2^2 < \infty, \tag{C9}$$

$$\int_{-\infty}^{\infty} dk\,\|dD(k)/dk\|_2^2 < \infty. \tag{C10}$$

One easily finds that

$$\int_{-\infty}^{\infty} dk\,\|A_x(k)\|_2^2 < \infty, \tag{C11}$$

as well as

$$\int dk\,\|dA_x(k)/dk\|_2^2 < \infty, \tag{C12}$$

and therefore

$$\int_{-\infty}^{\infty} dk\,\|E(k)\|_2^2 < \infty, \quad \int_{-\infty}^{\infty} dk\,\|dE(k)/dk\|_2^2 < \infty. \tag{C13}$$

Similarly, it follows from (C4), (C5), (C11), and (C12) that

$$\int_{-\infty}^{\infty} dk\,\|H(k)\|_2^2 < \infty, \quad \int_{-\infty}^{\infty} dk\,\|dH(k)/dk\|_2^2 < \infty. \tag{C14}$$

As a result, if we write $F = D + E + H$,

$$\int_{-\infty}^{\infty} dk\,\|F(k)\|_2^2 < \infty, \quad \int_{-\infty}^{\infty} dk\,\|dF(k)/dk\|_2^2 < \infty. \tag{C15}$$

Therefore, if $\mathscr{F}$ is the operator on $L^2(\mathbb{R}_+ \times S^2)$ whose kernel is $\mathscr{F}(\alpha,\theta,\beta,\theta') = (\frac{1}{2}\pi)\int_{-\infty}^{\infty} dk\, e^{ik(\alpha+\beta)}F(k,\theta,-\theta') = \hat{F}(\alpha+\beta,\theta,-\theta')$, then $\|\mathscr{F}\|_2^2 = \int_{\infty}^{0} d\alpha\,\alpha\|\hat{F}(\alpha)\|_2^2 \leqslant C < \infty$ by (C15); furthermore, C is independent of x. Together with Lemma 2.1 this proves Lemma 4.1.

[1] R. G. Newton, J. Math Phys. **21**, 1698 (1980). This paper will be referred to as II.

[2] Another solution has been given by L. D. Faddeev, Itogi Nauk. Tekh. Sov. Probl. Mat. **3**, 93 (1974) [J. Sov. Math. **5**, 334 (1976)].

[3] We remind the reader that x is a point in $\mathbb{R}^3$; θ is a point on S^2 (or a unit vector in $\mathbb{R}^3$); $A(k;\theta,\theta')$ is the scattering amplitude, which we regard as the kernel of an operator family $A(k)$ on $L^2(S^2)$. We use † for the Hermitian adjoint and * for the complex conjugate.

[4] $\|\cdot\|_2$ denotes the Hilbert-Schmidt norm, $\|A\|_2^2 = \mathrm{tr}\,A^\dagger A$, where tr denotes the trace on $L^2(\mathbb{R}_+ \times S^2)$.

[5] These are (3.4) and (3.6) of II.

[6] It also follows that if $\nu^2 = 1$ then $Q\mathscr{S}_x(k)p(k) = \nu p(-k)$; this leads to Lemma 6.1 of II.

[7] If E_x is the operator on $L^2(S^2)$ of multiplication by $\exp(-\kappa_m x\cdot\theta)$ and B^0 is the projection on $\mathscr{H}_1^0$, then $E_{-x}B^0E_x$ is a projection on $\mathscr{H}_1^x$ but it is not self-adjoint.

[8] D. Ludwig, Commun. Pure Appl. Math. **19**, 49 (1966), Theorem 4.9, p. 60.

[9] $\mathbb{C}^+$ is the upper half of the complex plane.

[10] See, for example, R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966), p. 348.

[11] The tilde denotes the operator whose kernel is the transpose.

A new direct proof of the expansion of a Coulomb-distorted plane wave in Coulomb-distorted spherical waves

Masao Furukawa
National Space Development Agency of Japan, Tokyo, 105, Japan

(Received 20 February 1981; accepted for publication 22 May 1981)

By employing specific properties of confluent hypergeometric functions, it has been directly proved that a Coulomb-distorted plane wave is expressible in superposition of Coulomb-distorted spherical waves.

PACS numbers: 03.65.Nk, 24.10.Fr, 02.30. − f

It is well known that a Coulomb-distorted plane wave $e^{ikz}F(-i\eta,1;ik(r-z))$ can be expanded in Coulomb-distorted spherical waves $\chi_l(\eta,kr)P_l(\cos\theta)$ $(l=0,1,2,\ldots)$ in the following form[1,2]:

$$e^{ikz}F(-i\eta,1;ik(r-z)) = \sum_{l=0}^{\infty}(2l+1)i^l\chi_l(\eta,kr)P_l(\cos\theta). \quad (1)$$

Here r and θ denote the spherical polar coordinates with $z=r\cos\theta$, k the wavenumber, η the Coulomb parameter, and $P_l(\cos\theta)$ the Legendre polynomial of the l th order, while $\chi_l(\eta,kr)$ represents a Coulomb function of the first kind defined by

$$\chi_l(\eta,kr) = \frac{\Gamma(l+1+i\eta)}{\Gamma(1+i\eta)}\frac{(2kr)^l e^{ikr}}{(2l+1)!} \times F(l+1+i\eta,2l+2;-2ikr). \quad (2)$$

An expression equivalent to Eq. (1) was first found by Gordon,[3] Mott,[4] and Temple[5] in the process of obtaining the Rutherford formula in wave mechanics. Their approach is, therefore, not so straightforward from the angle of demonstrating Eq. (1), and seems rather complicated. In this paper, a more direct proof of Eq. (1) will be presented by using the following properties of confluent hypergeometric functions:

$$F(\alpha,\gamma;z) = 1 + \sum_{m=1}^{\infty}\frac{\alpha(\alpha+1)\cdots(\alpha+m-1)}{\gamma(\gamma+1)\cdots(\gamma+m-1)}\frac{z^m}{m!}, \quad (3)$$

$$F(\alpha,\gamma;z) = e^z F(\gamma-\alpha,\gamma;-z), \quad (4)$$

$$\frac{d^l}{dz^l}F(\alpha,\gamma;z) = \frac{\alpha(\alpha+1)\cdots(\alpha+l-1)}{\gamma(\gamma+1)\cdots(\gamma+l-1)}F(\alpha+l,\gamma+l;z). \quad (5)$$

In order to demonstrate Eq. (1), it is convenient to employ new variables defined as $\rho=kr$ and $t=\cos\theta$. The problem is thereby attributed to determination of a coefficient, $a_l(\rho)$, in the following series:

$$e^{i\rho t}F(-i\eta,1,i\rho(1-t)) = \sum_{l=0}^{\infty}a_l(\rho)P_l(t). \quad (6)$$

Multiplying $P_l(t)$ on both sides of Eq. (6) and integrating over a range of -1 to 1, one obtains a basic expression to be calculated,

$$a_l(\rho) = \frac{2l+1}{2}\int_{-1}^{1}e^{i\rho t}F(-i\eta,1;i\rho(1-t))P_l(t)\,dt. \quad (7)$$

Applying Eq. (4), one has

$$\begin{aligned} a_l(\rho) &= \frac{2l+1}{2}\int_{-1}^{1}e^{i\rho}F(1+i\eta,1;i\rho(t-1))P_l(t)\,dt \\ &= \frac{2l+1}{2^{l+1}l!}e^{i\rho}\int_{-1}^{1}\left[\frac{d^l}{dt^l}F(1+i\eta,1;i\rho(t-1))\right] \\ &\quad\times(1-t^2)^l\,dt. \end{aligned} \quad (8)$$

Because of Eq. (5). Eq. (8) is transformed into

$$\begin{aligned} a_l(\rho) &= \frac{2l+1}{2^{l+1}l!}e^{i\rho}(i\rho)^l\frac{(1+i\eta)(2+i\eta)\cdots(l+i\eta)}{l!}\int_{-1}^{1}F(l+1+i\eta,l+1;i\rho(t-1))(1-t^2)^l\,dt \\ &= \frac{2l+1}{2^{l+1}(l!)^2}\frac{\Gamma(l+1+i\eta)}{\Gamma(1+i\eta)}e^{i\rho}(i\rho)^l\int_{-1}^{1}F(l+1+i\eta,l+1;i\rho(t-1))(1-t^2)^l\,dt. \end{aligned} \quad (9)$$

By using Eq. (3), the integration in Eq. (9) is carried out as follows:

$$\begin{aligned} &\int_{-1}^{1}F(l+1+i\eta,l+1;i\rho(t-1))(1-t^2)^l\,dt \\ &= \int_{-1}^{1}(1-t^2)^l dt + \sum_{m=1}^{\infty}\frac{(l+1+i\eta)(l+2+i\eta)\cdots(l+m+i\eta)}{(l+1)(l+2)\cdots(l+m)}\frac{(-i\rho)^m}{m!}\int_{-1}^{1}(1-t)^m(1-t^2)^l\,dt \\ &= 2^{2l+1}\frac{l!l!}{(2l+1)!} + \sum_{m=1}^{\infty}\frac{(l+1+i\eta)(l+2+i\eta)\cdots(l+m+i\eta)}{(l+1)(l+2)\cdots(l+m)}\frac{(-i\rho)^m}{m!}2^{2l+m+1}\frac{(l+m)!l!}{(2l+m+1)!} \\ &= 2^{2l+1}\frac{(l!)^2}{(2l+1)!}\left[1+\sum_{m=1}^{\infty}\frac{(l+1+i\eta)(l+2+i\eta)\cdots(l+m+i\eta)}{(2l+2)(2l+3)\cdots(2l+m+1)}\frac{(-2i\rho)^m}{m!}\right] \\ &= \frac{2^{2l+1}(l!)^2}{(2l+1)!}F(l+1+i\eta,2l+2;-2i\rho). \end{aligned} \quad (10)$$

Hence Eq. (9) is reduced to

$$a_l(\rho) = (2l+1)i^l \frac{\Gamma(l+1+i\eta)}{\Gamma(1+i\eta)} \frac{(2\rho)^l e^{i\rho}}{(2l+1)!} F(l+1+i\eta, 2l+2; -2i\rho) = (2l+1)i^l \chi_l(\eta,\rho). \tag{11}$$

Substitution of Eq. (11) into Eq. (6) yields

$$e^{i\rho t} F(-i\eta, 1; i\rho(1-t)) = \sum_{l=0}^{\infty} (2l+1) i^l \chi_l(\eta,\rho) P_l(t). \tag{12}$$

Thus Eq. (1) has been proved directly.

[1] K. Adler *et al.*, "Study of Nuclear Structure by Electromagnetic Excitation with Accelerated Ions," Rev. Mod. Phys. **28**(4), 432–542 (1956).
[2] A. Sommerfeld, "Über die Beugung und Bremsung der Elektronen," Ann. Physik (Leipzig) **11**, 257–330 (1931).
[3] W. Gordon, "Über den Stosszweier Punktladungen nach der Wellenmechanick," Z. Phys. **48**, 180–191 (1928).
[4] N. F. Mott, "The Solution of the Wave Equation for the Scattering of Particles by a Coulombian Center of Force," Proc. R. Soc. London **118**, 542–9 (1928).
[5] G. Temple, "The Scattering Power of a Bare Nucleus According to Wave Mechanics," Proc. R. Soc. London, **121**, 673–5 (1928).

Algebraic and geometric structure of linear filters and scattering systems. I[a)]

Robert Hermann
The Association for Physical and Systems Mathematics, 53 Jordan Road, Brookline, Massachusetts 02146

(Received 26 February 1981; accepted for publication 29 May 1981)

Some of the problems of linear filter and scattering theory are interpreted in terms of the theory of algebraic and analytical subvarieties of Grassmann manifolds.

PACS numbers: 03.65.Nk, 84.30.Vn, 02.40. – k

I. INTRODUCTION

In recent years, System Theory has arisen as a mathematically oriented discipline, attempting to provide unified methodology for diverse branches of engineering and physics, and perhaps playing the role in technology that mathematical physics has historically played in physics and chemistry. Mathematics has a strong role to play here, as, of course, it has in mathematical physics. Existing mathematics must be adapted to the problems, and information about the applications must be fed back to stimulate the development of new mathematics. As in physics, geometry (differential and algebraic) and Lie theory has played an increasing role in System Theory in recent years.

In Ref. 1, Martin and I were led to the introduction of certain concepts of algebraic geometry and complex manifold theory for the study of the finite dimensional linear, time-invariant, input-output systems. This work has been carried considerably further in the work of Byrnes and Duncan.[2] The purpose of this paper is to extend the study of the relations with algebraic geometry to other systems, particularly those of *transmission line theory*[3] and related "distributed parameter systems." What might be of interest to the wider mathematical physics community is that scattering problems can be treated with a related geometric formalism.

II. THE FREQUENCY VARIETY OF FINITE-DIMENSIONAL, LINEAR, TIME-INVARIANT INPUT-OUTPUT SYSTEMS

Let us briefly recall the situation of Ref. 1. Let U, Y, X be finite-dimensional vector spaces, over the complex numbers as field of scalars. Let

$$L(U,Y)$$

be the vector space of linear maps: $U \to Y$. Suppose we are given a triple

$$(A,B,C),$$

with

$$A \in L(X,X), \quad B \in L(U,X), \quad C \in L(X,Y).$$

Construct the input–output system

$$\frac{dx}{dt} = Ax + Bu, \quad y = Cx. \tag{2.1}$$

The *frequency response* determines the solution of Eqs. (2.1), which are eigenvalues of the operator d/dt. Algebraically, it is determined by

$$T(s) = C(s - A)^{-1}B. \tag{2.2}$$

a rational map of $\mathbb{C} \to L(U,Y)$ (in classical matrix language, a matrix of rational functions of the complex variable s.)

In Ref. 1, an algebraic variety was assigned to the system (2.1) and the frequency response function (2.2) in the following way: Let

$$W = U \oplus Y \tag{2.3}$$

be the direct sum of the input and output spaces,

$$n = \dim U \tag{2.4}$$

Let $G^n(W)$ be the Grassmann manifold[4] of n-dimensional linear subspace of the vector space W.

Define a mapping

$$\gamma: \mathbb{C} \to G^n(W) \tag{2.5}$$

of the complex numbers (parameterized by the complex variable s) into $G^n(W)$ as follows:

$$\gamma(s) = \{(u, T(s)u) : u \in U\}. \tag{2.6}$$

$\gamma(s)$ is essentially the *graph* of $T(s)$. It is shown in Ref. 1 that γ can also be defined at the poles of T, i.e., the eigenvalues of A, so as to be an analytic map of $\mathbb{C} \to G^n(W)$. Further, γ can be extended holomorphically to $s = \infty$, to obtain a holomorphic mapping of the extended complex plane [which we denote as $P_1(\mathbb{C})$, since it is isomorphic to the complex projective space of one-dimensional linear subspaces of $\mathbb{C}^2$] into $G^n(W)$. The image under γ is a *rational algebraic curve* in $G^n(W)$. In this paper it will be called the *frequency variety* of the system. The natural algebrogeometric invariants of the curve then turn out to be the natural system theoretic invariants of the system (2.1).

III. THE FREQUENCY VARIETY OF LINEAR, TIME-INVARIANT FILTERS

Let U and Y continue as finite-dimensional complex vector spaces. A (linear, time-invariant filter) is an integral operator transforming time-parameterized curves in U into curves of Y, of the following form:

$$y(t) = \int_0^t K(t-\tau)u(\tau)\, d\tau \tag{3.1}$$

Here, $t \to K(t)$ is a curve in $L(U,Y)$.

[a)]Supported by a grant from the Ames Research Center (NASA), #NSG-2402, the U.S. Army Research Office, #ILIG1102RHN7-05 MATH, and the National Science Foundation.

The *symbol* of the filter is the Laplace transform

$$\sigma(s) = \int_0^\infty e^{-st} K(t)\, dt \tag{3.2}$$

of the kernel of the integral operator. For the filters commonly encountered in engineering, this symbol will be initially defined by (3.3) as a holomorphic function in a region S of the complex s plane, and then may be analytically continued. Let us adopt the Grassmannian approach of Ref. 1 to the situation.

Again, let

$$W = U \oplus Y$$

$G^n(W)$ = Grassmann manifold of n-dimensional linear subspaces of W. Set

$$\gamma(s) = \{(u, \sigma(s)u : u \in U\}. \tag{3.3}$$

$s \to \gamma(s)$ then is a holomorphic mapping of S into $G^n(W)$.

Definition: The algebraic closure of the set $\gamma(S)$ in $G^n(W)$, i.e., the smallest algebraic subvariety[4,5] of $G^n(W)$ containing $\gamma(S)$, is called the *frequency variety* of the filter.

Remark: It is a basic property of algebraic varieties[4,5] that the set of algebraic subvarieties is closed under intersection (equivalent to the Hilbert basis theorem) which assures us that the "frequency variety" is a well-defined concept.

One can now refer to a table of Laplace transforms to have an idea of the possible diversity of frequency varieties.

A more general possibility is to define the "frequency variety" as the *analytic closure* of $\gamma(S)$, i. e., as the smallest analytic subvariety[5] of $G^n(W)$ containing $\gamma(S)$.

IV. ALGEBRAIC AND ANALYTIC VARIETIES ASSOCIATED WITH ONE-DIMENSIONAL WAVE EQUATIONS AND BOUNDARY VALUES

Let W be a finite dimensional complex vector space and let x and t be two real variables (physically, space and time). Consider a linear partial differential equation of the following form:

$$w_t = A(x)w_x + B(x)w, \qquad (x,t) \to A(x),\ B(x) \tag{4.1}$$

are assumed to be (for simplicity) real analytic maps:

$$R \to L(W,W).$$

Solutions $(x,t) \to w(x,t)$ to (4.1) are to be real analytic maps: $R^2 \to W$. Subscripts denote partial derivatives:

$$w^x \equiv \frac{\partial w}{\partial x} \equiv \partial_x w, \quad w_t \equiv \frac{\partial w}{\partial t} \equiv \partial_t w. \tag{4.2}$$

The system of differential equations (4.2) is time-translation invariant, i.e., if

$$(x,t) \to w(x,t)$$

is a solution, so is $\partial_t w$. Let us look for eigenvalues of ∂_t on the space of solutions with eigenvalues s. They are solutions of the following system of ordinary differential equations, with s as parameter:

$$w(x,t) = e^{ts}\tilde{w}(x,s), \tag{4.3}$$

$$s\tilde{w} = A\tilde{w}_x + B\tilde{w}. \tag{4.4}$$

For fixed s, (4.4) will be a system of linear, ordinary differential equations. Let us assume (for simplicity) that the space of solutions (analytic in x) forms an n-dimensional complex vector space, denoted as

$$\mathscr{S}(s).$$

Let Y be another complex vector space with

$$C: W \to Y$$

a linear map. Let $G(Y)$ be the Grassmann *space* of Y consisting of *all* linear subspaces of Y. $G(Y)$ is acted on by $GL(Y)$. The orbits are, of course, the Grassmann *manifolds* $G^n(Y)$ of subspaces of fixed dimension.

For two values of x, say

$$x = a, \quad x = b, \tag{4.5}$$

let $\gamma(s)$ be the following linear subspace of $Y \oplus Y$

$$\gamma(s) = (C(\tilde{w}(a)), C(\tilde{w}(b))) : \tilde{w} \in \mathscr{S}(s). \tag{4.6}$$

This determines a map γ from certain regions S of the complex plane to $G(Y \oplus Y)$. We will now define the *analytic frequency variety* of the boundary value system as the smallest analytic subvariety[5] of $G(Y \oplus Y)$ containing $\gamma(S)$. Of course, this is not, at this stage, completely meaningful as a definition, since it is not clear that the definition of "analytic variety" can be extended from the Grassmann manifolds to these more general spaces. The study of this point will be put off until more examples are available as data.

V. AN ILLUSTRATIVE LINEAR WAVE SYSTEM

To provide a class of examples, let us specialize the system (4.1) to one of the following form:

$$w_t = Aw_x, \tag{5.1}$$

where A is a linear map: $W \to W$ such that

$$A^2 = 1. \tag{5.2}$$

[(5.1) is thus a constant coefficient wave equation.] Thus, the equation determining $\mathscr{S}(s)$ is the following:

$$w_x = sAw \tag{5.3}$$

or

$$w(x,s) = \exp(sxA)w_0, \tag{5.4}$$

where w is an element of W.

Now, in view of our assumptions (5.2) about the algebraic property of A, we can calculate $\exp(sxA)$ explicitly:

$$\exp(sxA) = \cosh(sx) + \sinh(sx)A \tag{5.5}$$

Hence,

$$\mathscr{S}(s) = \text{set of curves in } W \text{ of the form} \quad x \to \cosh(sx) + \sinh(sx)Aw, \tag{5.6}$$

where w runs through W.

Thus,

$$\begin{aligned} \gamma(s) &= \{(w(a), w(b)) : s \to w(s) \text{ in } \mathscr{S}(s)\} \\ &= \{(\cosh(as)w_0 + \sinh(as)Aw_0, \\ &\qquad \cosh(bs)w_0 + \sinh(bs)Aw_0)\}, \end{aligned} \tag{5.7}$$

where w_0 runs through W. $\gamma(s)$ is thus an n-dimensional linear subspace of $W \oplus W$. The algebraic and analytical closure of $\gamma(S)$ is readily calculated, using the algebraic relations satisfied by the cosh and sinh functions.

In particular, note what happens if

$$b>0, \quad a=-b:$$

$$\begin{aligned}(w(-b),w(b)) &= (\cosh(bs)w_0 - \sinh(bs)Aw_0, \cosh(bs)w_0 \\ &\quad + \sinh(bs)Aw_0) \\ &= \cosh(bs)(w_0 - \tanh(bs)aw_0, w_0 + \tanh(bs)Aw_0) \\ &= \cosh(bs)(w_0,[1+\tanh(bs)A]/[1-\tanh(bs)A] - w_0).\end{aligned} \tag{5.8}$$

This proves:

Theorem 5.1: $\gamma(s)$ is the graph of the linear map:

$$[1+\tanh(bs)A]/[1-\tanh(bs)A]: W \to W. \tag{5.9}$$

We can now establish a relation to *scattering theory*: Set

$$s = ik, \quad k \in R. \tag{5.10}$$

Set

$$s(k) = [1+\tanh(ikb)A]/[1-\tanh(ikb)A].$$

$S(k)$ is the *scattering operator* of the system. If $W = \mathbb{C}^2$, with A a real, symmetric matrix, note that $S(k)$ is *unitary*. Note that it is a Cayley transform of a real symmetric matrix.

This parameterization of $\gamma(ik)$ as the graph of a *unitary* matrix is the "scattering picture" in classical circuit theory (see Ref. 3, Chap. 3). This suggests a geometric description of "scattering theory" for more general nonconstant coefficient systems.

VI. SCATTERING THEORY

Return to the general, nonconstant coefficient one-dimensional transmission line-wave equation:

$$w_t = A(x)w_x + B(x)w. \tag{6.1}$$

For $s \in \mathscr{S}$, let $\mathscr{S}(s)$ be the space of solutions of the equation:

$$s\tilde{w} = A(x)\tilde{w}_x + B(x)\tilde{w}. \tag{6.2}$$

Let Y be another vector space with

$$C: W \to Y$$

a linear map. Let b be a positive real number, and consider x over the interval

$$-b \leqslant x \leqslant b.$$

Let

$$s \to \gamma(s,b) \in G(Y \oplus Y)$$

be the boundary value map constructed in Sec. IV, i.e., $\gamma(s)$ is the space of vectors of the form

$$(C\tilde{w}(-b), C\tilde{w}(-a)), \tag{6.3}$$

where w runs over $\mathscr{S}(s)$, i.e., the solutions of (6.2).

Now, we are prepared to define the *scattering curve*

$$s \to \gamma(s) \in G(Y \oplus Y)$$

(heuristically) as

$$\gamma(s) = \lim_{b \to \infty} \gamma(s,b). \tag{6.4}$$

γ is called the *scattering curve*, and its algebraic or analytic closure is the *scattering variety*.

Of course, in the physics literature one finds scattering defined in terms of *matrices*.

Let

$$s \to \gamma(s,b).$$

Now, various subsets of the Grassmann spaces are *parametrized* by matrices.

For example, one might have a linear map

$$S(b,s): Y \to Y$$

such that

$$\gamma(s,b) = \{(y, S(b,s)y): y \in Y\}$$

i.e., $\gamma(s,b)$ is the *graph* of y. One might try to define "scattering matrix" as

$$S(s) = \lim_{b \to \infty} S(b,s). \tag{6.5}$$

This limit is a curve in the Grassmannian $G^n(W \oplus W)$. The *scattering operator* would, of course, then be defined as a curve

$$s \to S(s).$$

in $L(W,W)$, such that

$$\gamma(s,\infty) = \text{graph of } S.$$

Of course, discussing the rigorous conditions for the existence of these limits requires analytic technique, some of which can be developed with known methods, which will be considered in a later paper in this series.

We now turn to the relation between the formalism and the usual description of 1-*D* scattering for the inhomogeneous wave equation.

VII. SCATTERING THEORY FOR THE 1-D WAVE EQUATIONS[6]

Specialize now to the case

$$W = \mathbb{C}^2.$$

Let us put the following wave equation,

$$y_{tt} = y_{xx} - V(x)y, \tag{7.1}$$

into "system" form. Factor the wave operator:

$$\partial_{tt} - \partial_{xx} = (\partial_t + \partial_x)(\partial_t - \partial_x). \tag{7.2}$$

Then, (7.1) takes the form

$$(y_t - y_x)_t + (y_t - y_x)_x = Vy. \tag{7.3}$$

Put

$$z = y_t - y_x. \tag{7.4}$$

Then,

$$z_t + z_x = Vy. \tag{7.5}$$

Now, write (7.4)–(7.5) as an "evolution" system:

$$y_t = y_x + z, \quad z_t = Vy - z_x. \tag{7.6}$$

In matrix form,

$$w_t \equiv \begin{pmatrix} y \\ z \end{pmatrix}_t = \begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}_x + \begin{pmatrix} 0, & -1 \\ V, & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}. \tag{7.7}$$

Look for solutions of (7.7) that are eigenvectors of $\partial/\partial t$

with eigenvalue ik, with k real. Thus,

$$\begin{pmatrix} y \\ z \end{pmatrix} = e^{ik}\begin{pmatrix} \tilde{y}(x,k) \\ \tilde{z}(x,k) \end{pmatrix}$$

$$ik\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix}\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}_x + \begin{pmatrix} 0, & 1 \\ -V(x), & 0 \end{pmatrix}\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}, \quad (7.8)$$

where $\tilde{y}$, $\tilde{z}$ are functions of x and k.

Now, set up boundary conditions at $x = \pm a$. Let $\gamma_a(k)$ be the two-dimensional linear subspace of $\mathbb{C}^2 \oplus \mathbb{C}^2$ generated by the boundary values of the solutions $\mathscr{S}(k)$ of (7.8) at $x = \pm a$:

$$\gamma_a(k) = \left\{\begin{pmatrix} \tilde{y}(-a,k) \\ \tilde{z}(-a,k) \end{pmatrix} \oplus \begin{pmatrix} \tilde{y}(a,k) \\ \tilde{z}(a,k) \end{pmatrix} : y \in \mathscr{S}(k)\right\}. \quad (7.9)$$

Rewrite (7.8) as follows:

$$A\frac{\partial}{\partial x}\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} ik, & 1 \\ -V(x), & ik \end{pmatrix}\begin{pmatrix} \tilde{y} \\ \tilde{k} \end{pmatrix},$$

$$A = \begin{pmatrix} 1, & 0 \\ 0, & -1 \end{pmatrix}. \quad (7.10)$$

Note that

$$A = A^{-1}, \quad A\begin{pmatrix} ik, & 1 \\ -V, & ik \end{pmatrix} = \begin{pmatrix} ik, & -1 \\ V, & -ik \end{pmatrix}. \quad (7.11)$$

Hence, (7.9) takes the form

$$\frac{\partial}{\partial x}\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} ik, & 1 \\ -V, & -ik \end{pmatrix}\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}. \quad (7.12)$$

Let $(x,k,a) \rightarrow g(x,k,a)$ be the map from R^3 to the 2×2 matrix group $\mathrm{GL}(2,C)$ such that

$$\frac{\partial g}{\partial x} = \begin{pmatrix} ik, & 1 \\ -V(x), & -ik \end{pmatrix}g, \quad (7.13)$$

$$g(-a,k,a) = 1. \quad (7.14)$$

Since

$$\mathrm{tr}\begin{pmatrix} ik, & 1 \\ -V, & -ik \end{pmatrix} = 0,$$

g is a map from R^3 to $\mathrm{SL}(2,C)$, the group of determinant one real 2×2 matrices. If $V = 0$, notice that g lies in a compact subgroup of $\mathrm{SL}(2,C)$. Thus, if

$$x \rightarrow \begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}(x,k)$$

belongs to $\mathscr{S}(k)$, then

$$\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}(x,k) = g(x,k,a)\begin{pmatrix} \tilde{y} \\ \tilde{z} \end{pmatrix}(-a,k). \quad (7.15)$$

Thus, we see that we have proved

Theorem 7.1: Let $G^2(\mathbb{C}^4)$ be the Grassmann manifold of two-dimensional linear subspace of $\mathbb{C}^4$. Then, (ik,a) (as defined in Sec. VI) is the linear subspace

$$\{(w, g(a,k,a)w) : w \in \mathbb{C}^2\}. \quad (7.16)$$

The scattering matrix $S(k)$ is then defined as follows:

$$\lim_{a\to\infty} g(a,k,a) = S(k). \quad (7.17)$$

We can, of course, write the usual Volterra integral equation for the function

$$g(x,k,\infty) = \lim_{a\to\infty} g(x,k,a):$$

$$\partial_x(g) = \begin{pmatrix} ik, & 1 \\ 0, & -ik \end{pmatrix}g + \begin{pmatrix} 0, & 0 \\ -V, & 0 \end{pmatrix}g. \quad (7.18)$$

Hence

$$g(x,k,a) = 1 + \int_{-a}^{x} \exp\left((x-u)\begin{pmatrix} ik, & 1 \\ 0, & -ik \end{pmatrix}\right) - \begin{pmatrix} 0, & 0 \\ -V(u), & 0 \end{pmatrix} g(u,k,a)\, du, \quad (7.19)$$

$$g(x,k,\infty) = 1 + \int_{-\infty}^{x} \exp\left((x-u)\begin{pmatrix} ik, & 1 \\ 0, & -ik \end{pmatrix}\right) - \begin{pmatrix} 0, & 0 \\ -V(u), & 0 \end{pmatrix} g(u,k,\infty)\, du. \quad (7.20)$$

From this, the scattering matrix itself may be obtained from (7.17).

Remark: This general 2-vector formalism is not necessarily the optimal one for this example. One can follow Fadeev's formalism[8] better by introducing an "output" space

$$Y = \mathbb{C}$$

with

$$C: W \rightarrow Y$$

given by

$$C\begin{pmatrix} y \\ z \end{pmatrix} = y.$$

Of course, the case

$$\mathbb{C}^2 = W = Y, \quad C = \text{identity},$$

is essentially that described exhaustively by Ablowitz, Kaup, Newell, and Segur.[7]

VIII. FINAL REMARKS

The preceding example, when viewed in terms of this geometric-systems formulation, suggests many directions of generalization. Note the role played by the orbit structure and its closure of the action of various linear groups on Grassmannians. For example, in Sec. VII we dealt with $G^2(\mathbb{C}^4)$, with the usual *transitive* action of $\mathrm{GL}(4,C)$. $\mathrm{GL}(2,C)\times\mathrm{GL}(2,C)$ is a subgroup. A subgroup of this is the group $G = 1\times\mathrm{SL}(2,C)$. Basically, we are dealing with the orbit of this group and its closure. Of course, the usual "unitarity" of the S-matrix[7] has something to do with the action of the $1\times\mathrm{SU}(2,\mathbb{C})$ subgroup. A Lie-group theorist will see obvious possibilities of extending the formalism to more complicated systems and Lie groups.

Another source of fruitful speculation might be analogies between "inverse scattering" and the "identification-filtering" problem of system theory.[9]

[1]R. Hermann and C. Martin, SIAM J. Control Optim. **16**, 743–755 (1978).

[2]C. Byrnes and T. Duncan, "On certain topological invariants in system theory," in *New Directions in Applied Mathematics*, edited by P. Hilton and G. Young (Springer-Vergag, Berlin) to be published.

[3]R. Newcomb, *Linear Multiport Synthesis* (McGraw-Hill, New York, 1966).

1966).
[4]P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1976).
[5]H. Whitney, *Complex Analytic Varieties* (Addison-Wesley, Reading, Mass., 1972).
[6]I. Kay, Comm. Pure. Appl. Math. **13**, 371–393 (1960).
[7]M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249–315 (1974).
[8]L. D. Fadeev, "Properties of the *S*-matrix of the one-dimensional Schrödinger equations," Am. Math. Soc. Transl. Ser. 2 **65**, 139–166 (1967).
[9]M. Hazewinkel and J. Willems, in *Proceedings of the* 1980 *Les Arcs NATO Conference on Filtering* (Reidel, Dordrecht), to be published.

A perturbative look at the dynamics of extended systems in quantum field theory

G. Semenoff, H. Matsumoto, and H. Umezawa
Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2J1, Canada

(Received 7 January 1981; accepted for publication 27 March 1981)

The structure of a quantum field theory with an extended object is explored perturbatively. The perturbative expansion consists of coupled differential equations which are derived from the Heisenberg field equation. These equations are used to reduce the zero mode problem to a choice of boundary conditions. They are then integrated and the constraint of the equal time commutation relations is used to set the boundary conditions and derive commutation relations for the physical fields and the quantum coordinate. Using this information the quantal Hilbert space is constructed.

PACS numbers: 03.70. + k, 11.10. − z, 11.30. − j

I. INTRODUCTION

In recent years much effort has been devoted to extending the methodology of quantum field theory to the analysis of physical systems which have spatially inhomogeneous ground states. Quantum systems of this kind are said to contain extended objects.

Some motivation for the study of such systems is derived from the fact that many nonlinear field equations possess classical soliton solutions.[1] These are, for either dynamical or topological reasons, classically stable, nontrivial configurations of the basic fields. It is of considerable interest to study the role of these classical solutions in the corresponding quantum field theory.[2] From the point of view of the latter theory the classical nature of these solutions is not the result of a classical limit of the quantum theory. Instead, classical objects coexist with quanta and interact with them.

Furthermore, there are many examples of these types of extended structures in nature. Point defects and dislocations in crystals and vortex lines in superconductors are extended objects which are imbedded in and coexist with quantum systems. Grain boundaries and surfaces are also good examples. Many quasi-one-dimensional systems such as long polymer chains are known to possess solitonlike behavior. The study of systems such as these requires the development of a consistent methodology for the analysis of extended objects in quantum field theory.

The analysis typically begins with some classical solution of the Heisenberg field equation of the quantum field theory and treats this solution as a first approximation to the vacuum matrix element of a basic field in an expansion in Planck's constant $\hbar$ (a loop expansion). This first approximation is known as the tree approximation. Quantum corrections are to be added perturbatively. However, naive perturbation theory immediately encounters a technical problem, that of the so-called "zero modes". The field equation, when linearized about some static classical solution, possesses zero frequency eigensolutions, the zero modes. These modes arise when the classical solution is noninvariant under some symmetry transformations which leave the original field equation invariant. Since they have zero eigenvalues, the part of the Green's function corresponding to these solutions does not exist. This makes a direct integration of the inhomogeneous differential equations of perturbation theory impossible.

Several methods have been proposed for dealing with this problem. One of them, the collective coordinate method,[3] anticipates the appearance of the zero modes by performing a canonical transformation which elevates the symmetry parameters which give rise to the zero modes to the status of dynamical variables. This is accompanied by the imposition of certain constaints on the operator fields which can be arranged in such a way that a diagrammatic type of perturbation theory is possible. This method has been used for the calculation of quantum corrections to the ground-state energy and the soliton-soliton scattering amplitudes in some model quantum field theories.[4]

A systematic method for describing extended systems has been investigated within the orthodox formalism of quantum field theory. In this method (the boson method[5]) extended objects are created by a boson condensation process. The Heisenberg equation of a quantum field theory is first solved for the case of a spatially homogeneous ground state. Then certain extended objects are created in this system by means of the boson transformation which is the mathematical expression of a boson condensation. It has been shown that, in the tree approximation, the boson transformation leads to the corresponding classical field equations and their soliton solutions.[6]

For example, consider a quantum field theory consisting of a single boson Heisenberg field satisfying the field equation

$$\Lambda(\partial)\psi(x) = F[\psi(x)], \tag{1.1}$$

which interpolates single boson in- and out-fields $\rho^{\text{in}}(x)$ and $\rho^{\text{out}}(x)$ satisfying the equation

$$\lambda(\partial)\rho^{\text{in}}(x) = \lambda(\partial)\rho^{\text{out}}(x) = 0. \tag{1.2}$$

Let us assume that both field equations exhibit space- and time-translational symmetry. A solution of this quantum field theory is given when all matrix elements of the Heisenberg field $\psi(x)$ in the Fock space of the in-fields $\rho^{\text{in}}(x)$ are given. This is expressed compactly in an expression known as the dynamical map

$$\psi(x) = \psi[x, \rho^{\text{in}}]. \tag{1.3}$$

The boson transformation theorem states that the Heisenberg field

$$\psi^f(x) = \psi[x, \rho^{\text{in}} + f] \tag{1.4}$$

satisfies the field equation

$$\Lambda(\partial)\psi^f(x) = F[\psi^f(x)], \tag{1.5}$$

when $f(x)$ is a c-number function satisfying

$$\lambda(\partial)f(x) = 0. \tag{1.6}$$

The operator translation

$$\rho^{\text{in}}(x) \to \rho^{\text{in}}(x) + f(x) \tag{1.7}$$

is called the boson transformation. It corresponds to a condensation of the bosons $\rho^{\text{in}}(x)$. The Heisenberg field $\psi^f(x)$ completely describes the quantum system with extended structure.

In general, an essential feature of static extended objects is the fact that the boson transformation function $f(x)$ must have some singularities which prohibit its Fourier transform. Computations in this formalism, then, require particular care in the treatment of these singularities. This has been demonstrated explicitly in the calculation of the soliton solutions for some $(1+1)$-dimensional field theories.[6]

The physical fields $\rho^{\text{in}}(x)$ are modified by their interaction with the extended object. In general, besides the scattering states of these asymptotic particles, there appear bound states of the particles to the extended object and also a single quantum mechanical mode associated with the translation of the system. This mode is known as the quantum coordinate. Its appearance is a natural result of the canonical commutation relations of the Heisenberg fields.[7]

The quantal Hilbert space of the system with extended structure is therefore different from the Fock space of the fields $\rho^{\text{in}}(x)$. The physical fields are, in fact, complicated functionals of the boson transformation function $f(x)$. They consist of infinite summations of the physical fields $\rho^{\text{in}}(x)$ interacting through the many-point Green's functions of the homogeneous theory with the classical fields $f(x)$. These series may not converge, but in certain spatial regions may be asymptotic. The commutation relations are useful in defining the sums in these regions.

In Ref. 8 it was shown that, once the order parameter in the tree approximation is known, it is possible to derive coupled equations for the many-particle components of the dynamical maps of the Heisenberg fields in the tree approximation. These equations take the form of a perturbative expansion. They must be solved consistently together with the canonical commutation relations. This was done in Ref. 8 for the first few orders. The physical Hilbert space for the system is the direct product of the Hilbert space of the quantum coordinate and the Fock space of the particlelike states.

However, particularly subtle effects occur in higher orders of this perturbative expansion where the recoil of the extended system under scattering by particles is taken into account. The Hamiltonian of the system must contain terms which describe this interaction and thus is not a simple linear sum of the energies of the quantum soliton and the quantum particle modes even after the "diagonalization" is completed.[9]

It is the purpose of the present work to explore this regime perturbatively. The first result is a theorem which establishes a connection, at least in the tree approximation, between the quantum coordinate and the collective coordinate used in the collective coordinate method. This theorem states that, when the tree approximation is summed to all orders, the quantum coordinate Q appears in the dynamical map in the combination $x + Q$ with the spatial coordinate x. This gives a criterion by which the coupled equations of the perturbation theory are solved. Then the dynamical map is calculated to second order in the tree approximation for a $(1+1)$-dimensional boson model. In the course of converting the perturbative equations into the form of integral equations, the zero mode associated with translational invariance presents a technical problem. The usual Green's function method is applicable only to those components which are orthogonal to the zero-mode wavefunction. The components which are proportional to zero-mode wavefunction can, however, be determined to within the addition of a solution of the homogeneous equation. Solutions of the homogeneous equation which may be added at each order of the perturbative calculation as well as the commutation relations among the linear (or physical) operators are then determined using the canonical commutation relations. This is done both indirectly through the canonical momentum and Hamiltonian operators and also by direct perturbative calculation of the commutation relations. The physical Hilbert space is then constructed as a direct product of the Schrödinger picture realization of the quantum coordinate and the canonical momentum and the Fock-space realization of the quantum particle excitations.

The number representation of the physical particles is used to construct their Fock space. This Fock space does not manifest translational invariance. Completeness requires the presence of the quantum coordinate which serves to recover translational invariance. Thus the physical Hilbert space must be a direct product of these two subspaces.

II. THE QUANTUM COORDINATE

Consider a one-component boson field in $(1+1)$-dimensions satisfying the Heisenberg field equation

$$(\partial^2 + m^2)\psi(x) = F[\psi(x)]. \tag{2.1}$$

It is assumed that this equation, together with the equal-time canonical commutation relation

$$[\psi(x), \dot\psi(y)]_{x^0 = y^0} = i\,\delta(x-y), \tag{2.2}$$

can be realized in the Fock space of a single free boson field $\rho^0(x)$ with field equation

$$(\partial^2 + m^2)\rho^0(x) = 0. \tag{2.3}$$

Thus the boson transformation $\rho^0(x) \to \rho^0(x) + f(x)$ leads to the field equation

$$(\partial^2 + m^2)\psi^f(x) = F[\psi^f(x)]. \tag{2.4}$$

The following is a study of this equation in the tree approximation.

Some of the notation we will use is the following. The symbol dx denotes spatial integration and d^2x a space-time integration. The symbol x denotes either the space coordinate or the space-time coordinates. Whether or not the time coordinate is included should be clear from the context.

The vacuum expectation value of $\psi^f(x)$ in the Fock space of $\rho^0(x)$ is called the order parameter. It is the sum of all connections through the many-point Green's functions of the Heisenberg field $\psi(x)$ with the classical functions $f(x)$. We will, for purposes which will be clear later, denote the order parameter by $\psi_{-1}(x)$. One can obtain the n-particle term in the dynamical map of $\psi^f(x)$ by removing n of the functions $f(x)$ from the order parameter and replacing them with a normal-ordered product of n of the basic fields $\rho^0(x)$

$$\psi_{n-1}(x) = :\delta_f^n \psi_{-1}(x):, \tag{2.5}$$

where

$$\delta_f = \int d^2y\ \rho_0(y)\frac{\delta}{\delta f(y)}. \tag{2.6}$$

It is useful to consider a power-counting parameter λ

$$\psi_\lambda^f(x) = \sum_{n=-1}^{\infty} \lambda^n \psi_n(x). \tag{2.7}$$

In the tree approximation the order parameter satisfies the equation

$$(\partial^2 + m^2)\psi_{-1}(x) = F[\psi_{-1}(x)]. \tag{2.8}$$

Equation (2.4) leads to

$$(\partial^2 + m^2)\psi_n(x) = \delta_f^{n+1}F[\psi_{-1}(x)]$$

or

$$(\partial^2 + m^2)\psi_n(x) = \sum \frac{1}{l!}F_l[\psi_{-1}(x)]\psi_{\alpha_1}(x)\cdots\psi_{\alpha_l}(x), \tag{2.9}$$

where $\alpha_1 + \cdots + \alpha_l + l = n+1$; $\alpha_1,\cdots,\alpha_l \geqslant 0$; $l \geqslant 0$ and

$$F_l[\psi_{-1}(x)] = \frac{\partial^l F[\psi_{-1}(x)]}{\partial \psi_{-1}(x)^l}.$$

If the powers of λ are inserted in Eq. (2.9) we see that

$$(\partial^2 + m^2)\psi_\lambda^f(x) = \lambda^{-1}F[\lambda\psi_\lambda^f(x)], \tag{2.10}$$

that is, λ is like a scaling factor. The right-hand side of Eq. (2.9) contains a term which is linear in $\psi_n(x)$. When this term is subtracted from each side of the equation, the right-hand side contains components of order strictly less than n

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\psi_n(x) = \sum \frac{1}{l!}F_l[\psi_{-1}(x)]\psi_{\alpha_1}(x)\ldots\psi_{\alpha_l}(x), \tag{2.11}$$

where $\alpha_1 + \cdots + \alpha_l + l = n+1$; $\alpha_1,\cdots,\alpha_l \geqslant 0$; $l \geqslant 2$; $n \geqslant 0$. Equation (2.11) is supplemented by the classical field equation (2.8). We assume that $\psi_{-1}(x)$ behaves as $x \to \pm\infty$ in such a way that $F_1[\psi_{-1}(x)] \to 0$ and $\psi'_{-1}(x) \to 0$ faster than any polynomial. We also assume that $\psi_{-1}(x)$ is static. The quantum field $\psi_0(x)$ satisfies the linear field equation

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\psi_0(x) = 0, \tag{2.12}$$

possessing the well-known solution

$$\psi_0(x) = Q\psi'_{-1}(x) + \tilde{\psi}_0(x), \tag{2.13}$$

where the prime denotes differentiation by the spatial coordinate. The operator Q is the quantum coordinate and $\tilde{\psi}_0(x)$ is written as

$$\tilde{\psi}_0(x) = \sum_i \frac{1}{(2\omega_i)^{1/2}}\{u_i(x)e^{-i\omega_i t}\alpha_i + u_k^*(x)e^{i\omega_i t}\alpha_i^\dagger\} + \frac{1}{(2\pi)^{1/2}}\int \frac{dk}{(2\omega_k)^{1/2}}\{u_k(x)e^{-i\omega_k t}\alpha_k + u_k^*(x)e^{i\omega_k t}\alpha_k^\dagger\}, \tag{2.14}$$

where $u_i(x)$, ω_i and $u_k(x)$, ω_k are the bound-state and scattering-state eigenfunctions and eigenfrequencies respectively of Eq. (2.12). Note that Q depends on time, t. The function $\psi'_{-1}(x)$ is commonly referred to as the zero mode or translation mode.

Theorem 1: $\ddot{Q} = 0.$ (2.15)

Proof: $\psi_0(x)$ satisfies Eq. (2.12). Since $\psi'_{-1}(x)$ satisfies Eq. (2.12) and is time-dependent, $\ddot{Q} = 0$ in order that $Q\psi'_{-1}(x)$ satisfies Eq. (2.12). Q. E. D.

From Eq. (2.15), Q is given by

$$Q = q + \dot{Q}t. \tag{2.16}$$

In the following we will show, by the iteration of the field equation with (2.13) as the first order term, that Q in the dynamical map appears in the combination $x + Q$.

Definition: $\tilde{\psi}(x) = \psi(x)$ when $Q = \dot{Q} = 0.$ (2.17)

Therefore $\tilde{\psi}_n$ satisfies Eq. (2.11) with ψ_i replaced by $\tilde{\psi}_i$.

Theorem 2: When $\dot{Q} = 0$, the following relation holds:

$$\psi_n(x) = \sum_{k=0}^{n+1} \frac{1}{k!}(Q\cdot\nabla)^k \tilde{\psi}_{n-k}(x), \tag{2.18}$$

where we have used the symbol ∇ to denote the spatial derivative to emphasize the fact that this theorem is valid for any number of space dimensions.

Proof: For $n = -1$ and $n = 0$, Eq. (2.18) leads to ψ_{-1} and Eq. (2.13). Suppose that (2.18) is true for $\psi_m(x)$ when $m < n$. When we define $\tilde{\psi}_\lambda(x)$ by

$$\tilde{\psi}_\lambda(x) = \sum_{n=-1}^{\infty} \lambda^n \tilde{\psi}_n(x), \tag{2.19}$$

we see from Eq. (2.19) with $\psi(x)$ replaced by $\tilde{\psi}_i$ that $\tilde{\psi}_\lambda(x)$ satisfies

$$\Lambda(\partial_x)\tilde{\psi}_\lambda(x) = \lambda^{-1}F[\lambda\tilde{\psi}_\lambda(x)]. \tag{2.20}$$

Here, we have replaced the Klein–Gordon operator $(\partial^2 + m^2)$ by $\Lambda(\partial)$ in order to emphasize that the theorem is valid regardless of the form of the field operator. If $\dot{Q} = 0$, Eq. (2.20) is also valid for $\tilde{\psi}_\lambda(x + \lambda Q)$, since this is a simple space translation

$$\Lambda(\partial_x)\tilde{\psi}_\lambda(x + \lambda Q) = \lambda^{-1}F[\lambda\tilde{\psi}_\lambda(x + \lambda Q)]. \tag{2.21}$$

Note that Eq. (2.21) follows from Eq. (2.20) if

$$\Lambda(\partial_x)(Q\cdot\nabla)^l g(x) = (Q\cdot\nabla)^l \Lambda(\partial_x)g(x) \tag{2.22}$$

even when $\dot{Q} \neq 0$. $\tilde{\psi}_\lambda(x + \lambda Q)$ is given by an expansion in λ as

$$\tilde{\psi}_0(x) = \sum_i \frac{1}{(2\omega_i)^{1/2}}\{u_i(x)e^{-i\omega_i t}\alpha_i + u_k^*(x)e^{i\omega_i t}\alpha_i^\dagger\} = \sum_{l=-1}^{\infty} \lambda^l \sum_{k=0}^{\infty} \frac{(\lambda Q\nabla)^k}{k!}\tilde{\psi}_l(x)$$

$$= \sum_{n=-1}^{\infty} \lambda^n \bar{\psi}_n(x), \tag{2.23}$$

with

$$\bar{\psi}_n(x) = \sum_{k=0}^{n+1} \frac{(Q\nabla)^k}{k!} \tilde{\psi}_{n-k}(x). \tag{2.24}$$

The λ^n-th term of the right-hand side of Eq. (2.21) is calculated as

$$\begin{aligned} &\lambda^{-1} F[\lambda \tilde{\psi}_\lambda(x + \lambda Q)] \\ &= \lambda^{-1} F[\bar{\psi}_{-1}(x) + \lambda \bar{\psi}_0(x) + \dots + \lambda^{n+1} \bar{\psi}_n(x) + 0(\lambda^{n+2})] \\ &= \dots + \lambda^n \Big\{ F_1[\bar{\psi}_{-1}(x)] \bar{\psi}_n(x) \\ &\quad + \sum_{l=2}^{n+1} \frac{1}{l!} F_l[\bar{\psi}_{-1}(x)] \bar{\psi}_{\alpha_1}(x) \dots \bar{\psi}_{\alpha_l}(x) \Big\} + 0(\lambda^{n+1}), \end{aligned} \tag{2.25}$$

where, in the summation, $\alpha_1 + \dots + \alpha_l + l = n + 1$. Therefore comparing the λ^n-th terms of Eq. (2.21), we have

$$\begin{aligned} &\{\Lambda(\partial_x) - F_1[\psi_{-1}(x)]\} \bar{\psi}_n(x) \\ &= \sum_{l=2}^{n} \frac{1}{l!} F_l[\bar{\psi}_{-1}(x)] \bar{\psi}_{\alpha_1}(x) \cdots \bar{\psi}_{\alpha_l}(x). \end{aligned} \tag{2.26}$$

Since α_i in the right-hand side of Eq. (2.26) satisfies $0 \leqslant \alpha_i < n - 1$,

$$\bar{\psi}_{\alpha_i}(x) = \psi_{\alpha_i}(x). \tag{2.27}$$

From Eqs. (2.9) and (2.26), we have

$$\begin{aligned} &\{\Lambda(\partial_x) - F_1[\psi_{-1}(x)]\} \bar{\psi}_n(x) \\ &= \{\Lambda(\partial_x) - F_1[\psi_{-1}(x)]\} \psi_n(x). \end{aligned} \tag{2.28}$$

By requiring that the homogeneous terms which satisfy $\{\Lambda(\partial_x) - F_1[\psi_{-1}(x)]\} \delta\psi_n(x) = 0$ vanish when $\dot{Q} = 0$, we have

$$\psi_n(x) = \bar{\psi}_n(x). \tag{2.29}$$

This completes the proof of the theorem.

Given Theorem 2, it is easy to prove the following theorem.

Theorem 3. $\psi_\lambda(x) = \sum_{n=-1}^{\infty} \lambda^n \tilde{\psi}_n(x + \lambda Q). \quad (2.30)$

Proof. A combination of Eq. (2.7), (2.24), and (2.29) leads to

$$\begin{aligned} \psi_\lambda(x) &= \sum_{n=-1}^{\infty} \sum_{\eta=0}^{n+1} \frac{1}{\eta!} (Q\cdot\nabla)^\eta \tilde{\psi}_{n-\eta}(x) \lambda^n \\ &= \sum_{n=-1}^{\infty} \sum_{\eta=0}^{\infty} \frac{\lambda^\eta}{\eta!} (Q\cdot\nabla)^\eta \tilde{\psi}_{n-\eta}(x) \lambda^{n-\eta} \\ &= \sum_{n=-1}^{\infty} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} (Q\cdot\nabla)^k \tilde{\psi}_n(x) \lambda^n \\ &= \sum_{n=-1}^{\infty} \lambda^n \tilde{\psi}_n(x + \lambda Q). \end{aligned}$$

Thus the theorem is proved and

$$\psi_\lambda(x,t) = \tilde{\psi}_\lambda(x + \lambda Q, t). \tag{2.31}$$

In this way we see that when the time derivatives of Q are ignored Q appears everywhere in an additive combination with the space coordinate.

The remaining task is to determine how the presence of $\dot{Q}$ modifies the result of Eq. (2.31). A solution of this problem has already been given in Refs. 8 and 9. For the sake of completeness we will repeat the derivation here. In the proof of Theorem 2, the replacement of Q by $(X - x)$ with X^μ being a function of x^μ, Q, and $\dot{Q}$ does not change the steps when

$$\begin{aligned} &\Lambda(\partial)(X - x)^{\mu_1} \cdots (X - x)^{\mu_n} \partial_{\mu_1} \cdots \partial_{\mu_n} g(x) \\ &= (X - x)^{\mu_1} \cdots (X - x)^{\mu_n} \partial_{\mu_1} \cdots \partial_{\mu_n} \Lambda(\partial) g(x), \end{aligned} \tag{2.32}$$

similar to Eq. (2.22) holds. This replacement is considered as a partial summation of the perturbation series to include the effect of $\dot{Q}$. It has the following particular property:

$$\begin{aligned} &\left(\partial_\nu + \frac{\partial}{\partial X^\nu}\right) \{(X - x)^{\mu_1} \cdots (X - x)^{\mu_n} \partial_{\mu_1} \cdots \partial_{\mu_n} g(x)\} \\ &= (X - x)^{\mu_1} \cdots (X - x)^{\mu_n} \partial_{\mu_1} \cdots \partial_{\mu_n} \partial_\nu g(x). \end{aligned} \tag{2.33}$$

Here it is understood that ∂_ν does not operate on x in X. Given the replacement (2.32), the property of Eq. (2.33), and the notation

$$\partial / \partial X_\mu = D^\mu \tag{2.34}$$

an equation analogous to Eq. (2.28) can be written as

$$\begin{aligned} &\{\Lambda(\partial) - F_1[\psi_{-1}(x)]\} \psi_n(x) = \{\Lambda(D + \partial) - F_1[\psi_{-1}(x)]\} \\ &\times \sum_{k=0}^{n+1} \frac{1}{k!} (X - x)^{\mu_1} \cdots (X - x)^{\mu_k} \partial_{\mu_1} \cdots \partial_{\mu_k} \tilde{\psi}_{n-k}(x). \end{aligned}$$

Summing both sides over n, we obtain

$$\begin{aligned} &\{\Lambda(\partial) - F_1[\psi_{-1}(x)]\} \psi(x) \\ &= \{\Lambda(D) - F_1[\psi_{-1}(x)]\} \tilde{\psi}(X). \end{aligned} \tag{2.35}$$

Here in the right-hand side, we have $\Lambda(D)$ instead of $\Lambda(D + \partial)$, because $\tilde{\psi}(X)$ contains x only through X. If one now determines $X^\mu(= X, T)$ so that

$$\Lambda(\partial) = \Lambda(D), \tag{2.36}$$

$X^0 \to t$ and $X^i \to x + Q$ as $\dot{Q} \to 0$, and x and Q always appear as $x + Q$ in X, Eqs. (2.35) and (2.36) lead to

$$\psi^f(x,t) = \tilde{\psi}(X,T). \tag{2.37}$$

In the case of $\Lambda(\partial) = \partial^2 + m^2$ in $(1 + 1)$-dimensions we have

$$T = (1 - \dot{Q}^2)^{1/2} t + [\dot{Q}/(1 - \dot{Q}^2)^{1/2}](x + Q), \tag{2.38}$$

$$X = [1/(1 - \dot{Q}^2)]^{1/2} (x + Q). \tag{2.39}$$

In this section, using the perturbative development, we have explicitly shown that the quantum coordinate Q appears in the dynamical map in the combination $x + Q$ with the spatial coordinate x. Section III is devoted to a further exploration of the perturbative development of Eq. (2.37).

III. A PERTURBATIVE LOOK AT THE DYNAMICAL MAP

Consider the field equations corresponding to the first few orders of Eq. (2.9):

$$\begin{aligned} &\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\} \psi_1(x) \\ &= F_2[\psi_{-1}(x)] [\psi_0(x)]^2 / 2!, \end{aligned} \tag{3.1}$$

$$\begin{aligned} &\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\} \psi_2(x) \\ &= F_3[\psi_{-1}(x)] [\psi_0(x)]^3 / 3! + F_2[\psi_{-1}(x)] \psi_0(x) \psi_1(x). \end{aligned} \tag{3.2}$$

We seek a solution of these equations which is consistent with the replacement $x \to x + Q$. Substitution of Eq. (2.13) into Eq. (3.1) leads to

$$
\begin{aligned}
&\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\psi_1(x) \\
&\quad = (Q^2/2!)F_2[\psi_{-1}(x)]\psi'_{-1}(x)^2 + QF_2[\psi_{-1}(x)]\psi'_{-1}(x)\tilde{\psi}_0(x) + F_2[\psi_{-1}(x)][\tilde{\psi}_0(x)]^2/2! \\
&\quad = (Q^2/2!)\{F[\psi_{-1}(x)]'' - F_1[\psi_{-1}(x)]\psi''_{-1}(x)\} + Q\{(F_1[\psi_{-1}(x)]\tilde{\psi}_0(x))' - F_1[\psi_{-1}(x)]\tilde{\psi}'_0(x)\} \\
&\qquad + \{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\tilde{\psi}_1(x) \\
&\quad = (Q^2/2!)\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\psi''_{-1}(x) + Q\{\partial^2 - m^2 - F_1[\psi_{-1}(x)]\}\tilde{\psi}'_0(x) + \{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\tilde{\psi}_1(x) \\
&\quad = \{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\{(Q^2/2!)\psi''_{-1}(x) + Q\tilde{\psi}'_0(x) + \tilde{\psi}_1(x)\} - (2\dot{Q}^2/2!)\psi''_{-1}(x) - 2\dot{Q}\tilde{\psi}'_0(x).
\end{aligned}
\tag{3.3}
$$

Using the relations

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}x\psi'_{-1}(x) = -2\psi''_{-1}(x) \tag{3.4}$$

and

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}x\dot{\tilde{\psi}}_0(x) = -2\dot{\tilde{\psi}}'_0(x), \tag{3.5}$$

Eq. (3.3) may be written as

$$
\begin{aligned}
&\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\psi_1(x) \\
&\quad = \{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\{(Q^2/2!)\psi''_{-1}(x) \\
&\qquad + (\dot{Q}^2/2!)x\psi'_{-1}(x) + Q\tilde{\psi}'_0(x) + \dot{Q}x\dot{\tilde{\psi}}_0(x) + \tilde{\psi}_1(x)\}
\end{aligned}
\tag{3.6}
$$

and therefore

$$
\begin{aligned}
\psi_1(x) = {} & (Q^2/2!)\psi''_{-1}(x) + (\dot{Q}^2/2!)x\tilde{\psi}'_{-1}(x) \\
& + Q\tilde{\psi}'_0(x) + \dot{Q}x\dot{\tilde{\psi}}_0(x) + \tilde{\psi}_1(x).
\end{aligned}
\tag{3.7}
$$

Now, consider Eq. (3.2). A series of manipulations similar to those used in deriving Eq. (3.3) leads to

$$
\begin{aligned}
&\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\psi_2(x) \\
&= \{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\bigg\{\frac{Q^3}{3!}\psi'''_{-1}(x) + Q\frac{\dot{Q}^2}{2!}x\psi''_{-1}(x) \\
&\quad + \frac{Q^2}{2!}\tilde{\psi}''_0(x) + Q\dot{Q}x\tilde{\psi}'_0(x) + \frac{\dot{Q}^2}{2!}x\tilde{\psi}'_0(x) + Q\tilde{\psi}'_1(x) \\
&\quad + \dot{Q}x\dot{\tilde{\psi}}_1(x) + \tilde{\psi}_2(x)\bigg\} - 2\,\dot{Q}^2x\tilde{\psi}'_0(x).
\end{aligned}
\tag{3.8}
$$

Use of the identities

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}x^2\ddot{\tilde{\psi}}_0(x) = -4x\ddot{\tilde{\psi}}'_0(x) - 2\ddot{\tilde{\psi}}_0(x) \tag{3.9}$$

and

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}t\dot{\tilde{\psi}}_0(x) = 2\ddot{\tilde{\psi}}_0(x) \tag{3.10}$$

leads to

$$
\begin{aligned}
\psi_2(x) = {} & \frac{Q^3}{3!}\psi'''_{-1}(x) + Q\frac{\dot{Q}^2}{2!}x\psi''_{-1}(x)\ \frac{Q^2}{2!}\tilde{\psi}''_0(x) \\
& + Q\dot{Q}x\dot{\tilde{\psi}}'_0(x) + \frac{\dot{Q}^2}{2!}x\tilde{\psi}'_0(x) + \frac{\dot{Q}^2}{2!}x^2\ddot{\tilde{\psi}}_0(x) \\
& + \frac{\dot{Q}^2}{2!}t\dot{\tilde{\psi}}_0(x) + \dot{Q}x\dot{\tilde{\psi}}_1(x) + Q\tilde{\psi}'_1(x) + \tilde{\psi}_2(x).
\end{aligned}
\tag{3.11}
$$

However, Eq. (3.11) does not lead to $(\partial/\partial Q)\psi_2(x) = (\partial/\partial x)\psi_1(x)$, which is necessary if x and Q appear in the form $(x + Q)$. The difference appears through the explicit x dependence of $\psi_1(x)$

$$\psi'_1(x) - \frac{\partial}{\partial Q}\psi_2(x) = \frac{\dot{Q}^2}{2!}\psi'_{-1}(x) + \dot{Q}\dot{\tilde{\psi}}_0(x). \tag{3.12}$$

This difference must be remedied by introducing into $\psi_2(x)$ appropriate terms which satisfy the free field equation. The first term on the right-hand side of Eq. (3.12) can be compensated by adding $Q(\dot{Q}^2/2!)\psi'_{-1}(x)$ to $\psi_2(x)$. The second term can be compensated by adding $(Q - \dot{Q}t)\dot{Q}\dot{\tilde{\psi}}_0(x)$. In this way we arrive at

$$
\begin{aligned}
\psi_2(x) = {} & \frac{Q^3}{3!}\psi'''_{-1}(x) + Q\frac{\dot{Q}^2}{2!}x\psi''_{-1}(x) + Q\frac{\dot{Q}^2}{2!}\psi'_{-1}(x) \\
& + \frac{Q^2}{2!}\tilde{\psi}''_0(x) + Q\dot{Q}x\dot{\tilde{\psi}}'_0(x) + Q\dot{Q}\dot{\tilde{\psi}}_0(x) + \frac{\dot{Q}^2}{2!}x\tilde{\psi}'_0(x) \\
& + \frac{\dot{Q}^2}{2!}x^2\ddot{\tilde{\psi}}_0(x) - \frac{\dot{Q}^2}{2!}t\dot{\tilde{\psi}}_0(x) + \dot{Q}x\dot{\tilde{\psi}}_1(x) + Q\tilde{\psi}'_1(x) + \tilde{\psi}_2(x)
\end{aligned}
\tag{3.13}
$$

The dynamical map is

$$\psi^f(x) = \psi_{-1}(x) + Q\psi'_{-1}(x) + \tilde{\psi}_0(x) + \psi_1(x) + \psi_2(x) + \cdots. \tag{3.14}$$

The solution determined here is consistent with a power series expansion of Eqs. (2.37), (2.38), and (2.39): $\psi^f(x) = \tilde{\psi}^f(X,T)$.

IV. THE CANONICAL MOMENTUM

Consider the quantity[10]

$$P_\lambda = \int dx\ \ \dot{\psi}^f_\lambda(x)\psi^{f\prime}_\lambda(x). \tag{4.1}$$

The field equation (2.4) leads to the conservation law

$$\dot{P}_\lambda = 0 \tag{4.2}$$

and the equal-time canonical commutation relation of Eq. (2.2) leads to

$$[\psi^f_\lambda(x), P_\lambda] = i\psi^{f\prime}_\lambda(x) \tag{4.3}$$

and

$$[\dot{\psi}^f_\lambda(x), P_\lambda] = i\dot{\psi}^{f\prime}_\lambda(x). \tag{4.4}$$

Thus P generates spatial translations and can be identified with the canonical momentum of Q. Since x and Q appear only in the combination $x + Q$, conditions which are sufficient to yield Eqs. (4.3) and (4.4) are that

$$[Q, P_\lambda] = i/\lambda \tag{4.5}$$

and

$$[\tilde{\psi}_0(x), P_\lambda] = [\dot{\tilde{\psi}}_0(x), P_\lambda] = 0. \tag{4.6}$$

Combining Eqs. (2.7) and (4.1) leads to

$$P_n = \int dx\ \ \{\Sigma\, \dot{\psi}_k(x)\psi'_l(x)\}, \tag{4.7}$$

where $k + l = n; k \geqslant 0;\ \ l \geqslant -1$. The first few orders of Eq. (4.7) are

$$P_{-1} = \int dx\ \dot{\tilde{\psi}}_0(x)\psi'_{-1}(x), \tag{4.8}$$

$$P_0 = \int dx\ \{\dot{\tilde{\psi}}_1(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x)\}, \tag{4.9}$$

$$P_1 = \int dx\ \{\dot{\tilde{\psi}}_2(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_1(x)\tilde{\psi}'_0(x) + \dot{\tilde{\psi}}_0(x)\tilde{\psi}'_1(x)\}. \tag{4.10}$$

Equation (2.13) leads to

$$P_{-1} = M\dot{Q},\ M = \int dx\ \psi'_{-1}(x)\psi'_{-1}(x). \tag{4.11}$$

Also, Eqs. (2.13), (3.7), and (4.9) lead to

$$\begin{aligned} P_0 = \int dx\ \{&[Q\dot{Q}\psi''_{-1}(x) + \dot{Q}\tilde{\psi}'_0(x) + Q\dot{\tilde{\psi}}'_0(x) \\ &+ \dot{Q}x\ddot{\tilde{\psi}}_0(x) + \dot{\tilde{\psi}}_1(x)]\psi'_{-1}(x) + [\dot{Q}\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)] \\ &\times[Q\psi''_{-1}(x) + \tilde{\psi}'_0(x)]\} \end{aligned} \tag{4.12}$$

or

$$\begin{aligned} P_0 = {}& 2Q\dot{Q}\int dx\ \psi'_{-1}(x)\psi''_{-1}(x) + Q\int dx\ \{\dot{\tilde{\psi}}'_0(x)\psi'_{-1}(x) \\ &+ \dot{\tilde{\psi}}_0(x)\psi''_{-1}(x)\} + \dot{Q}\int dx\ \{\tilde{\psi}'_0(x)\psi'_{-1}(x) + x\ddot{\tilde{\psi}}_0(x)\psi'_{-1}(x) \\ &+ \tilde{\psi}'_0(x)\psi'_{-1}(x)\} + \int dx\ \{\dot{\tilde{\psi}}_1(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x)\}. \end{aligned} \tag{4.13}$$

Using the field equation for $\psi_0(x)$ and integrating by parts reduces Eq. (4.13) to

$$P_0 = \tilde{P}_0 = \int dx[\dot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + \dot{\tilde{\psi}}_1(x)\psi'_{-1}(x)], \tag{4.14}$$

where $\tilde{P}_n = P_n$ with $Q = \dot{Q} = 0$, that is

$$\tilde{P}_n = \int dx \sum_{l=0}^{n+1} \dot{\tilde{\psi}}_l(x)\tilde{\psi}'_{n-l}(x). \tag{4.15}$$

Combining Eqs. (2.13), (3.7), (3.13), and (4.10) results in

$$\begin{aligned} P_1 = {}& \dot{Q}^3\int dx\ \{x\,\psi''_{-1}(x)\psi'_{-1}(x) + \psi'_{-1}(x)\psi'_{-1}(x)\} + \dot{Q}^2\int dx\ \{x\tilde{\psi}'_0(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)\psi'_{-1}(x) + \tfrac{1}{2}x\dot{\tilde{\psi}}'_0\psi'_{-1}(x) \\ &+ \tfrac{1}{2}x^2\ddot{\tilde{\psi}}_0(x)\psi'_{-1}(x) - \tfrac{1}{2}t\ddot{\tilde{\psi}}_0(x)\psi'_{-1}(x) - \tfrac{1}{2}\dot{\tilde{\psi}}_0(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)\psi'_{-1}(x) + x\dot{\tilde{\psi}}'_0(x)\psi'_{-1}(x) + \tfrac{1}{2}x\dot{\tilde{\psi}}_0(x)\psi''_{-1}(x) \\ &+ \tfrac{1}{2}\dot{\tilde{\psi}}_0(x)\psi'_{-1}(x)\} + \dot{Q}\int dx\ \{x\ddot{\tilde{\psi}}_1(x)\psi'_{-1}(x) + \tilde{\psi}'_1(x)\tilde{\psi}'_{-1}(x)\tilde{\psi}'_0(x)\tilde{\psi}'_0(x) + x\ddot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + \psi'_{-1}(x)\tilde{\psi}'_1(x) \\ &+ \dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) + x\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}'_0(x)\} + \int dx\ \{\dot{\tilde{\psi}}_2(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_1(x)\tilde{\psi}'_0(x) + \dot{\tilde{\psi}}_0(x)\tilde{\psi}'_1(x)\}. \end{aligned} \tag{4.16}$$

Here the terms containing Q without time derivatives are dropped automatically, since the space integration of $(x + Q)$ in Eq. (4.1) guarantees the disappearance of Q in P_λ. This can also be checked by explicit calculation, as was done for Eq. (4.13).

Use of the free field Eq. (2.12) for $\tilde{\psi}_0(x)$ and $\psi'_{-1}(x)$, the orthogonality of $\psi'_{-1}(x)$ and $\tilde{\psi}_0(x)$ and integration by parts reduces Eq. (4.16) to

$$\begin{aligned} P_1 = {}& \frac{M}{2}\dot{Q}^3 + \dot{Q}\int dx\ \{x\ddot{\tilde{\psi}}_1(x)\psi'_{-1}(x) + 2\tilde{\psi}'_1(x)\psi'_{-1}(x) \\ &+ \tilde{\psi}'_0(x)\tilde{\psi}'_0(x) + x\ddot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + \tfrac{1}{2}\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x)\} + \tilde{P}_1. \end{aligned} \tag{4.17}$$

The field equation (3.1) for $\tilde{\psi}_1(x)$ leads to

$$\begin{aligned} \int dx\ x\ddot{\tilde{\psi}}_1(x)\psi'_{-1}(x) = {}& \int dx\ \left\{x\left(\frac{\partial^2}{\partial x^2} - m^2 + F_1[\psi_{-1}(x)]\right)\right. \\ &\times\tilde{\psi}_1(x)\psi'_{-1}(x) + xF_2[\psi_{-1}(x)]\psi'_{-1}(x)(\tilde{\psi}_0(x))^2/2!\} \\ = {}& \int dx\ \{x[\tilde{\psi}''_1(x)\psi'_{-1}(x) - \tilde{\psi}_1(x)\psi'''_{-1}(x)] \\ &- F_1[\psi_{-1}(x)][\tilde{\psi}_0(x)]^2/2! - xF_1[\psi_{-1}(x)]\tilde{\psi}_0(x)\tilde{\psi}'_0(x)\} \\ = {}& \int dx\ \{-2\tilde{\psi}'_1(x)\psi'_{-1}(x) - \tfrac{1}{2}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_0(x) \\ &- x\ddot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + \tilde{\psi}''_0(x)\tilde{\psi}_0(x)\}, \end{aligned} \tag{4.18}$$

and, upon substituting Eq. (4.18) into (4.17),

$$P_1 = (M/2)\dot{Q}^3 + (\dot{Q}/2)\tilde{H}_0 + \tilde{P}_1, \tag{4.19}$$

where

$$\tilde{H}_0 = \tfrac{1}{2}\int dx\ \{\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \ddot{\tilde{\psi}}_0(x)\tilde{\psi}_0(x)\}. \tag{4.20}$$

The structure of the quantities $\tilde{P}_n$ is closely related to the dynamical maps of $\tilde{\psi}_n$, which obey the equation

$$\begin{aligned} &\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\tilde{\psi}_n(x) \\ &= \sum\frac{1}{l!}F_l[\psi_{-1}(x)]\tilde{\psi}_{\alpha_1}(x)\ldots\tilde{\psi}_{\alpha_l}(x), \end{aligned} \tag{4.21}$$

where $\alpha_1 + \cdots + \alpha_l + l = m + 1$; $n\geqslant0$; $l\geqslant2$; $\alpha_1,\cdots,\alpha_l\geqslant0$.

Equation (4.21) can be integrated formally using the Green's function

$$\begin{aligned} g(x,y) = \int\frac{d\omega}{2\pi}\ e^{i\omega(x_0 - y_0)}\Bigg[&\frac{1}{M}\frac{\psi'_{-1}(x)\psi'_{-1}(y)}{-\omega^2} \\ &+ \sum_i\frac{u_i(x)u_i^*(y)}{\omega_i^2 - \omega^2} + \int\frac{dk}{2\pi}\frac{u_k(x)u_k^*(y)}{\omega_k^2 - \omega^2}\Bigg]. \end{aligned} \tag{4.22}$$

Though no eigenfunctions of the homogeneous equation (2.12) appear in the right-hand side of Eq. (4.21), the existence of the zero mode leads to possible nonoscillating terms in the right-hand side of Eq. (4.21). Therefore one needs special care for the component proportional to $\psi'_{-1}(x)$. Moreover, the operator $\tilde{\psi}_n$ can have additional homogeneous terms which are proportional to the eigenfunctions of Eq. (2.12). The renormalization of wavefunctions with nonzero frequency is not considered because the tree approximation is used. Therefore we are concerned only with the zero-mode wave function.

We write $\tilde{\psi}_n(x)$, by extracting the $\psi'_{-1}(x)$-component, as

$$\tilde{\psi}_n(x,t) = \alpha_{n-1}(t)\psi'_{-1}(x) + \hat{\tilde{\psi}}_n(x,t), \tag{4.23}$$

where

$$\int dx \;\; \psi'_{-1}(x)\hat{\tilde{\psi}}_n(x,t) = 0 \tag{4.24}$$

and

$$\alpha_{n-1}(t) = \frac{1}{M}\int dx \;\; \psi'_{-1}(x)\tilde{\psi}_n(x). \tag{4.25}$$

Equation (4.21) can be rewritten as

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\hat{\tilde{\psi}}_n(x) = \sum_{l=2}^{\infty}\frac{1}{l!}F_l[\psi_{-1}(x)] \times \tilde{\psi}_{\alpha_1}(x)\ldots\tilde{\psi}_{\alpha_l}(x) - \ddot{\alpha}_{n-1}(t)\psi'_{-1}(x). \tag{4.26}$$

When $\ddot{\alpha}_{n-1}(t)$ is chosen as

$$\ddot{\alpha}_{n-1}(t) = \frac{1}{M}\int dx \;\; \psi'_{-1}(x)\sum_{l=2}^{\infty}\frac{1}{l!}F_l[\psi_{-1}(x)]\tilde{\psi}_{\alpha_1}\ldots\tilde{\psi}_{\alpha_l}$$
$$= -\frac{1}{M}\int dx \;\; \sum_{l=2}^{\infty}\frac{1}{(l-1)!}F_{l-1}[\psi_{-1}(x)]\tilde{\psi}_{\alpha_1}\ldots\tilde{\psi}_{\alpha_{l-1}}\tilde{\psi}'_{\alpha_l}, \tag{4.27}$$

the right-hand side of Eq. (4.26) contains no $\psi'_{-1}(x)$ component. Equation (4.26) can be integrated for $\tilde{\psi}_n$ using the Green's function

$$\tilde{g}(x,y) = \int\frac{d\omega}{2\pi}\;\; e^{i\omega(x_0 - y_0)}\left[\sum_i\frac{u_i(x)u_i^*(y)}{\omega_i^2 - \omega^2} + \int\frac{dk}{2\pi}\;\;\frac{u_k(x)u_k^*(y)}{\omega_k^2 - \omega^2}\right]. \tag{4.28}$$

Using Eq. (2.9) in Eq. (4.27) leads to

$$\ddot{\alpha}_{n-1}(t) = -\frac{1}{M}\int dx \;\; \sum(\partial^2 + m^2)\tilde{\psi}_k(x)\tilde{\psi}'_l(x)$$

or

$$\ddot{\alpha}_{n-1}(x) = -\frac{1}{M}\frac{\partial}{\partial t}\int dx \;\; \sum\dot{\tilde{\psi}}_k(x)\tilde{\psi}'_l(x), \tag{4.29}$$

where $k + l = n - 1$; $k, l \geqslant 0$. Thus

$$\ddot{\alpha}_{n-1}(t) = -\frac{1}{M}\int dx \;\; \sum\dot{\tilde{\psi}}_k(x)\tilde{\psi}'_l(x) + \beta_{n-1}, \tag{4.30}$$

where

$$\ddot{\beta}_n = 0. \tag{4.31}$$

Now consider the cases $n = 1$ and $n = 2$ of Eq. (4.30).

$$\dot{\alpha}_0(t) = -\frac{1}{M}\int dx \;\; \dot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + \dot{\beta}_0, \tag{4.32}$$

$$\dot{\alpha}_1(t) = -\frac{1}{M}\int dx \;\; \{\dot{\tilde{\psi}}_0(x)\psi'_1(x) + \dot{\tilde{\psi}}_1(x)\tilde{\psi}'_0(x)\} + \dot{\beta}_1. \tag{4.33}$$

Using the identity

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}x\tilde{\psi}_0(x) = -2\tilde{\psi}'_0(x), \tag{4.34}$$

Eq. (4.32) can be written as

$$\dot{\alpha}_0(t) = \frac{1}{2M}\int dx \;\; \dot{\tilde{\psi}}_0(x)\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}x\tilde{\psi}_0(x) + \dot{\beta}_0 \tag{4.35}$$

$$= \frac{\partial}{\partial t}\frac{1}{2M}\int dx \;\; x\{\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \tilde{\psi}_0(x)\ddot{\tilde{\psi}}_0(x)\} + \dot{\beta}_0, \tag{4.36}$$

or

$$\alpha_0(t) = \frac{1}{2M}\int dx \;\; x\{\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \tilde{\psi}_0(x)\ddot{\tilde{\psi}}_0(x)\} + \beta_0. \tag{4.37}$$

Similarly, using the identity

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}x\tilde{\psi}_1(x) = x\frac{1}{2!}F_2[\psi_{-1}(x)](\tilde{\psi}_0(x))^2 - 2\tilde{\psi}'_1(x) \tag{4.38}$$

and Eq. (4.34), Eq. (4.33) can be written as

$$\dot{\alpha}_1(t) = \frac{1}{M}\frac{\partial}{\partial t}\int dx \;\; \left[x\left[\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_1(x) - \tfrac{2}{3}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_1(x) - \tfrac{1}{3}\tilde{\psi}_0(x)\ddot{\tilde{\psi}}_1(x)\right] - \tfrac{1}{3}\tilde{\psi}'_0(x)\tilde{\psi}_1(x)\right] + \dot{\beta}_1 \tag{4.39}$$

or

$$\alpha_1(t) = \frac{1}{M}\int \int dx \;\; \left[x\left[\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_1(x) - \tfrac{2}{3}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_1(x) - \tfrac{1}{3}\tilde{\psi}_0(x)\ddot{\tilde{\psi}}_1(x)\right] - \tfrac{1}{3}\tilde{\psi}'_0(x)\tilde{\psi}_1(x)\right] + \beta_1. \tag{4.40}$$

The quantities β_n must be constructed from the operators $\{\alpha_i, \alpha_i^\dagger, \alpha_k, \alpha_k^\dagger\}$ defined in Eq. (2.14). As will be shown in the following, the choice of β_n depends on the choice of the boundary conditions. Equation (4.31) leads to

$$\{\partial^2 + m^2 - F_1[\psi_{-1}(x)]\}\beta_n\psi'_{-1}(x) = 0, \tag{4.41}$$

that is, $\beta_n\psi'_{-1}(x)$ is a solution of the homogeneous free field equation. Therefore $\beta_{n-1}\psi_{-1}(x)$ appears in $\tilde{\psi}_n(x)$ in the same way as any solution of the homogeneous free field equation which may be added when integrating Eq. (4.21). In this way, all ambiguities arising from the presence of the so-called zero mode have been reduced to the choice of boundary conditions. These boundary conditions must be chosen in a way which is consistent with the solution that we are seeking. This solution corresponds to an expression for the Heisenberg field $\psi(x)$ in terms of the physical field $\tilde{\psi}_0(x)$ and the operators Q and P such that the commutation relations between members of the set $\{Q, P, \tilde{\psi}_0(x), \dot{\tilde{\psi}}_0(x)\}$ lead to the canonical commutation relation of Eq. (2.2).

The only commutation relations between members of the set $\{Q, P, \tilde{\psi}_0(x), \dot{\tilde{\psi}}_0(x)\}$ which have so far been specified are those in Eqs. (4.5) and (4.6). Physical considerations dictate one more requirement. The states of the system corresponding to $\tilde{\psi}_0(x)$ should be particlelike. That is, we require that

$$[\alpha_i, \alpha_j^\dagger] = \delta_{ij}, [\alpha_k, \alpha_l^\dagger] = \delta(k - l)$$

and

$$[\alpha_i, \alpha_j] = [\alpha_k, \alpha_l] = 0. \tag{4.42}$$

Equations (4.42), and (2.14) and the completeness of the set of functions $\{\psi'_{-1}(x), u_i(x), u_k(x)\}$ lead to

$$[\tilde{\psi}_0(x), \dot{\tilde{\psi}}_0(y)]_{x^0=y^0} = i\left\{\delta(x-y) - \frac{1}{M}\psi'_{-1}(x)\psi'_{-1}(y)\right\} \quad (4.43)$$

and

$$[\tilde{\psi}_0(x), \tilde{\psi}_0(y)]_{x^0=y^0} = [\tilde{\psi}_0(x), \tilde{\psi}_0(y)]_{x^0=y^0} = 0. \quad (4.44)$$

Using the free field Eq. (2.12), it is possible to show that

$$\dot{\tilde{H}}_0 = 0, \quad (4.45)$$

where $\tilde{H}_0$ is given in Eq. (4.20). This fact, together with Eqs. (4.43) and (4.44) leads to

$$[\tilde{\psi}_0(x), \tilde{H}_0] = i\dot{\tilde{\psi}}_0(x) \quad (4.46)$$

and

$$[\dot{\tilde{\psi}}_0(x), \tilde{H}_0] = i\ddot{\tilde{\psi}}_0(x). \quad (4.47)$$

Thus $\tilde{H}_0$ generates the time translation of the fields $\tilde{\psi}_0(x)$ and $\dot{\tilde{\psi}}_0(x)$.

Equation (4.15) leads to

$$\tilde{P}_n = \int dx \left\{\sum \dot{\tilde{\psi}}_k(x)\tilde{\psi}'_l(x) + \dot{\tilde{\psi}}_{n+1}(x)\psi'_{-1}(x)\right\}, \quad (4.48)$$

which, using Eqs. (4.25) and (4.30), reduces to

$$\tilde{P}_n = M\dot{\beta}_n. \quad (4.49)$$

This shows that appearance of $\tilde{P}_n$ is directly related to the boundary conditions.

We have now determined the canonical momentum to the first order as

$$P = (1/\lambda)M\dot{Q} + M\dot{\beta}_0 + \lambda(M/2)\dot{Q}^3 + \lambda\dot{Q}\tilde{H}_0 + \lambda M\dot{\beta}_1 + \cdots. \quad (4.50)$$

In Sec. V the Hamiltonian will be calculated. The Hamiltonian must generate the time translation of the Heisenberg fields. This is a result of the field equation and the canonical commutation relations for the Heisenberg fields. This places certain requirements on β_n and the commutation relations for the physical fields with Q.

V. THE HAMILTONIAN

Consider the quantity

$$H = \int dx \;\{\tfrac{1}{2}[\dot{\psi}^f(x)]^2 + \tfrac{1}{2}[\psi^{f\prime}(x)]^2 + \tfrac{1}{2}m^2[\psi^f(x)]^2 + V[\psi^f(x)]\}, \quad (5.1)$$

where $V[\psi^f(x)]$ is a local function of $\psi^f(x)$ such that

$$V_1[\psi^f(x)] = -F[\psi^f(x)] \quad (5.2)$$

and $V[\psi_{-1}(x)] \to 0$ as $x \to \pm\infty$. Using the Heisenberg field equation (2.1), it is possible to show that

$$\dot{H} = 0. \quad (5.3)$$

Then, using the commutation relation of Eq. (2.2) leads to

$$[\psi(x), H] = i\,\dot{\psi}(x), \quad (5.4)$$

$$[\dot{\psi}(x), H] = i\,\ddot{\psi}(x), \quad (5.5)$$

and H is the canonical Hamiltonian.

When the power-counting parameter is included, Eq. (5.1) becomes

$$H_\lambda = \int dx \;\{\tfrac{1}{2}[\dot{\psi}^f_\lambda(x)]^2 + \tfrac{1}{2}[\psi^{f\prime}_\lambda(x)]^2 + \tfrac{1}{2}m^2[\psi^f_\lambda(x)]^2 + \lambda^{-2}V[\lambda\psi^f_\lambda(x)]\}. \quad (5.6)$$

The interaction term $V[\lambda\psi^f(x)]$ can be expanded about the classical solution of the Heisenberg equation to get

$$H_n = \int dx \left\{\sum[\tfrac{1}{2}\dot{\psi}_k(x)\dot{\psi}_l(x) + \tfrac{1}{2}\psi'_k(x)\psi'_l(x) + \tfrac{1}{2}m^2\psi_k(x)\psi_l(x)] + \sum\frac{1}{l!}V_l[\psi_{-1}(x)]\psi_{\alpha_1}\cdots\psi_{\alpha_l}(x)\right\}, \quad (5.7)$$

where $k + l = n$ in the first summation and $\alpha_1 + \cdots + \alpha_l + l = n + 2$, $\alpha_1,\cdots,\alpha_l \geqslant 0$ in the second summation. Equation (5.7) leads to the following relations:

$$H_{-2} = \int dx \;\{\tfrac{1}{2}\psi'_{-1}(x)\psi'_{-1}(x) + \tfrac{1}{2}m^2[\psi_{-1}(x)]^2 + V[\psi_{-1}(x)]\}, \quad (5.8)$$

$$H_{-1} = \int dx \;\{\psi'_0(x)\psi'_{-1}(x) + m^2\psi_0(x)\psi_{-1}(x) - F[\psi_{-1}(x)]\psi_0(x)\}, \quad (5.9)$$

$$H_0 = \int dx \left\{\tfrac{1}{2}[\dot{\psi}_0(x)]^2 + \psi'_{-1}(x)\psi'_1(x) + \tfrac{1}{2}\psi'_0(x)\psi'_0(x) + m^2\psi_{-1}(x)\psi_1(x) + \frac{m^2}{2}[\psi_0(x)]^2 - F[\psi_{-1}(x)]\psi_1(x) - F_1[\psi_{-1}(x)][\psi_0(x)]^2/2!\right\}, \quad (5.10)$$

$$H_1 = \int dx \;\{\dot{\psi}_0(x)\dot{\psi}_1(x) + \psi'_{-1}(x)\psi'_2(x) + \psi'_0(x)\psi'_1(x) + m^2\psi_{-1}(x)\psi_2(x) + m^2\psi_0(x)\psi_1(x) - F[\psi_{-1}(x)]\psi_2(x) - F_1[\psi_{-1}(x)]\psi_0(x)\psi_1(x) - F_2[\psi_{-1}(x)][\psi_0(x)]^3/3!\}, \quad (5.11)$$

$$H_2 = \int dx \;\{\dot{\psi}_0(x)\dot{\psi}_2(x) + \tfrac{1}{2}[\dot{\psi}_1(x)]^2 + \psi'_{-1}(x)\psi'_3(x) + \psi'_0(x)\psi'_2(x) + \tfrac{1}{2}\psi'_1(x)\psi'_1(x) + m^2\psi_{-1}(x)\psi_3(x) + m^2\psi_0(x)\psi_2(x) + \tfrac{1}{2}m^2\psi_1(x)\psi_1(x) - F[\psi_{-1}(x)]\psi_3(x) - F_1[\psi_{-1}(x)](\psi_0(x)\psi_2(x) + [\psi_1(x)]^2/2!) - F_2[\psi_{-1}(x)][\psi_0(x)]^2/2!\psi_1(x) - F_3[\psi_{-1}(x)][\psi_0(x)]^4/4!\} \quad (5.12)$$

Use of Eqs. (2.12), (2.13), (3.1), and (3.2) leads to

$$H_{-2} = M, \quad (5.13)$$

$$H_{-1} = 0, \quad (5.14)$$

$$H_0 = \int dx \;\{\tfrac{1}{2}\dot{\psi}_0(x)\dot{\psi}_0(x) - \tfrac{1}{2}\ddot{\psi}_0(x)\psi_0(x)\}, \quad (5.15)$$

$$H_1 = \int dx \;\{\dot{\psi}_0(x)\dot{\psi}_1(x) - \tfrac{1}{3}\psi_0(x)\ddot{\psi}_1(x) - \tfrac{2}{3}\ddot{\psi}_0(x)\psi_1(x)\}, \quad (5.16)$$

$$H_2 = \int dx \;\{\dot{\psi}_2(x)\dot{\psi}_0(x) - \tfrac{1}{4}\ddot{\psi}_2(x)\psi_0(x) - \tfrac{3}{4}\psi_2(x)\ddot{\psi}_0(x) + \tfrac{1}{2}\dot{\psi}_1(x)\dot{\psi}_1(x) - \tfrac{1}{2}\psi_1(x)\ddot{\psi}_1(x)\}. \quad (5.17)$$

Equations (2.12) and (4.20) lead to

$$H_0 = \tfrac{1}{2}M\dot{Q}^2 + \tilde{H}_0. \quad (5.18)$$

Upon substitution of Eqs. (2.13) and (3.7), Eq. (5.16) becomes

$$
\begin{aligned}
H_1 = \dot{Q}^2 \int dx\ & \{\psi'_{-1}(x)\tilde{\psi}'_0(x) + x\psi'_{-1}(x)\ddot{\tilde{\psi}}_0(x) \\
& - \tfrac{1}{3}\psi''_{-1}(x)\tilde{\psi}_0(x) - \tfrac{1}{3}x\psi'_{-1}(x)\ddot{\tilde{\psi}}_0(x)\} \\
+ \dot{Q}\int dx\ & \{\psi'_{-1}(x)\dot{\tilde{\psi}}_1(x) + \dot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + x\dot{\tilde{\psi}}_0(x)\ddot{\tilde{\psi}}_0(x) \\
& - \tfrac{2}{3}\tilde{\psi}_0(x)\dot{\tilde{\psi}}'_0(x) - \tfrac{1}{3}x\tilde{\psi}_0(x)\dddot{\tilde{\psi}}_0(x) - \tfrac{2}{3}x\ddot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x)\} \\
+ \int dx\ & \{\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_1(x) - \tfrac{1}{3}\tilde{\psi}_0(x)\ddot{\tilde{\psi}}_1(x) - \tfrac{2}{3}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_1(x)\}\,. \qquad (5.19)
\end{aligned}
$$

Here terms containing Q are automatically dropped, since the space integration in Eq. (5.1) guarantees the disappearance of Q. They are also shown to disappear by explicit calculation. Use of Eqs. (2.12), (2.13), (3.1), and (3.2) leads to

$$H_1 = M\dot{Q}\dot{\beta}_0. \qquad (5.20)$$

We define $\tilde{H}_n$ by $\tilde{H}_n = H_n$, when $Q = \dot{Q} = 0$. Since $\tilde{\psi}_n$ satisfies the field equation, $\tilde{H}_n$ can be shown to satisfy

$$\dot{\tilde{H}}_n = 0.$$

$\tilde{H}_n$ contains zero-frequency components only; the latter components for $n \geqslant 1$ originate from $\alpha_{n-1}(t)\psi'_{-1}(x)$ $(n \geqslant 1)$. The $\alpha_0\psi'_{-1}(x)$ in the last term in Eq. (5.19) drops out because of the orthogonality of $\psi'_{-1}(x)$ and $\tilde{\psi}_0$.

A combination of Eqs. (2.13), (3.7), (3.13), and (5.17) leads to

$$
\begin{aligned}
H_2 = {} & \dot{Q}^4\int dx\ \{\tfrac{1}{2}x\psi''_{-1}(x)\psi'_{-1}(x) + \tfrac{1}{2}\psi'_{-1}(x)\psi'_{-1}(x) - \tfrac{1}{4}x\psi''_{-1}(x)\psi'_{-1}(x)\} + \dot{Q}^3\int dx\ \{\tfrac{1}{2}x\psi''_{-1}(x)\dot{\tilde{\psi}}_0(x) + \tfrac{1}{2}\psi'_{-1}(x)\dot{\tilde{\psi}}_0(x) \\
& + x\dot{\tilde{\psi}}_0(x)\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)\psi'_{-1}(x) + \tfrac{1}{2}x\dot{\tilde{\psi}}'_0(x)\psi'_{-1}(x) + \tfrac{1}{2}x^2\dddot{\tilde{\psi}}_0(x)\psi'_{-1}(x) - \tfrac{1}{2}\dot{\tilde{\psi}}_0(x)\psi'_{-1}(x) - \tfrac{1}{2}t\ddot{\tilde{\psi}}_0(x)\psi'_{-1}(x) \\
& - \tfrac{1}{2}x\psi''_{-1}(x)\dot{\tilde{\psi}}_0(x) - \tfrac{1}{2}x\dot{\tilde{\psi}}'_0(x)\psi'_{-1}(x) - \tfrac{1}{4}x^2\dddot{\tilde{\psi}}_0(x)\psi'_{-1}(x)\} + \dot{Q}^2\int dx\ \{x\dot{\tilde{\psi}}'_0(x)\dot{\tilde{\psi}}_0(x) + \dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) + \tfrac{1}{2}x\dot{\tilde{\psi}}'_0(x)\dot{\tilde{\psi}}_0(x) \\
& + \tfrac{1}{2}x^2\dddot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \tfrac{1}{2}\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \tfrac{1}{2}t\ddot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) + x\ddot{\tilde{\psi}}_1(x)\psi'_{-1}(x) + \tilde{\psi}'_1(x)\psi'_{-1}(x) - \tfrac{1}{4}\tilde{\psi}''_0(x)\tilde{\psi}_0(x) - \tfrac{1}{2}x\ddot{\tilde{\psi}}'_0(x)\tilde{\psi}_0(x) \\
& - \tfrac{1}{2}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_0(x) - \tfrac{1}{8}x\ddot{\tilde{\psi}}'_0(x)\tilde{\psi}_0(x) - \tfrac{1}{8}x^2\dddot{\tilde{\psi}}_0(x)\tilde{\psi}_0(x) + \tfrac{1}{4}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_0(x) + \tfrac{1}{8}t\dddot{\tilde{\psi}}_0(x)\tilde{\psi}_0(x) - \tfrac{3}{8}x\tilde{\psi}'_0(x)\ddot{\tilde{\psi}}_0(x) \\
& - \tfrac{3}{8}x^2\ddot{\tilde{\psi}}_0(x)\ddot{\tilde{\psi}}_0(x) + \tfrac{3}{8}t\dot{\tilde{\psi}}_0(x)\ddot{\tilde{\psi}}_0(x) + \tfrac{1}{2}\tilde{\psi}'_0(x)\tilde{\psi}'_0(x) + \tfrac{1}{2}x^2\ddot{\tilde{\psi}}_0(x)\ddot{\tilde{\psi}}_0(x) + x\tilde{\psi}'_0(x)\ddot{\tilde{\psi}}_0(x) - \tfrac{1}{2}\psi''_{-1}(x)\tilde{\psi}_1(x) - x\dot{\tilde{\psi}}'_0(x)\dot{\tilde{\psi}}_0(x) \\
& - \tfrac{1}{2}x^2\dddot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \tfrac{1}{4}x\ddot{\tilde{\psi}}_1(x)\psi'_{-1}(x)\} + \dot{Q}\int dx\ \{x\ddot{\tilde{\psi}}_1(x)\dot{\tilde{\psi}}_0(x) + \tilde{\psi}'_1(x)\dot{\tilde{\psi}}_0(x) + \dot{\tilde{\psi}}_2(x)\psi'_{-1}(x) - \tfrac{1}{4}x\dddot{\tilde{\psi}}_1(x)\tilde{\psi}_0(x) \\
& - \tfrac{3}{4}x\dot{\tilde{\psi}}_1(x)\ddot{\tilde{\psi}}_0(x) + \tilde{\psi}'_0(x)\dot{\tilde{\psi}}_1(x) + x\ddot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_1(x) - \dot{\tilde{\psi}}'_0(x)\tilde{\psi}_1(x) - \tfrac{1}{2}x\dddot{\tilde{\psi}}_0(x)\tilde{\psi}_1(x) - \tfrac{1}{2}x\ddot{\tilde{\psi}}_1(x)\dot{\tilde{\psi}}_0(x) - \tfrac{1}{2}\dot{\tilde{\psi}}'_1(x)\tilde{\psi}_0(x)\} \\
& + \int dx\ \{\dot{\tilde{\psi}}_2(x)\dot{\tilde{\psi}}_0(x) - \tfrac{1}{4}\ddot{\tilde{\psi}}_2(x)\tilde{\psi}_0(x) - \tfrac{3}{4}\tilde{\psi}_2(x)\ddot{\tilde{\psi}}_0(x) + \tfrac{1}{2}\dot{\tilde{\psi}}_1(x)\dot{\tilde{\psi}}_1(x) - \tfrac{1}{2}\ddot{\tilde{\psi}}_1(x)\tilde{\psi}_1(x)\}. \qquad (5.21)
\end{aligned}
$$

This expression can be reduced to

$$H_2 = \tfrac{3}{8}M\dot{Q}^4 + \tfrac{1}{2}\dot{Q}^2 H_0 + M\dot{Q}\dot{\beta}_1 + \tilde{H}_2(\beta_0). \qquad (5.22)$$

The last term in Eq. (5.21) cannot be concluded to be zero immediately because of the terms $\beta_0(t)\,\psi_{-1}(x)$ in $\tilde{\psi}_1(x)$ and $\beta_1(t)\,\psi'_{-1}(x)$ in $\tilde{\psi}_2(x)$. The $\beta_1(t)\,\psi'_{-1}(x)$ terms drop out because of the orthogonality of $\psi'_{-1}(x)$ and $\tilde{\psi}_0(x)$. The Hamiltonian to second order is

$$H = (1/\lambda^2)M + \tfrac{1}{2}M\dot{Q}^2 + \tilde{H}_0 + \lambda M\dot{Q}\dot{\beta}_0 + \lambda^2\{\tfrac{3}{8}M\dot{Q}^4 + \tfrac{1}{2}\dot{Q}^2\tilde{H}_0 + M\dot{Q}\dot{\beta}_1 + H_2(\beta_0)\} + \cdots. \qquad (5.23)$$

When Q and P are chosen as independent dynamical variables, $\dot{Q}$ is a dependent variable and may be determined from a formal inversion of Eq. (4.50),

$$\dot{Q} = \hat{P}/M - \lambda\dot{\beta}_0 - \lambda^2\{\tfrac{1}{2}(\hat{P}^3/M^3) + (\hat{P}/M^2)\tilde{H}_0 + \dot{\beta}_1\} + \cdots, \qquad (5.24)$$

where $\hat{P} = \lambda P_\lambda$. Note that $\hat{P}$ is independent of λ.

Combining Eqs. (5.23) and (5.24) leads to

$$H_\lambda = \frac{M}{\lambda^2}\left\{1 + \tfrac{1}{2}\lambda^2\frac{\hat{P}^2}{M^2} - \tfrac{1}{8}\lambda^4\frac{\hat{P}^4}{M^4}\right\} + \left\{1 - \tfrac{1}{2}\lambda^2\frac{\hat{P}^2}{M^2}\right\}\tilde{H}_0 + \lambda^2\left[\tilde{H}_2(\beta_0) - \frac{M\dot{\beta}_0^2}{2}\right] + \cdots. \qquad (5.25)$$

Combining Eqs. (2.13), (3.7), (3.13), (3.14), and (5.24) leads to

$$
\begin{aligned}
\psi^f_\lambda(x) = {} & \frac{1}{\lambda}\psi_{-1}(x) + \{Q\psi'_{-1}(x) + \tilde{\psi}_0(x)\} + \lambda\left\{\frac{Q^2}{2!}\psi''_{-1}(x) + \frac{1}{2!}\frac{\hat{P}^2}{M^2}x\psi'_{-1}(x) + Q\tilde{\psi}'_0(x) + \frac{\hat{P}}{M}x\dot{\tilde{\psi}}_0(x) + \tilde{\psi}_1(x)\right\} \\
& + \lambda^2\left\{\frac{Q^3}{3!}\psi'''_{-1}(x) + \frac{1}{2}Q\frac{\hat{P}^2}{M^2}x\psi''_{-1}(x) + \frac{1}{2!}Q\frac{\hat{P}^2}{M^2}\psi'_{-1}(x) - \frac{\hat{P}}{M}\beta_0 x\psi'_{-1}(x) + \frac{Q^2}{2!}\tilde{\psi}''_0(x) + Q\frac{\hat{P}}{M}x\dot{\tilde{\psi}}'_0(x) + Q\frac{\hat{P}}{M}\dot{\tilde{\psi}}_0(x)\right. \\
& \left. + \frac{1}{2}\frac{\hat{P}^2}{M^2}x\tilde{\psi}'_0(x) + \frac{1}{2}x^2\frac{\hat{P}^2}{M^2}\tilde{\psi}''_0(x) - \frac{1}{2}t\frac{\hat{P}^2}{M^2}\dot{\tilde{\psi}}_0(x) - x\dot{\beta}_0\dot{\tilde{\psi}}_0(x) + x\frac{\hat{P}}{M}\dot{\tilde{\psi}}_1(x) + Q\tilde{\psi}'_1(x) + \tilde{\psi}_2(x) + \ldots\right\}, \qquad (5.26)
\end{aligned}
$$

and

$$\Pi^f_\lambda(x) = \left\{\frac{\hat{P}}{M}\psi'_{-1}(x) + \dot{\tilde{\psi}}_0(x)\right\} + \lambda\left\{Q\frac{\hat{P}}{M}\psi''_{-1}(x) - \dot{\beta}_0\psi'_{-1}(x) + \frac{\hat{P}}{M}\tilde{\psi}'_0(x) + Q\dot{\tilde{\psi}}'_0(x) + x\frac{\hat{P}}{M}\ddot{\tilde{\psi}}_0(x) + \dot{\tilde{\psi}}_1(x)\right\}$$

$$+\lambda^2\Big\{\frac{1}{2}Q^2\frac{\hat P}{M}\psi'''_{-1}(x)+\frac{1}{2}x\frac{\hat P^3}{M^3}\psi''_{-1}(x)-\frac{\hat P}{M^2}\tilde H_0\psi'_{-1}(x)-\dot\beta_1\psi'_{-1}(x)-Q\dot\beta_0\psi''_{-1}(x)+\frac{Q^2}{2!}\dot{\tilde\psi}''_0(x)+Q\frac{\hat P}{M}\tilde\psi''_0(x)$$
$$+\frac{\hat P^2}{M^2}x\dot{\tilde\psi}'_0(x)+xQ\frac{\hat P}{M}\ddot{\tilde\psi}'_0(x)+\frac{\hat P^2}{M^2}\dot{\tilde\psi}_0(x)+Q\frac{\hat P}{M}\ddot{\tilde\psi}_0(x)+\frac{1}{2}x\frac{\hat P^2}{M^2}\dot{\tilde\psi}'_0(x)+\frac{1}{2}x^2\frac{\hat P^2}{M^2}\dddot{\tilde\psi}_0(x)-\frac{1}{2}\frac{\hat P^2}{M^2}\dot{\tilde\psi}_0(x)$$
$$-\frac{1}{2}t\frac{\hat P^2}{M^2}\ddot{\tilde\psi}_0(x)-\dot\beta_0\tilde\psi'_0(x)-x\dot\beta_0\ddot{\tilde\psi}_0(x)+x\frac{\hat P}{M}\ddot{\tilde\psi}_1(x)+\frac{\hat P}{M}\tilde\psi'_1(x)+Q\dot{\tilde\psi}'_1(x)+\dot{\tilde\psi}_2(x)\Big\}+\cdots, \tag{5.27}$$

where $\Pi^f_\lambda(x)=\dot\psi^f_\lambda(x)$.

Equations (5.25), (5.26), and (5.27) must together satisfy Eq. (5.4)

$$i\,\Pi^f_\lambda(x)=[\psi^f_\lambda(x),H_\lambda]. \tag{5.28}$$

Equation (5.28) leads to the following conditions:

$$\dot\beta_0=\dot\beta_1=\tilde H_2(\beta_0)=0 \tag{5.29}$$

and

$$[Q,\tilde H_0]=0. \tag{5.30}$$

These conditions are also sufficient to yield

$$i\,\psi^{f\prime}(x)=[\psi^f(x),P], \tag{5.31}$$

$$i\,\Pi^{f\prime}(x)=[\Pi^f(x),P]. \tag{5.32}$$

Combining Eqs (5.24), (5.25), (5.29), and (5.30) leads to

$$i\,\dot Q=[Q,H]. \tag{5.33}$$

Thus, in contrast to the case of $\tilde\psi_0(x)$ in Eqs. (4.46) and (4.47), the time derivative of Q is generated by the full Hamiltonian.

In summary, we have now arrived at the following results:

$$\dot Q_\lambda=\frac{\hat P}{M}-\lambda^2\Big\{\tfrac{1}{2}\frac{\hat P^3}{M^3}+\frac{\hat P}{M^2}\tilde H_0\Big\}+O(\lambda^3), \tag{5.34}$$

$$H_\lambda=\frac{M}{\lambda^2}\Big\{1+\tfrac{1}{2}\lambda^2\frac{\hat P^2}{M^2}-\tfrac{1}{8}\lambda^4\frac{\hat P^4}{M^4}\Big\}+\Big\{1-\tfrac{1}{2}\lambda^2\frac{\hat P^2}{M^2}\Big\}\tilde H_0+O(\lambda^3), \tag{5.35}$$

$$\psi^f_\lambda(x)=\frac{1}{\lambda}\psi_{-1}(x)+\{Q\psi'_{-1}(x)+\tilde\psi_0(x)\}+\lambda\Big\{\frac{Q^2}{2!}\psi''_{-1}(x)+\frac{\hat P^2}{2!M^2}x\psi'_{-1}(x)+Q\tilde\psi'_0(x)+\frac{\hat P}{M}x\dot{\tilde\psi}_0(x)+\tilde\psi_1(x)\Big\}+\lambda^2$$
$$\times\Big\{\frac{Q^3}{3!}\psi'''_{-1}(x)+\frac{Q\hat P^2}{2!M^2}x\psi''_{-1}(x)+\frac{Q\hat P^2}{2!M^2}\psi'_{-1}(x)+\frac{Q^2}{2!}\tilde\psi''_0(x)+\frac{Q\hat P}{M}x\dot{\tilde\psi}'_0(x)+\frac{Q\hat P}{M}\dot{\tilde\psi}_0(x)+\frac{\hat P^2}{2!M^2}x\tilde\psi'_0(x)+\frac{\hat P^2}{2!M^2}x^2\ddot{\tilde\psi}_0(x)$$
$$-\frac{\hat P^2}{2!M^2}t\dot{\tilde\psi}_0(x)+\frac{\hat P}{M}x\dot{\tilde\psi}_1(x)+Q\tilde\psi'_1(x)+\tilde\psi_2(x)\Big\}+O(\lambda^3), \tag{5.36}$$

$$\Pi^f_\lambda(x)=\Big\{\frac{\hat P}{M}\psi'_{-1}(x)+\dot{\tilde\psi}_0(x)\Big\}+\lambda\Big\{\frac{Q\hat P}{M}\psi''_{-1}(x)+Q\dot{\tilde\psi}'_0(x)+\frac{\hat P}{M}\tilde\psi'_0(x)+\frac{\hat P}{M}x\ddot{\tilde\psi}_0(x)+\dot{\tilde\psi}_1(x)\Big\}+\lambda^2\Big\{\frac{Q^2\hat P}{2!M}\psi'''_{-1}(x)$$
$$+\frac{\hat P^3}{2!M^3}x\psi''_{-1}(x)-\frac{\hat P}{M^2}\tilde H_0\psi'_{-1}(x)+\frac{Q^2}{2!}\dot{\tilde\psi}''_0(x)+\frac{Q\hat P}{M}\tilde\psi''_0(x)+\frac{Q\hat P}{M}x\ddot{\tilde\psi}'_0(x)+\frac{3\hat P^2}{2!M^2}x\tilde\psi'_0(x)+\frac{Q\hat P}{M}\ddot{\tilde\psi}_0(x)$$
$$+\frac{\hat P^2}{2!M^2}\dot{\tilde\psi}_0(x)+\frac{\hat P^2}{2!M^2}x^2\dddot{\tilde\psi}_0(x)-\frac{\hat P^2}{2!M^2}t\ddot{\tilde\psi}_0(x)+\frac{\hat P}{M}x\ddot{\tilde\psi}_1(x)+Q\dot{\tilde\psi}'_1(x)+\frac{\hat P}{M}\tilde\psi'_1(x)+\dot{\tilde\psi}_2(x)\Big\}+O(\lambda^3), \tag{5.37}$$

$$[Q,\hat P]=i, \tag{5.38}$$

$$[\tilde\psi_0(x),\dot{\tilde\psi}_0(y)]_{x^0=y^0}=i\{\delta(x-y)-[\psi'_{-1}(x)\psi'_{-1}(y)]/M\}, \tag{5.39}$$

$$[P,\tilde\psi_0(x)]=[P,\dot{\tilde\psi}_0(x)]=[\tilde\psi_0(x),\tilde\psi_0(y)]_{x^0=y^0}=[\dot{\tilde\psi}_0(x),\dot{\tilde\psi}_0(y)]_{x^0=y^0}=[Q,\tilde H_0]=0, \tag{5.40}$$

$$[Q,H]=i\dot Q, \tag{5.41}$$

$$[\tilde\psi_0(x),\tilde H_0]=i\,\dot{\tilde\psi}_0(x). \tag{5.42}$$

In Sec. VI, we will use the information from Eqs. (5.34) to (5.42) to compute the commutation relation for the Heisenberg fields. This will impose further requirements on the commutation relations between the set of operators $\{Q,\hat P,\tilde\psi_0(x),\dot{\tilde\psi}_0(x)\}$. We notice, finally, that the expansions of $\dot Q$ and H in Eqs. (5.34) and (5.35) are consistent with expansions of

$$\dot Q=\hat P/[\hat P^2+(M+\tilde H_0)^2]^{1/2}\text{ or }P=[\dot Q(M+\tilde H_0)]/(1-\dot Q^2)^{1/2} \tag{5.43}$$

and

$$H=[\hat P^2+(M+\tilde H_0)^2]^{1/2}=(M+\tilde H_0)/(1-\dot Q^2)^{1/2}, \tag{5.44}$$

which can also be derived from general arguments using the Lorentz symmetry.[11]

VI. THE CANONICAL COMMUTATION RELATIONS

In this section we investigate the canonical equal time commutation relation

$$[\psi^f_\lambda(x),\Pi^f_\lambda(y)]_{x^0=y^0}=i\,\delta(x-y). \tag{6.1}$$

In the course of the calculation, we will obtain the condition imposed on the commutation relation between Q and $\tilde\psi_0(x)$ by Eq. (6.1). It is shown that Q_0 can be chosen independently of $\tilde\psi_0(x)$ as the canonical conjugate of $\hat P$.

We separate the spatial component of $\psi'_{-1}(x)$ as follows:

$$\psi^f_\lambda(x)=(1/M)\Psi_\lambda(t)\psi'_{-1}(x)+\hat\psi^f_\lambda(x);$$
$$\Pi^f_\lambda(x)=(1/M)\Pi_\lambda(t)\psi'_{-1}(x)+\hat\Pi^f_\lambda(x). \tag{6.2}$$

The condition (6.1) leads to

$$[\Psi_\lambda(t),\Pi_\lambda(t)]=iM. \tag{6.3}$$

The operators $\Psi_\lambda(t)$ and $\Pi_\lambda(t)$ are expanded as

$$\Psi_\lambda(t) = QM + \lambda\Psi_1 + \lambda^2\Psi_2 + \cdots, \tag{6.4a}$$

$$\Pi_\lambda(t) = \hat{P} + \lambda\Pi_1 + \lambda^2\Pi_2 + \cdots. \tag{6.4b}$$

From the dynamical maps of ψ^f in Eq. (5.36) and Π^f in Eq. (5.37), we have

$$\Psi_1 = \int dx\ \psi'_{-1}(x)\Big[\frac{\hat{P}^2}{2M^2}x\psi'_{-1}(x) + Q\tilde\psi_0'(x) + \frac{\hat{P}}{M}x\dot{\tilde\psi}_0(x) + \tilde\psi_1(x)\Big], \tag{6.5}$$

$$\Psi_2 = \int dx\ \psi'_{-1}(x)\Big[\frac{Q^3}{3!}\psi'''_{-1}(x) + \frac{Q\hat{P}^2}{4M^2}\psi'_{-1}(x) + \frac{Q^2}{2!}\tilde\psi_0''(x) + \frac{Q\hat{P}}{M}x\tilde\psi_0'(x) - \frac{3}{2}\frac{\hat{P}^2}{M^2}x\tilde\psi_0'(x) + \frac{\hat{P}}{M}x\tilde\psi_1(x) + Q\tilde\psi_1'(x) + \tilde\psi_2(x)\Big], \tag{6.6}$$

$$\Pi_1 = \int dx\ \psi'_{-1}(x)\Big[Q\dot{\tilde\psi}_0'(x) - \frac{\hat{P}}{M}\tilde\psi_0'(x) + \dot{\tilde\psi}_1(x)\Big], \tag{6.7}$$

$$\Pi_2 = \int dx\ \psi'_{-1}(x) \times\Big[\frac{Q^2\hat{P}}{2M}\psi'''_{-1}(x) + \frac{\hat{P}^3}{2M^3}x\psi''_{-1}(x) - \frac{\hat{P}}{M^2}\tilde{H}_0\psi'_{-1}(x) + \frac{Q^2}{2}\dot{\tilde\psi}_0''(x) + \frac{Q\hat{P}}{M}(\tilde\psi_0''(x) + x\dddot{\tilde\psi}_0'(x)) - \frac{\hat{P}^2}{2M^2}x\dot{\tilde\psi}_0'(x) + \frac{\hat{P}}{M}x\tilde\psi_1(x) + Q\dot{\tilde\psi}_1'(x) + \frac{\hat{P}}{M}\tilde\psi_1'(x) + \tilde\psi_2(x)\Big]. \tag{6.8}$$

To get Eqs. (6.5) ~ (6.8), the following relations are used:

$$\int dx\ x\psi'_{-1}(x)\psi''_{-1}(x) = -\tfrac{1}{2}\int dx\ \psi'_{-1}(x)\psi'_{-1}(x), \tag{6.9}$$

$$\int dx\ x^2\psi'_{-1}(x)\dddot{\tilde\psi}_0(x) = -4\int dx\ x\psi'_{-1}(x)\dot{\tilde\psi}_0'(x), \tag{6.10}$$

$$\int dx\ x^2\psi'_{-1}(x)\ddddot{\tilde\psi}_0(x) = -4\int dx\ x\psi'_{-1}(x)\dot{\tilde\psi}_0'(x), \tag{6.11}$$

$$\int dx\ x\psi'_{-1}(x)\ddot{\tilde\psi}_0(x) = -2\int dx\ \psi'_{-1}(x)\tilde\psi_0'(x). \tag{6.12}$$

The derivation of Eqs. (6.9)–(6.12) uses Eqs. (2.12), (3.4), and (3.5). Equation (6.3) leads to

$$[MQ, \hat{P}] = iM, \tag{6.13}$$

$$[MQ, \Pi_1] + [\Psi_1, \hat{P}] = 0, \tag{6.14}$$

$$[MQ, \Pi_2] + [\Psi_1, \Pi_1] + [\Psi_2, \hat{P}] = 0. \tag{6.15}$$

From Eq. (5.38) we see that Eq. (6.13) is satisfied. In order to evaluate the left-hand side of Eq. (6.14), we need some information about the commutation relation between Q and $\tilde\psi_0(x)$ and $\dot{\tilde\psi}_0(x)$. We will proceed with the assumption that

$$[Q, \tilde\psi_0(x)] = O(\lambda^2),$$
$$[Q, \dot{\tilde\psi}_0(x)] = O(\lambda^2), \tag{6.16}$$

at equal times. This assumption has already been used in obtaining Eq. (6.15). If Eqs. (6.14) and (6.15) can be reduced to identities, assumption (6.16) will be justified.

Equation (6.14) is also satisfied with the assumption (6.16). Since Eq. (6.14) contains $[Q, \tilde\psi_0(x)]$ and $[Q, \dot{\tilde\psi}_0(x)]$ independently, Eq. (6.16) is, in fact, a sufficient condition for Eq. (6.14).

Using Eqs. (5.38), (5.39), (6.16), and (4.18), Eq. (6.15) is reduced to

$$Q\Big\{2\int dx\ \psi'_{-1}(x)[\tilde\psi_0''(x) + x\dddot{\tilde\psi}_0'(x)] + \int dx\ dy\ \psi'_{-1}(x)\psi'_{-1}(y)\Big(\frac{1}{i}[\tilde\psi_0'(x),\dot{\tilde\psi}_1(y)] + \frac{1}{i}[\tilde\psi_1(x),\dot{\tilde\psi}_0'(y)]\Big)\Big\} + \frac{\hat{P}}{M}\int dx\ dy\ \psi'_{-1}(x)\psi'_{-1}(y) \times\Big\{\frac{x}{i}[\dot{\tilde\psi}_0(x),\dot{\tilde\psi}_1(y)] - \frac{1}{i}[\tilde\psi_1(x),\tilde\psi_0'(y)] - \frac{x}{M}\psi'_{-1}(x)\dot{\tilde\psi}_0'(y)\Big\} - M\tilde{H}_0 - \tfrac{1}{2}\int dx\ [\ddot{\tilde\psi}_0(x)\tilde\psi_0(x) + 2x\ddot{\tilde\psi}_0(x)\tilde\psi_0'(x) - 2\tilde\psi_0''(x)\tilde\psi^0(x)] + \int dx\ dy\ \psi'_{-1}(x)\psi'_{-1}(y)\Big\{-\frac{1}{M}\tilde\psi_0'(x)\tilde\psi_0'(y) - \frac{x}{M}\dot{\tilde\psi}_0(x)\dot{\tilde\psi}_0(y) + \frac{1}{i}[\tilde\psi_1(x),\dot{\tilde\psi}_1(y)]\Big\} = 0. \tag{6.17}$$

It is necessary to consider the commutation relations between $(\tilde\psi_0, \tilde\psi_1)$ and $(\tilde\psi_1, \dot{\tilde\psi}_1)$. From $\dot\beta_0 = 0$ [See Eqs. (5.29) and (4.30)] we have

$$\int dx\ \psi'_{-1}(x)\,\dot{\tilde\psi}_1(x) = -\int dx\ \dot{\tilde\psi}_0(x)\tilde\psi_0'(x) \tag{6.18}$$

and from the definition of α_0, we have [See Eqs. (4.25) and (4.32)]

$$\int dx\ \psi'_{-1}(x)\tilde\psi_1(x) = M\beta_0 + \tfrac{1}{2}\int dx\ x[\dot{\tilde\psi}_0(x)\dot{\tilde\psi}_0(x) - \tilde\psi_0(x)\ddot{\tilde\psi}_0(x)]. \tag{6.19}$$

Then the following commutation relations are obtained:

$$\int dx\ dy\ \psi'_{-1}(x)\psi'_{-1}(y)\frac{1}{i}[\tilde\psi_0'(x),\dot{\tilde\psi}_1(y)] = \int dx\ \psi''_{-1}(x)\tilde\psi_0'(x), \tag{6.20}$$

$$\int dx\ dy\ x\psi'_{-1}(x)\psi'_{-1}(y)\frac{1}{i}[\dot{\tilde\psi}_0(x),\dot{\tilde\psi}_1(y)] = -\int dx\ x\psi'_{-1}(x)\dot{\tilde\psi}_0'(x), \tag{6.21}$$

$$\int dx\ \psi'_{-1}(x)\frac{1}{i}[\tilde\psi_1(x),\dot{\tilde\psi}_0(y)] = \frac{M}{i}[\beta_0,\dot{\tilde\psi}_0(y)] - y\ddot{\tilde\psi}_0(y) - \tilde\psi_0'(y) + \frac{1}{M}\psi'_{-1}(y)\int dx\ \psi''_{-1}(x)\,\tilde\psi_0(x), \tag{6.22}$$

$$\int dx\ \psi'_{-1}(x)\frac{1}{i}[\tilde\psi_1(x),\tilde\psi_0(y)] = \frac{M}{i}[\beta_0,\tilde\psi_0(y)] - y\dot{\tilde\psi}_0(y) + \frac{1}{M}\psi'_{-1}(y)\int dx\ x\psi'_{-1}(x)\dot{\tilde\psi}_0(x), \tag{6.23}$$

and

$$\int dx\ dy\ \psi'_{-1}(x)\psi'_{-1}(y)\frac{1}{i}[\tilde{\psi}_1(x), \dot{\psi}_1(y)]$$

$$= \int dx\ [x\ \ddot{\tilde{\psi}}_0(x)\tilde{\psi}'_0(x) + \tilde{\psi}''_0(x)\tilde{\psi}_0(x) + \tfrac{1}{2}\dot{\tilde{\psi}}_0(x)\dot{\psi}_0(x)]$$

$$+\frac{1}{M}\int dx\ \psi'_{-1}(x)\psi'_{-1}(y)[\tilde{\psi}'_0(x)\ \tilde{\psi}'_0(y) + x\ \dot{\psi}_0(x)\dot{\tilde{\psi}}'_0(y)]. \tag{6.24}$$

When these equations are substituted in Eq. (6.17) and the definition (4.20) of $\tilde{H}_0$ is used, we arrive at the two conditions

$$\int dx\ \psi'_{-1}(x)\ [\beta_0, \dot{\tilde{\psi}}'_0(x)] = 0,$$

$$\int dx\ \psi'_{-1}(x)\ [\beta_0, \tilde{\psi}'_0(x)] = 0. \tag{6.25}$$

We also see that the assumption (6.16) is justified.

Now we turn our attention to the commutation relation between $\Psi_\lambda(t)$ and $\hat{\Pi}^f_\lambda(x)$

between $\Psi_\lambda(t)$ and $\Pi^f_\lambda(x)$

$$[\Psi_\lambda(t), \hat{\Pi}^f_\lambda(y)]_{t=t_y} = 0. \tag{6.26}$$

This leads to the equations

$$[QM, \hat{\Pi}_1(y)] + [\Psi_1, \dot{\tilde{\psi}}_0(y)] = 0, \tag{6.27}$$

$$[QM, \hat{\Pi}_2(y)] + [\Psi_1, \hat{\Pi}_1(y)]$$
$$+ [\Psi_2, \dot{\tilde{\psi}}_0(y)] + [QM, \dot{\tilde{\psi}}_0(y)]_2 = 0, \tag{6.28}$$

where Ψ_1 and Ψ_2 are given in Eqs. (6.5) and (6.6) and $\hat{\Pi}_i(y)$ is

$$\hat{\Pi}_i(y) = \int dz\ \mathscr{P}(y,z)\Pi_i(z), \tag{6.29}$$

$$\Pi^f_\lambda(x) = \sum_{n=0}^{\infty}\lambda^n\Pi_n(x), \tag{6.30}$$

with

$$\mathscr{P}(y,z) = \delta(y-z) - (1/M)\psi'_{-1}(y)\psi'_{-1}(z). \tag{6.31}$$

$\Pi_n(x)$ is given by Eq. (5.37). The last term in Eq. (6.28) appears because $[Q, \tilde{\psi}_0(x)] = 0(\lambda^2)$. The subscript 2 on the last term denotes the second order part of that commutator. Equation (6.27) reduces to

$$\int dy\ \mathscr{P}(z,y)\{\tilde{\psi}'_0(y) + y\ddot{\tilde{\psi}}_0(y)\}$$
$$+\int dx\ \psi'_{-1}(x)\frac{1}{i}[\tilde{\psi}_1(x), \dot{\tilde{\psi}}_0(z)] = 0, \tag{6.32}$$

which, using Eq. (6.22), gives the condition

$$[\beta_0, \dot{\tilde{\psi}}_0(z)] = 0. \tag{6.33}$$

Equation (6.28) can be reduced to

$$Q\int dy\ \mathscr{P}(z,y)\Big[\tilde{\psi}''_0(y) + y\ddot{\tilde{\psi}}'_0(y) + \ddot{\tilde{\psi}}_0(y) + \int dx\ \psi'_{-1}(x)\Big\{\frac{1}{i}[\tilde{\psi}'_1(x), \dot{\tilde{\psi}}_0(y)] + \frac{1}{M}\tilde{\psi}'_0(x)\psi''_{-1}(y) + \frac{1}{i}[\tilde{\psi}'_0(x), \dot{\tilde{\psi}}_1(y)]$$
$$+\frac{1}{i}[\tilde{\psi}_1(x), \dot{\tilde{\psi}}'_0(y)]\Big\}\Big] + \frac{\hat{P}}{M}\int dy\ \mathscr{P}(z,y)\Big[3y\ \dot{\tilde{\psi}}'_0(y) + \dot{\tilde{\psi}}_0(y) + y^2\ \dddot{\tilde{\psi}}_0(y) - t\ddot{\tilde{\psi}}_0(y) + \int dx\psi'_{-1}(x)\Big\{\frac{x}{i}[\dot{\tilde{\psi}}_1(x), \dot{\tilde{\psi}}_0(y)]$$
$$+\frac{x}{i}[\dot{\tilde{\psi}}_0(x), \dot{\tilde{\psi}}_1(y)] + \frac{1}{i}[\tilde{\psi}_1(x), \tilde{\psi}'_0(y)] + \frac{y}{i}[\tilde{\psi}_1(x), \ddot{\tilde{\psi}}_0(y)] - \frac{1}{M}x\dot{\tilde{\psi}}_0(x)\psi''_{-1}(y) - \frac{1}{M}x\tilde{\psi}'_{-1}(x)\dot{\tilde{\psi}}'_0(y)\Big\}\Big]$$
$$+\int dy\ \mathscr{P}(z,y)\Big[y\ \ddot{\tilde{\psi}}_1(y) + \tilde{\psi}'_1(y) + dx\ \int\psi'_{-1}(x)\Big\{\frac{1}{i}[\tilde{\psi}_2(x), \dot{\tilde{\psi}}_0(y)] + \frac{1}{M}\tilde{\psi}'_0(x)\tilde{\psi}'_0(y) + \frac{1}{M}\tilde{\psi}'_0(x)y\ddot{\tilde{\psi}}_0(y) - \frac{x}{M}\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}'_0(y)$$
$$+\frac{1}{i}[\tilde{\psi}_1(x), \dot{\tilde{\psi}}_1(y)]\Big\}\Big] + \frac{1}{i}[QM, \tilde{\psi}_0(z)]_2 = 0. \tag{6.34}$$

Using Eq. (6.22), it can be shown that the term proportional to Q in Eq. (6.34) reduces to

$$\int dy\ \mathscr{P}(z,y)\int dx\ \psi''_{-1}(x)\frac{1}{i}\{[\tilde{\psi}_1(x),\dot{\tilde{\psi}}_0(y)]$$
$$+[\tilde{\psi}_0(x), \dot{\tilde{\psi}}_1(y)]\} = 0. \tag{6.35}$$

Since $\hat{\psi}_1(x)$ is given by

$$\hat{\psi}_1(x) = \int d^2y\tilde{g}(x,y)F_2[\psi_{-1}(y)][\tilde{\psi}_0(y)]^2/2!, \tag{6.36}$$

Eq. (6.35) can be written as

$$\int dy\ \mathscr{P}(z,y)\int dx\ d^2u\ \psi''_{-1}(x)\,F_2[\psi_{-1}(u)]\tilde{\psi}_0(u)$$
$$\times\{\tilde{g}(x,u)\,\dot{\Delta}(u,y) + \dot{\tilde{g}}(y,u)\Delta(x,u)\} = 0, \tag{6.37}$$

where

$$i\Delta(x,y) = [\tilde{\psi}_0(x), \tilde{\psi}_0(y)]. \tag{6.38}$$

If we choose, say, a retarded Green's function

$$\tilde{g}(x,y) = \theta(x^0 - y^0)\Delta(x,y), \tag{6.39}$$

then Eq. (6.37) is satisfied immediately. By similar methods, the term of Eq. (6.34) which is proportional to $\hat{P}$ can be reduced to

$$\frac{\hat{P}}{M}\int dy\ \mathscr{P}(z,y)\Big\{-t\ddot{\tilde{\psi}}_0(y) + yM\frac{1}{i}[\beta_0, \ddot{\tilde{\psi}}_0(y)]$$
$$+M\frac{1}{i}[\beta_0, \tilde{\psi}'_0(y)]\Big\}. \tag{6.40}$$

This term can be made consistent by the following requirements:

$$[\beta_0, \dot{\tilde{\psi}}_0(y)] = 0 \tag{6.41}$$

and

$$[Q, \dot{\tilde{\psi}}_0(x)] = i\lambda^2 t(\hat{P}/M^2)\ddot{\tilde{\psi}}_0(x) + \cdots. \tag{6.42}$$

Equations (6.33) and (6.41) together with the fact that β_0 must be constructed from operators from the set $\{\alpha_i, \alpha_i^\dagger, \alpha_k, \alpha_k^\dagger\}$ imply that β_0 is a c number, which we set equal to zero.

In order to calculate the second and third terms on the left-hand side of Eq. (6.28) we need to evaluate the following commutator. Using Eqs. (4.37) and (4.40), we may write

$$\begin{aligned}\int dx\ \psi'_{-1}(x)\frac{1}{i}\{[\tilde{\psi}_2(x), \dot{\tilde{\psi}}_0(y)] + [\tilde{\psi}_1(x), \dot{\tilde{\psi}}_1(y)]\}\\ = \frac{1}{i}\int dx x\{[\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_0(x) - \tilde{\psi}_0(x)\ddot{\tilde{\psi}}_0(x), \tfrac{1}{2}\dot{\tilde{\psi}}_1(y)]\\ + \Big[\dot{\tilde{\psi}}_0(x)\dot{\tilde{\psi}}_1(x) - \frac{2}{3}\ddot{\tilde{\psi}}_0(x)\tilde{\psi}_1(x)\\ - \frac{1}{3}\tilde{\psi}_0(x)\ddot{\tilde{\psi}}_1(x), \dot{\tilde{\psi}}_0(y)\Big]\Big\}\\ - \frac{1}{3}\int dx\ [\tilde{\psi}_0'(x)\tilde{\psi}_1(x), \dot{\tilde{\psi}}_0(y)] + \frac{M}{i}[\beta_1, \dot{\tilde{\psi}}_0(y)].\end{aligned} \tag{6.43}$$

Equation (6.43) together with Eq. (6.36), (6.38), and (6.39) leads to

$$\begin{aligned}\int dx\ \psi'_{-1}(x)\frac{1}{i}\{[\tilde{\psi}_2(x), \dot{\tilde{\psi}}_0(y)] + [\tilde{\psi}_1(x), \dot{\tilde{\psi}}_1(y)]\}\\ = \dot{\tilde{\psi}}_0'(y)\frac{1}{M}\int dx\psi'_{-1}(x)\dot{\tilde{\psi}}_0(x) - \{y\ \ddot{\tilde{\psi}}_0(y) + \tilde{\psi}_0'(y)\}\\ \times\frac{1}{M}\int dx\psi'_{-1}(x)\tilde{\psi}_0'(x) - y\ddot{\tilde{\psi}}_1(y) - \tilde{\psi}_1'(y)\\ + \frac{M}{i}[\beta_1, \dot{\tilde{\psi}}_0(y)],\end{aligned} \tag{6.44}$$

which together with the last term of Eq. (6.34) leads to

$$[\beta_1, \dot{\tilde{\psi}}_0(y)] = 0. \tag{6.45}$$

We have thus shown that Eqs. (6.27) and (6.28) lead to the requirements of Eqs. (6.33), (6.41), (6.42), and (6.45).

The commutation relation between $\hat{\psi}^f(x)$ and $\Pi(t)$ is

$$[\hat{\psi}^f_\lambda(x), \Pi_\lambda(t)] = 0, \tag{6.46}$$

which leads to the relations

$$[\tilde{\psi}_0(x), \Pi_1] + [\hat{\psi}_1(x), \hat{P}] = 0, \tag{6.47}$$

$$[\tilde{\psi}_0(x), \Pi_2] + [\hat{\psi}_1(x), \Pi_1] + [\hat{\psi}_2(x), \hat{P}] = 0. \tag{6.48}$$

Here Π_1 and Π_2 are given in Eqs. (6.7) and (6.8) and

$$\hat{\psi}_\lambda(x) = \int dy\ \mathscr{P}(x, y)\psi_\lambda(y).$$

[See Eqs. (6.2) and (6.31)]

Equation (6.47) can immediately be shown to be satisfied. Equation (6.48) can be reduced to

$$\begin{aligned}\int dz\ \mathscr{P}(x,z)\Big\{Q\Big\{\tilde{\psi}_0''(z) + \int dy\ \psi'_{-1}(y)\Big(-\frac{1}{M}\psi''_{-1}(z)\tilde{\psi}_0'(y)\\ - i[\tilde{\psi}_0'(z), \dot{\tilde{\psi}}_1(y)] - i[\tilde{\psi}_0(z), \dot{\tilde{\psi}}_1'(y)]\\ - i[\tilde{\psi}_1(z), \dot{\tilde{\psi}}_0'(y)]\Big)\Big\} + \frac{P}{M}\Big\{z\dot{\tilde{\psi}}_0'(z)\\ + \int dy\psi'_{-1}(y)\Big(-\frac{1}{2M}z\psi'_{-1}(z)\dot{\tilde{\psi}}_0'(y)\\ - zi[\tilde{\psi}_0(z), \ddot{\tilde{\psi}}_1(y)] - yi[\tilde{\psi}_0(z), \ddot{\tilde{\psi}}_1(y)]\\ - i[\tilde{\psi}_0(z), \tilde{\psi}_1'(y)] + i[\tilde{\psi}_1(z), \tilde{\psi}_0'(y)]\Big)\Big\} + \Big\{\tilde{\psi}_1'(z)\\ + \int dy\ \psi'_{-1}(y)\Big(-\frac{1}{M}\tilde{\psi}_0'(z)\tilde{\psi}_0'(y) - \frac{1}{M}z\dot{\tilde{\psi}}_0(z)\dot{\tilde{\psi}}_0'(y)\\ - i[\tilde{\psi}_0(z), \dot{\tilde{\psi}}_2(y)] - i[\tilde{\psi}_1(z), \dot{\tilde{\psi}}_1(y)]\Big)\Big\}\Big\} = 0.\end{aligned} \tag{6.49}$$

the terms proportional to Q in Eq. (6.49) are easily shown to cancel. Using the identity

$$\begin{aligned}\int dy\ y\psi'_{-1}(y)\frac{1}{i}[\tilde{\psi}_0(z), \ddot{\tilde{\psi}}_1(y)]\\ = 2\int dy\psi''_{-1}(y)\frac{1}{i}[\tilde{\psi}_0(z), \tilde{\psi}_1(y)]),\end{aligned} \tag{6.50}$$

the terms proportional to $\hat{P}$ in Eq. (6.49) can also be easily shown to cancel. The remaining expression can then be shown to be identically zero. In this way, Eq. (6.46) is satisfied.

Finally, from Eq. (6.1) it is possible to derive the identity

$$[\hat{\psi}^f(x), \hat{\Pi}^f(y)] = i\mathscr{P}(x, y). \tag{6.51}$$

It is possible but tedious to show that this relation is satisfied. The details are not written here as they are very long and yield no new information. In summary, we have shown that the commutation relation of Eq. (6.1) requires the following conditions:

$$[\beta_1, \dot{\tilde{\psi}}_0(y)] = 0, \tag{6.52}$$

$$\beta_0 = 0, \tag{6.53}$$

$$[Q, \dot{\tilde{\psi}}_0(x)] = i\lambda^2 t(\hat{P}/M^2)\ddot{\tilde{\psi}}_0(x) + \cdots. \tag{6.54}$$

Since $\dot{\beta}_1 = 0$, Eq. (6.52) leads to

$$\beta_1 = 0. \tag{6.55}$$

From Eqs. (5.24) and (5.42), we can derive

$$[\dot{Q}, \dot{\tilde{\psi}}_0(x)] = i\lambda^2(\hat{P}/M^2)\ddot{\tilde{\psi}}_0(x) + \cdots \tag{6.56}$$

and

$$[\dot{Q}, \dot{\tilde{\psi}}_0(x)] = i\lambda^2(\hat{P}/M^2)\ddot{\tilde{\psi}}_0(x) + \cdots. \tag{6.57}$$

Equations (6.54) and (6.57) lead to

$$[\ddot{Q}, \dot{\tilde{\psi}}(x)] = i\lambda^2 t(\hat{P}/m^2)\dddot{\tilde{\psi}}_0(x) + \cdots \tag{6.58}$$

which using the field equation for $\tilde{\psi}_0(x)$ and the fact that $\tilde{\psi}_0(x)$ is orthogonal to $\psi'_{-1}(x)$, leads to

$$[Q, \tilde{\psi}_0(x)] = i\lambda^2 t\frac{\hat{P}}{M^2}\dot{\tilde{\psi}}_0(x) + \cdots. \tag{6.59}$$

From the above results, we can see that the set of physical operators $\{\hat{P}, Q, \tilde{\psi}_0(x), \dot{\tilde{\psi}}_0(x)\}$ obey the algebra

$$[Q, \hat{P}] = i \tag{6.60}$$

$$[\tilde{\psi}_0(x), \dot{\tilde{\psi}}_0(y)]_{x^0=y^0} = i\{\delta(x-y) - \psi'_{-1}(x)\psi'_{-1}(y)/M\}, \tag{6.61}$$

$$[Q(t), \tilde{\psi}_0(x)]_{t=x^0} = i\lambda^2 t(\hat{P}/M^2)\dot{\tilde{\psi}}_0(x) + \ldots, \tag{6.62}$$

$$[Q(t), \dot{\tilde{\psi}}_0(x)]_{t=x^0} = i\lambda^2 t(\hat{P}/M^2)\ddot{\tilde{\psi}}_0(x) + \ldots, \tag{6.63}$$

where all other equal-time commutation relations vanish. If we construct the Schrödinger picture operator, q, by

$$q = e^{-iHt}Q(t)e^{iHt} = Q(t) - t\dot{Q}, \tag{6.64}$$

then Eqs. (6.52) and (6.56) lead to

$$[q, \tilde{\psi}_0(x)] = 0 \tag{6.65}$$

and

$$[q, \dot{\tilde{\psi}}_0(x)] = 0. \tag{6.66}$$

Also, since $[\hat{P}, H] = 0$,

$$[q, \hat{P}] = i. \tag{6.67}$$

Thus we see that the set of operators $\{\hat{P}, q\}$ commute with the set $\{\tilde{\psi}_0(x), \dot{\tilde{\psi}}_0(x)\}$. The Hilbert space of the system is now constructed in the following way: take the direct product of the quantum-mechanical realization of the operators $\{q, \hat{P} = (1/i)(\partial/\partial q)\}$ in the Schrödinger picture and the Fock representation of the operators $\{\tilde{\psi}_0(x), \tilde{\psi}_0(x)\}$ constructed cyclically by operations of the creation operators $\{\alpha_i^\dagger, \alpha_k^\dagger\}$ on the vacuum defined by $\alpha_i|0\rangle = \alpha_k|0\rangle = 0$.

VII. DISCUSSION AND CONCLUSIONS

In this paper we have thus found a consistent quantum-field-theoretical picture of a quantum system with an extended object. In this picture we have constructed the quantal Hilbert space and performed a perturbative calculation of the dynamical map. In the course of the calculation the zero-mode problem was avoided by using certain integral relations. Also, ambiguities arising in the choice of boundary conditions were resolved by requiring that the Heisenberg fields satisfy the canonical commutation relations.

We have seen that, to the order of approximation considered, the presence of q and P in the dynamical map induces only the following replacements of the explicit space-time variables:

$$x \to x + q + (P/M)t + (P^2/M^2)(x+q) - (P/M^2)\tilde{H}_0 + \cdots,$$
$$t \to t + (P/M)(x+q) + (P^2/2!M^2)t + \cdots,$$

which are the lower terms of a Taylor expansion of X and T given in Eqs. (2.38) and (2.39). The quantum coordinate q and the spatial coordinate x always appear in the combination $(x+q)$.

The operator quantity $\psi_2(x)$ describes particle-particle and particle-extended object scattering. When a particle scatters from the extended object, the object recoils. That is, its velocity $\dot{Q}$ changes. However, the total momentum of the system is conserved. Therefore $\dot{Q}$ depends on both the total momentum and the particle number. This is seen in Eq. (5.24). This gives rise to the noncommutativity of Q with the physical field $\psi_0(x)$ seen in Eqs. (6.62) and (6.63). The perturbative results for $\dot{Q}$ and H are consistent with the expansions of the expressions

$$\dot{Q} = P/[P^2 + (M+H_0)^2]^{1/2} \text{ or}$$
$$P = [(M+H_0)\dot{Q}]/(1-\dot{Q}^2)^{1/2},$$
$$H = [P^2 + (M+H_0)^2]^{1/2} \text{ or}$$
$$H = (M+H_0)/(1-\dot{Q}^2)^{1/2},$$

which can also be derived by general arguments using Lorentz symmetry.[11] The extended objects' position can, however, be specified without interference from the particle number at one particular time. This necessitates the use of the Schrödinger picture in describing the quantum state of the extended object. The quantum expectation value of the position at any given time can be found using Eq. (6.64).

We have shown in the tree approximation how the presence of the quantum coordinate leads simply to a replacement of the space-time variables (t, x) by some combination of these and the quantum-mechanical operators. This makes contact with the collective coordinate method. The constraint and gauge condition used in that method may correspond to the choice of boundary conditions in our formalism. However, this comparison needs further study.

All results presented here are in the tree approximation. A method for inserting the quantum corrections was partially given in Refs. 8 and 12. However, a more careful treatment of the ordering of the operators Q, $\dot{Q}$, $\tilde{\psi}_0(x)$, and $\dot{\tilde{\psi}}_0(x)$ appears to be necessary. Work in this direction is in progress.

Certain $(1+1)$-dimensional classical nonlinear field theories, such as the sine-Gordon theory, are exactly integrable and possess an infinite number of conservation laws. These theories have been studied in the collective coordinate formalism.[4] An analysis of the role of the conservation laws in these theories in the formalism presented in this paper would be interesting.

ACKNOWLEDGMENTS

This work was supported by the Natural Sciences and Engineering Research Council, Canada and the Faculty of Science, University of Alberta.

[1] For a review see, for example, A. C. Scott, F. Chu, and D. McLaughlin, Proc. IEEE **61**, 1443 (1973).

[2] For reviews and further references, see R. Rajaraman, Phys. Rep. C **21**, 227 (1975); J. L. Gervais and A. Neveu, Phys. Rep. C **23**, 237 (1976); R. Jackiw, Rev. Mod. Phys. **49**, 681 (1977). H. Umezawa and H. Matsumoto, *Symmetries in Science*, edited by B. Gruber and R. S. Millman (Plenum, New York, 1980), p. 411.

[3] G. Wentzel, Helv. Phys. Acta **13**, 269 (1940); D. Bohm and D. Pines, Phys. Rev. **92**, 609 (1953); J. L. Gervais and A. Neveu, Phys. Rev. D **11**, 2943 (1975); C. G. Callan Jr. and D. J. Gross, Nucl. Phys. B **93**, 29 (1975); N. M. Christ and T. D. Lee, Phys. Rev. D **12**, 1606 (1975); E. Tomboulis, Phys. Rev. D **12**, 1678 (1975); A. Hosoya and K. Kikkawa, Nucl. Phys. B **101**, 271 (1975); J. L. Gervais and G. Jevicki, Nucl. Phys. B **110**, 93 (1975); V. E. Korepin, P. P. Kulish, and L. D. Fadeev, JETP Lett. **21**, 138 (1975); E. Tomboulis and G. Woo, Ann. Phys. **98**, 1 (1976); J. L. Gervais, Acta Phys. Austriaca Suppl. XVIII, 385 (1977); C. L. Hammer, J. E. Shrauner, and B. De Facio, Phys. Rev. D **19**, 667 (1979).

[4] J. L. Gervais, A. Jevicki, and B. Sakita, Phys. Rev. D **12**, 1038 (1975); V. E. Korepin and L. D. Fadeev, Theor. Math. Phys. **25**, 1039 (1975); J. L. Gervais and A. Jevicki, Nucl. Phys. B **110**, 113 (1976); H. de Vega, Nucl. Phys. B **115**, 411 (1976).

[5] L. Leplae, F. Mancini, and H. Umezawa, Phys. Rev. B **2**, 3594 (1970); H. Matsumoto, N. J. Papastamatiou, and H. Umezawa, Nucl. Phys. B **82**, 45 (1974); **97**, 90 (1975).

[6] H. Matsumoto, P. Sodano, and H. Umezawa, Phys. Rev. D **19**, 511 (1979);

G. Oberlechner, M. Umezawa, and Ch. Zenses, Lett. Nuovo Cimento **23**, 641 (1978).
[7]H. Matsumoto, G. Oberlechner, H. Umezawa, and M. Umezawa, J. Math. Phys. **20**, 2088 (1979).
[8]H. Matsumoto, G. Semenoff, H. Umezawa, and M. Umezawa, J. Math. Phys. **21**, 1761 (1980).
[9]H. Matsumoto, N. J. Papastamatiou, H. Umezawa, and M. Umezawa, Phys. Rev. D **23**, 1339 (1981).
[10]The usual definition of the momentum is Eq. (4.1) with a minus sign. The change of the sign in P leads to the change of the sign in Q, replacing $(x + Q)$ by $(x - Q)$.
[11]H. Matsumoto, N. J. Papastamatiou, G. Semenoff, and H. Umezawa, "The Asymptotic Condition and the Hamiltonian in Quantum Field Theory with Extended Objects," Phys. Rev. D **24**, 406 (1981).
[12]G. Semenoff, H. Matsumoto, and H. Umezawa, Prog. Theor. Phys. **63**, 1393 (1980).

Incompressibility in relativistic continuous media

J. J. Ferrando and J. Olivert
Departamento de Mecánica y Astronomia, Facultad de Matemáticas, Burjassot (València), Spain

(Received 18 February 1981; accepted for publication 24 April 1981)

A formalism of continuous media in general relativity is given and the concept of shock waves is defined using manifolds with boundary and the transversality theory of submanifolds. A definition of relativistic incompressibility is proposed and one gets that the shock waves are longitudinal with the speed of light.

PACS numbers: 04.20. – q, 04.40. + c

1. INTRODUCTION

In this paper we propose to study the concept of incompressibility in General Relativity. This topic has been studied for perfect fluids (Eddington,[1] Lichnerowicz,[2] B. Coll,[3] Olivert[4]), but it has not been treated with whole generality.

Our aim is to extend the concept of dynamic volume given in[3,4] and require that its spatial projection does not vary along the stream lines. Then we shall examine the relativistic waves in continuous media.

When one deals with extended bodies in General Relativity, it is usual to avoid possible definitions of such bodies. Nevertheless, in the scientific literature, one finds some suggestions in order to structure them; for instance, Cattaneo[5] with his "ideal state," Choquet-Bruhat and Lamoureux-Brousse[6] with the "frame state," Beiglböck,[7] and so on.

In this paper, we start with the idea of scheme given by Lichnerowicz.[8] We introduce the concept of local scheme which, intersecting with the Landau manifold of an observer,[9] allows a possible definition of a relativistic continuous medium.

According to the notations used in this paper, let us consider the space–time as a connected pseudo-Riemannian manifold X, of Hausdorff type and four dimensions, in which a metric tensor field of hyperbolic signature (3,1) has been defined. The coordinate x^4 of a point, belonging to a coordinate neighborhood of a given local chart F, coincides with the coordinate time in the system described by F. It is assumed that the measure units have been chosen such that the constant light velocity in the vacuum has the value 1.

We notice that the indices represented by latin letters take values 1 to 4 and the Greeck indices are restricted the the values 1, 2, 3.

Finally, we represent by T_pX the tangent space of X at p, and by ϕ_{*m} the derived linear function at $m \in X$ of a differentiable function ϕ defined in X. We write $M \pitchfork N$ to say that the two submanifolds M and N are transverse.

2. LOCAL DOMAINS AND SIMULTANEITY IN THE SPACE-TIME

Let D be a domain of the space–time X. Being D an open manifold, D is a locally compact and connected manifold with boundary V. This manifold of dimension 4 is called a local domain.

The study of the local domain in the Minkowski space allows us to give, in Theorem 2.1, a structure for certain sets of simultaneous points, which are subsets of a domain D.

A local domain $\mathbf{V}$ of Minkowski space $\mathbf{X}$ is called a transversally spatial domain if its boundary $\partial \mathbf{V}$ cuts transversally[10] with the whole spacelike hyperplane which contains a point of $\mathrm{Int}\mathbf{V}$ (interior of $\mathbf{V}$).

Lemma 2.1: For each $\mathbf{p}$ of an open set $\mathbf{D}$ of Minkowski space, there is a transversally spatial and convex local domain $\mathbf{V} \subset \mathbf{D}$ such that $\mathbf{p} \in \mathrm{Int}\mathbf{V}$.

Proof: It is easy to prove that a closed ball of center $\mathbf{p}$ is a local domain satisfying the adequate conditions.

Lemma 2.2: Let $\mathbf{D}$ be an open set and π a spacelike hyperplane of Minkowski space $\mathbf{X}$. Then there is a local domain $\mathbf{V} \subset \mathbf{D}$ such that $\mathbf{M} = \mathbf{V} \cap \pi$ is a compact and convex manifold with boundary of 3 dimensions immersed in $\mathbf{R}^3$.

Proof: As the dimensions of $\mathbf{V}$ and $\mathbf{X}$ are the same,

$$T_x\mathbf{V} + T_x\pi = T_x\mathbf{X}, \quad \forall x \in \pi \cap \mathbf{V}.$$

That is,

$$\mathbf{V} \pitchfork \pi. \tag{1.1}$$

Therefore, as $\mathbf{V}$ is transversally spatial, π spacelike, and $\mathbf{p} \in \mathrm{Int}\mathbf{V} \cap \pi$, we get

$$\partial \mathbf{V} \pitchfork \pi. \tag{1.2}$$

From (2.1) and (2.2), one derives[10] that $\mathbf{M} = \mathbf{V} \cap \pi$ is a manifold with boundary of dimension 3.

Finally, as $\mathbf{V}$ is compact and π is closed, $\mathbf{M}$ is compact into $\pi \simeq \mathbf{R}^3$ and it is connected, inasmuch as $\mathbf{V}$ and π are convex.

Theorem 2.1: Let D be a domain and C an observer (a differentiable timelike curve) of the space–time X. For each $p \in C$ such that $L_p \cap D \neq \phi$ (L_p being the Landau manifold of p), there is a local domain $V \subset D$ so that $M_p = L_p \cap V$ is a compact connected Riemannian manifold with boundary of dimension 3.

Proof: Let $p \in C$ be such that $L_p \cap D \neq \phi$. As the space–time is a linear connected manifold, at each point $p \in X$ there exist a neighborhood B_p diffeomorphic to a neighborhood of $0 \in T_pX = \mathbf{X}$ by the map $\exp_p$. Then $\mathbf{D}_p = \exp_p^{-1}(B_p \cap D)$ is an open set of Minkowski space which satisfies

$$\mathbf{E}_p \cap \mathbf{D}_p = \exp_p^{-1}(L_p \cap D) \neq \phi,$$

where $\mathbf{E}_p$ is the physical space of C at p. $\mathbf{E}_p$ is a spacelike hyperplane.

Then we can apply Lemma 2.2. There is a local domain $\mathbf{V} \subset \mathbf{D}_p$ so that $\mathbf{M}_p = \mathbf{V} \cap \mathbf{E}_p$ is a compact connected Riemannian manifold with boundary of 3 dimensions.

Let us consider the local domain $V = \exp_p \mathbf{V} \subset D$. Then $M_p = V \cap L_p$ is a compact connected manifold with boundary of 3 dimensions, as it is the image of $\mathbf{M}_p$ by $\exp_p$.

Therefore, we can restrict M_p to the positive definite metric of the Landau manifold: $\bar{\gamma}^0 = i^*\bar{g}$, where i is the natural injection of L_p in X.

3. ALMOST-THERMODYNAMIC MATERIAL SCHEMES

Following Lichernowicz,[8] a scheme is a domain D of the space–time in which one has defined a tensor field of second order which is called the energy-momentum tensor $\bar{T}$. We say that D is normal is $\bar{T}$ is symmetric and has a timelike eigenvector; that is, there exist a scalar field r and a unitary timelike vector field $\bar{u}$ so that

$$\bar{T}(\bar{u}) = -r\bar{u}. \qquad (3.1)$$

Let us now solve the following question: As a physical body with a material distribution has a density, is it possible to endow a material density to a point set of a scheme D (i. e., the M_p of Theorem 2.1)? In this case, M_p would generalize the concept of continuous medium; we take it as a compact connected manifold with boundary of 3 dimensions immersed in $\mathbf{R}^3$, on which a scalar field $\sigma > 0$ is defined as the density of medium.

Theorem 3.1: Let D be a normal scheme with energy–momentum tensor $\bar{T}$ such that $r > 0$; and C an observer of space–time. For each $p \epsilon C$ such that $L_p \cap D \neq \phi$, there is a local scheme V (local domain of a scheme) of D so that $M_p = L_p \cap V$ is a compact connected Riemannian manifold endowed with a scalar field $\sigma > 0$.

Proof: Due to Theorem 2.1, it only remains to show the existence of the field σ. But $r = \bar{T}(\bar{u},\bar{u}) > 0$, by hypothesis. Then, if i is the injection from M_p in D,

$$\sigma = i^*r = r \circ i > 0$$

is a positive scalar field on M_p.

The schemes which satisfy the hypotheses of Theorem 3.1 with unique unitary timelike eigenvector $\bar{u}$ are called material schemes and the sets M_p are relativistic continuous media (diffeomorphic, on the other hand, to a continuous medium, according to the construction made in Theorem 2.1). The scalar field r is the so-called proper mass–energy density of the scheme. The tensor field $\bar{\tau} = i^*\bar{T}$ is a symmetric tensor field of second order on M_p which generalizes the classic stress tensor. The stream lines of the scheme are integral curves of the field $\bar{u}$. This field, which by definition of material scheme is unique, is a so-called 4-velocity of the scheme.

The relativistic stress tensor $\bar{t}$ is defined as the spatial projection (by means of Eckart tensor $\bar{\gamma} = \bar{g} + \bar{u} \otimes \bar{u}$) of the energy–momentum tensor. The spatial components of $\bar{t}$ in the local inertial proper system of point $p \in D$ (event of a stream line or observer) coincides at p with the components of $\bar{\tau}$ in the chart induced in M_p.

From the definition of $\bar{t}$, one derives that the energy–momentum tensor takes the expression

$$\bar{T} = \rho(1+\epsilon)\bar{u} \otimes \bar{u} + \bar{t}, \qquad (3.2)$$

where we have adopted the Taub hypothesis[11] over the decomposition of the proper mass–energy density in two terms: ρ is the mass–energy density and ϵ is the specific internal energy. This introduction of the thermodynamic variable ϵ leads us to say that an almost-thermodynamic material scheme is such that its energy–momentum tensor takes the expression (3.2).

Let us assume that there is matter conservation in an almost-thermodynamic material scheme, a condition which can be expressed by the continuity equation

$$\nabla_i(\rho u^i) = 0 \Leftrightarrow u^i \nabla_i \rho + d_i^{\,i} = 0, \qquad (3.3)$$

where $\bar{d} = \frac{1}{2} L_{\bar{u}} \bar{\gamma}$ is the deformation tensor of the scheme and $L_{\bar{u}}$ is the Lie derivative. The tensor $\bar{d}$ has the components

$$d_{ij} = 1/2\,[\gamma_i^k \nabla_k u_j + \gamma_j^k \nabla_k u_i]. \qquad (3.4)$$

The time and space components of the conservation equations $\nabla_i T_j^{\,i} = 0$ can be written as

$$\rho u^i \nabla_i \epsilon + t^{ij} d_{ij} = 0, \qquad (3.5)$$

$$\rho f_j^r u^k \nabla_k u^j + \gamma_i^r \gamma_j^k \nabla_k t^{ij} = 0, \qquad (3.6)$$

where $\bar{f} = (1+\epsilon)\bar{\gamma} + (1/\rho)\bar{t}$ is the tensor index[12] of the scheme (below we shall give its physical significance).

4. INCOMPRESSIBILITY OF AN ALMOST-THERMODYNAMIC MATERIAL SCHEME

Let us recall that in perfect fluids one has characterized the incompressibility for isoentropic processes by

$$\nabla_{\bar{u}} k = 0, \qquad (4.1)$$

k being the dynamic volume defined by

$$k = f/\rho \qquad (4.2)$$

and f the fluid index[2] which, at once, is given as a function of the pressure, the specific internal energy, and the material density according to the expression

$$f = 1 + \epsilon + p/\rho. \qquad (4.3)$$

The generalization of these concepts is suggested by the definition of the tensor index, which is given by Maugin[12] for elastic media, and we extend it to almost-thermodynamic material schemes:

$$\bar{f} = (1+\epsilon)\bar{\gamma} + (1/\rho)\bar{t}. \qquad (4.4)$$

From (4.4), we propose the following.

Definition 4.1: We call dynamic volume of an almost-thermodynamic material scheme to the 2-tensor

$$\bar{k} = v\bar{f},$$

where v is the specific volume.

From this definition and (4.4), one derives

$$\bar{k} = v\bar{\gamma} + v\bar{e} \qquad (4.5)$$

and $\bar{e} = \epsilon\bar{\gamma} + v\bar{t}$ is called the tensor enthalpy, according to the analogy with classical thermodynamics.

Formula (4.5) hints that the dynamic volume is the sum of a term which stands for the specific volume and another term, a function of the enthalpy, which is called enthalpic volume. This enthalpic volume reduces to zero in classical physics ($c \to \infty$), as one can prove if we correct the above mentioned formulas in order to make them homogeneous.

Definition 4.2: We say that an almost-thermodynamic material scheme D is incompressible if the spatial change of dynamic volume along the stream lines is null, i. e.,

$$\gamma_i^r \gamma_j^l \nabla_{\bar{u}} k^{ij} = 0. \tag{4.6}$$

Let us see that this idea of incompressibility generalizes the one which is given by (4.1) for perfect fluids. For them, the relativistic stress tensor is given by $\bar{t} = p\bar{\gamma}$; therefore, Eq. (4.4) can be written as $\bar{f} = f\bar{\gamma}$, and therefore

$$\bar{k} = k\bar{\gamma}. \tag{4.7}$$

Transporting (4.7) in the definition of incompressibility (4.6), one gets

$$\gamma^{rl} \nabla_{\bar{u}} k = 0, \tag{4.8}$$

hence (4.1) is obviously satisfied.

Theorem 4.1: The condition of incompressibility (4.6) leads to the expression

$$\gamma_i^r \gamma_j^l u^k \nabla_k t^{ij} + (2t^{rl} + \rho(1+\epsilon)\gamma^{rl})\gamma_i^k \nabla_k u^i$$
$$= \gamma^{rl} t^{ki} d_{ki}.$$

Proof: If we develop the covariant derivative of the dynamic volume using $v = 1/\rho$, we obtain

$$\nabla_{\bar{u}} k^{ij} = 1/\rho^2 \{\rho\gamma^{ij} \nabla_{\bar{u}} \epsilon + \rho(1+\epsilon)(u^i \nabla_{\bar{u}} u^j + u^j \nabla_{\bar{u}} u^i)$$
$$+ \nabla_{\bar{u}} t^{ij} - ([2/\rho] t^{ij} + (1+\epsilon)\gamma^{ij}) \nabla_{\bar{u}} \rho\}. \tag{4.9}$$

From this expression, by virtue of (4.6), one derives

$$\rho\gamma^{rl} u^k \nabla_k \epsilon + \gamma_i^r \gamma_j^l u^k \nabla_k t^{ij}$$
$$- ([2/\rho] t^{rl} + (1+\epsilon)\gamma^{rl}) u^k \nabla_k \rho = 0. \tag{4.10}$$

Using Eqs. (3.3) and (3.5), we eliminate the derivatives of ϵ and ρ, and the theorem is proved.

5. RELATIVISTIC WAVES IN THE INCOMPRESSIBLE ALMOST-THERMODYNAMIC SCHEMES

In order to study the relativistic waves for incompressible almost-thermodynamic schemes, we introduce some properties of the timelike characteristic hypersurfaces.

Theorem 5.1: Let W be a timelike hypersurface of a material scheme D, and p a point of D such that $L_p \cap D \neq \phi$. Then $H_p = W \cap L_p$ is a regular submanifold of dimension two.

Proof: It is easy to see that the tangent vectors to L_p are spacelike. Then, as W is timelike, one has that, in each $q \epsilon W \cap L_p$, there is a timelike vector tangent to W which will not be tangent to L_p. This, together with the fact that W and L_p are submanifolds of three dimensions and D of four dimensions, leads to the transversality condition

$$T_q L_p + T_q W = T_q D$$

and it tells us that H_p is a regular submanifold of dimension two.

Definition: We call shock waves of a timelike characteristic hypersurface W, of an observer C, the submanifolds

$$H_p = L_p \cap W, \quad p \in C.$$

If B_p is the domain of $\exp_p^{-1}$, then $\mathbf{W}_p = \exp_p^{-1}(W \cap B_p)$ is a hypersurface into the Minkowski space $T_p X$ associated with W at p. Thereby, we can consider the set

$$\mathbf{H}_p = \mathbf{W}_p \cap \mathbf{E}_p = \exp_p^{-1} H_p, \tag{5.1}$$

which is a submanifold of dimension two in $T_p X$, and it will be called a "spacelike shock wave" of the observer C.

Once the shock waves are defined, let us introduce their speed. This leads us to the well-known expression given by Lichnerowicz[2,8] and Pham Mau Quan.[13] Hereby we proceed to derive from the definitions some properties which attempt to fill a vacuum existing in this respect in the scientific literature.

Let p be a point of timelike hypersurface W and C the stream line of the scheme which contains p. In order to associate a speed to the shock waves $H_p, p \in C$, let us take a vector $\bar{w}$ tangent to W at p orthogonal to H_p. Let us see that $\bar{w}$ is timelike. As

$$T_p H_p = \mathbf{E}_p \cap T_p W, \tag{5.2}$$

then $\{\bar{h}_1, \bar{h}_2, \bar{w}\}$ will be a basis of $T_p W$ if $\{\bar{h}_1, \bar{h}_2\}$ is a basis of $T_p H_p$, because $\bar{w}$ is orthogonal to H_p. Therefore, $\bar{w}$ is timelike, inasmuch as W. If we consider that $\bar{w}$ is unitary, it is called a four-velocity of H_p.

By virtue of (5.2), there is a unitary vector $\bar{\lambda} \in \mathbf{E}_p$ orthogonal to H_p. This vector gives us the direction of the observed propagation. Hereafter we shall express $T_p H_p$ by $\mathbf{H}_\lambda$.

Theorem 5.2: The vectors $\bar{l}$ (orthogonal to W at p), $\bar{w}$, $\bar{\lambda}$, and $\bar{u}$ (four-velocity of the observer at p) are in the same two-plane; that is, $\bar{l}$ and $\bar{w}$ can be written as linear combinations of $\bar{\lambda}$ and $\bar{u}$.

Proof: Let us define the vector

$$\bar{L} = \bar{l} - a\bar{u} \quad \text{with} \quad a = -\bar{g}(\bar{u}, \bar{l}). \tag{5.3}$$

This vector belongs to the physical space $\mathbf{E}_p$, inasmuch as

$$\bar{g}(\bar{u}, \bar{L}) = 0. \tag{5.4}$$

But by virtue of (5.2), all $\bar{v} \in \mathbf{H}_\lambda$ is in $\mathbf{E}_p$ and in $T_p W$; hence it will be orthogonal to $\bar{L}$. Thus, $\bar{L}$ has the direction of $\bar{\lambda}$. This implies

$$\bar{l} = \|\bar{L}\| \bar{\lambda} + a\bar{u}. \tag{5.5}$$

With $b = -\bar{g}(\bar{w}, \bar{u})$, $\bar{w} - b\bar{u}$ belongs to the physical space $\mathbf{E}_p$. Moreover, as $\bar{w}$ is orthogonal to H_p, it will be orthogonal to all vectors of $\mathbf{H}_\lambda$, and so

$$\bar{g}(\bar{v}, \bar{w} - b\bar{u}) = 0, \quad \forall \bar{v} \in \mathbf{H}_\lambda.$$

That is, if $\bar{w} - b\bar{u} \neq 0$, it will have the direction of $\bar{\lambda}$ and there will be a $c \in R$ such that $c\bar{\lambda} = \bar{w} - b\bar{u}$, and hence

$$\bar{w} = b\bar{u} + c\bar{\lambda}. \tag{5.6}$$

If $\bar{w} - b\bar{u} = 0$, it is enough to take $c = 0$ so that (5.6) is satisfied.

The expressions (5.5) and (5.6) prove the theorem.

The speed $\mathbf{U}$ of propagation of the shock wave is the relative in the Minkowski space between the moving point of four-velocity $\bar{\mathbf{w}}$ and the observer of four-velocity $\bar{\mathbf{u}}$, where

$$\bar{\mathbf{u}} = (\exp_p^{-1})_* \bar{u}, \quad \bar{\mathbf{w}} = (\exp_p^{-1})_* \bar{w}. \tag{5.7}$$

As $\exp_p$ conserves the metric at p,

$$b = -w^i u_i = -\mathbf{w}^i \mathbf{u}_i = 1/(1 - \mathbf{U}^2)^{1/2}. \tag{5.8}$$

From (5.6) and (5.8), we get

$$c^2 = \mathbf{U}^2/(1 - \mathbf{U}^2). \tag{5.9}$$

On the other hand, as $\bar{w}$ and $\bar{l}$ are orthogonal, the expressions (5.5) and (5.6) lead to

$$a/\|\bar{L}\| = c/b. \tag{5.10}$$

Substituting (5.8) and (5.9) into (5.10), we arrive at

$$\mathbf{U} = a/\|\bar{L}\|. \tag{5.11}$$

[We have taken the positive root in (5.9). One would arrive at the same final result taking the negative root.]

Now, as W is a characteristic hypersurface, we may apply the Hadamard discontinuities, so that if ϕ is one of the tensor components, of any type, with discontinuities at the first derivatives, one satisfies[14]

$$[\phi,_i] = \delta\phi l_i. \tag{5.12}$$

As the Christoffel symbols are continuous, one gets

$$[\nabla_i\phi] = \delta\phi l_i. \tag{5.13}$$

Furthermore, using (5.5), we obtain

$$[\gamma_j^i\nabla_i\phi] = \|\bar{L}\|\lambda_j\delta\phi, \tag{5.14}$$

and keeping in mind (5.3) and (5.11), we derive

$$[u^i\nabla_i\phi] = -\mathbf{U}\|\bar{L}\|\delta\phi. \tag{5.15}$$

It is easy to verify that the components $\delta\phi$ yield to the lowering and raising of indices, because $\bar{g}$ is continuous in X.

If the W is a characteristic hypersurface by the discontinuities of the first derivatives of the four-velocity, by virtue of (5.12) it leads us for each component to

$$[\partial_j u^i] = \delta u^i l_j. \tag{5.16}$$

With theses discontinuities, we define the vector $\delta\bar{u} = (\delta u^i)$, which will be called "infinitesimal discontinuity of the four-velocity $\bar{u}$".

It is easy to prove that

$$u_i\delta u^i = 0, \tag{5.17}$$

and hence $\delta\bar{u}$ is a vector of the physical space.

With the decomposition made by Maugin,[12]

$$\delta\bar{u} = \delta\bar{u}_\perp + \bar{\lambda}\delta u_\|, \tag{5.18}$$

where $\delta u_\| = \lambda_i\delta u^i$, $\delta u_\perp^i = S_j^i\delta u^j$, $\bar{S} = \bar{\gamma} - \bar{\lambda}\otimes\bar{\lambda}$, we come to formulate the following

Theorem 5.3: Let $\bar{P}$ be a spacelike symmetric tensor of second order ($P_{ij}u^i = 0$). The discontinuity system of $\delta\bar{u}$,

$$P_{ri}\delta u^i = 0, \tag{5.19}$$

may be written in the form

$$P_{\perp ri}\delta u_\perp^i + p_r\delta u_\| = 0, \tag{5.20}$$

$$p_i\delta u_\perp^i + P_\|\delta u_\| = 0, \tag{5.21}$$

$\bar{P}_\perp$, $\bar{p}$, $P_\|$ are defined as

$$P_{\perp ri} = S_r^k S_i^l P_{kl}, \tag{5.22}$$

$$P_\| = P_{kl}\lambda^k\lambda^l, \tag{5.23}$$

$$p_i = \lambda^k P_{kl} S_i^l. \tag{5.24}$$

Proof: The use of (5.22) leads to

$$\bar{P} = \bar{P}_\perp + \bar{p}\otimes\bar{\lambda} + \lambda\otimes\bar{p} + P_\|\bar{\lambda}\otimes\bar{\lambda} \tag{5.25}$$

and from (5.24) to

$$p_i = P_{ik}\lambda^k - P_\|\lambda_i. \tag{5.26}$$

Placing (5.25) and (5.18) into (5.19), one gets

$$P_{ri}\delta u_\perp^i + p_r\delta u_\| + \lambda_r p_i\delta u^i + P_\|\lambda_r\delta u_\| = 0. \tag{5.27}$$

The substitution of (5.26) into (5.27) leads to (5.20). Finally, by means of the contraction of (5.19), one obtains (5.21).

Let us study the relativistic waves in the incompressible almost-thermodynamic scheme. For it, we analyze the equations by virtue of which the Cauchy problem is established.

As we are not going to study the gravitational waves, we assume that the second derivatives of g_{ij} are continuous, and thus the use of either the Einstein field equations or material waves will not be necessary, so that $\mathbf{U}\neq 0$.

Infinitesimal discontinuities can exist at the derivatives of the four-velocity components, of the relativistic stress tensor, of the material mass density, and of the specific internal energy; i. e.,

$$\delta u^i,\ \delta t_{ij},\ \delta\rho,\ \delta\epsilon. \tag{5.28}$$

Nevertheless, from the conservation equations (3.5) and (3.6), the continuity equation (3.3), and the equation given by Theorem 4.1, one has that not all these discontinuites are independent; but they satisfy the following.

Theorem 5.4: The study of discontinuities (5.28) reduces to the system

$$P_{ri}u^i = 0,$$

with

$$P_{ri} = \rho\mathbf{U}^2 f_{ri} + \lambda_r\lambda_k t_i^k - (2t_r^k\lambda_k + \rho(1+\epsilon)\lambda_r)\lambda_i.$$

Proof: Calculating discontinuities in the formula of Theorem 4.1, keeping in mind (5.14), (5.15), and (3.4), one has

$$-\gamma_i^r\gamma_j^l\mathbf{U}\delta t^{ij} = \gamma^{rl}t_i^k\lambda_k - (2t^{rl} + \rho(1+\epsilon)\gamma^{rl})\lambda_i\delta u^i, \tag{5.29}$$

and from (3.6) we obtain

$$-\rho f_i^r\mathbf{U}\delta u^i + \gamma_i^r\lambda_j\delta t^{ij} = 0. \tag{5.30}$$

Equations (5.29) and (5.30) lead us to the statement of the theorem.

Let us restrict the principal shock waves. We use the following definition given by Maugin[12]: We say that a shock wave is principal if $\bar{\lambda}$ (propagation direction) is the eigenvector of the relativistic stress tensor $\bar{t}$:

$$t_{ij}\lambda^j = t_\|\lambda_i. \tag{5.31}$$

Theorem 5.5: In a principal shock wave, one satisfies

$$P_{\perp ri}\delta u_\perp^i = 0, \quad P_\|\delta u_\| = 0, \tag{5.32}$$

where

$$P_{\perp ri} = \rho\mathbf{U}^2 S_r^s S_i^j f_{sj}, \tag{5.33}$$

$$P_\| = (\mathbf{U}^2 - 1)(\rho(1+\epsilon) + t_\|). \tag{5.34}$$

Proof: Due to (5.31), $\bar{P}$ is symmetric and takes the form

$$\bar{P} = \rho\mathbf{U}^2\bar{f} - (\rho(1+\epsilon) + t_\|)\bar{\lambda}\otimes\bar{\lambda}. \tag{5.35}$$

Then, we may apply Theorem 5.3 to Theorem 5.4, and hence one obtains straightforwardly the expressions (5.32), because of

$$p_j = S_j^i P_{ri}\lambda^r = 0$$

by virtue of (5.35). Further, (5.33) and (5.34) are derived from the definition given in Theorem 5.3.

Theorem 5.6: In an incompressible almost-thermodynamic scheme the principal shock waves are longitudinal and their speeds are $\mathbf{U} = 1$.

Proof: Due to the uniqueness of the four-velocity of a material scheme

$$\rho(1+\epsilon)+t_\alpha \neq 0, \quad \alpha = 1,2,3. \tag{5.36}$$

From (5.32) and (5.33) one derives, by virtue of $\mathbf{U}\neq 0$, $\delta\bar{u}_\perp \in \mathbf{H}_\lambda$,

$$\rho f_{ri}\delta u_\perp^i = 0. \tag{5.37}$$

But from (5.36), the system (5.37) leads to $\delta\bar{u}_\perp = 0$; that is, to the longitudinal shock waves

$$\delta\bar{u} = \delta u_\parallel \bar{\lambda}, \quad \delta u_\parallel \neq 0.$$

This condition carries us with (5.32) to

$$P_\parallel = 0. \tag{5.38}$$

Finally, the expression (5.34) of $P_\parallel$, with (5.36) and (5.38), implies the unity value for the speed $\mathbf{U}$ of shock waves.

6. DISCUSSION

In this paper we have studied what we can understand as a relativistic continuous medium in General Relativity. For it, we have made use of the theory of manifolds with boundary and use them to satisfy some topologic properties in order to verify the integrability conditions in these manifolds. The integration is essential to definition in these continuous media; for instance, their mass, center-of-mass, and so on.

Nevertheless, our exposition must be considered as a first order approximation to the reality, because the boundaries of the natural extended bodies are not always differentiable at all points. It is suggestive to extend our definition to the manifolds with corners; but these mathematical structures have not been deeply studied so very few properties are known to be applied to this topic.

Once the material schemes have been defined, we have restricted the discussion to the so-called almost-thermodynamic material schemes, in order to avoid introducing in this paper the concepts of enthropy and temperature.

We have obtained an incompressibility condition for the almost-thermodynamic schemes, following the line labelled by Lichnerowicz.[2] With it, the transversal shock waves disappear (at most, they are material: $\mathbf{U}=0$), and the longitudinal shock waves speed takes the value 1.

Therefore, if the almost-thermodynamic scheme is a perfect fluid, the incompressibility condition reduces to the known one in B. Coll[3] and J. Olivert[4] for isoenthropic processes.

Likewise, it is easy to prove that the incompressibility condition, in the hypoelastic media,[12] leads us to consistent results:

$$\rho(\lambda+2\mu)+2t_\parallel = \rho(1+\epsilon), \tag{6.1}$$

$$\rho\mu + t_\alpha = 0, \ \alpha = 2,3, \tag{6.2}$$

where λ, μ are the Lamé coefficients. With it, one gets the same results for the speeds for shock waves as this paper.

The expressions (6.1) and (6.2) are obtained establishing the equations for the Hadamard discontinuity described in Sec. 5, by replacing the space component of conservation equation (3.6) by the relativistic Hooke law given by Choquet-Bruhat and Lamoureux-Brouse,[6] equivalent to the one proposed by Maugin[12] when the continuity equation is satisfied.

ACKNOWLEDGMENTS

We would like to thank Professor A. M. Naveira and Dr. A. Montesinos for suggestions related to the mathematical content of this work, and especially Professor J. Bernabeu for his care in reading the manuscript and suggestions about its physical content.

[1] A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge U. P., Cambridge, 1923).

[2] A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).

[3] B. Coll, " Fronts de combustion en magnétodydrodynamique relativiste," Ann. Inst. Henri Poincaré **25**, 363-91 (1967).

[4] J. Olivert, "Estudio relativista de un fluido perfecto termodinámico de masa material variable en procesos reversibles," Col. Math. **28**, Fasc. 2, (1976)

[5] M. C. Cattaneo, "Essai d'une théorie relativiste de l'élasticité," C. R. Acad. Sci. Ser. A **272**, 1421 (1971).

[6] Y. Choquet-Bruhat and L. Lamoureux-Brouse, "Sur les équations de l'élasticité relativiste," C. R. Acad. Sci. Ser. A **276**, 1317 (1976).

[7] W. Beiglböck, "The Center-of-mass in Einsteins Theory of Gravitation," Commun. Math. Phys. **5**, 106-130 (1967).

[8] A. Lichnerowicz, *Theories relativistes de la gravitation et de l'électromagnétisme* (Masson, Paris, 1955).

[9] J. Olivert," On the local simultaneity in general ralativity," J. Math. Phys. **21**, 1783 (1980).

[10] V. Guillemin and A. Pollack, *Differential Topology* (Prentice-Hall, Englewood Cliffs, New Jersey, 1974).

[11] A. H. Taub, in *Hydrodynamics and General Relativity*, edited by R. Wasserman and Ch. P. Well, (Academic, New York, 1963).

[12] G. A. Maugin, "Ondes infinitésimales principales dans lesmilieux élastiques précontraints en relativité," C. R. Acad. Sci. Ser. A **284**, 393 (1977).

[13] Phan Mau Quan, " Sur les équations des fluides chargés inductifs en relativité général," Rend. Mat. Roma **6**, 2 (1969).

[14] V. D. Zakharov, *Gravitational waves in Einstein's theory* (Israel Progressive Scientific Translations, Jerusalem-London, 1973).

Classical solutions of the equations of supergravity

Robert J. Finkelstein and Jaegu Kim
Department of Physics, University of California, Los Angeles, California 90024

(Received 13 February 1981; accepted for publication 15 May 1981)

We describe an algorithm to solve the classical equations of supergravity for the coupled fields of helicity 2 and 3/2. The algorithm depends on an expansion in the associated Grassman algebra and leads to a sequence of coupled equations that may be solved in a step by step manner. The procedure begins with solutions of the usual empty space Einstein equations but the subsequent equations are all linear differential equations. To complete the method, it is necessary to generalize the supplementary conditions on the vector spinor field from flat to curved space. The algorithm also permits a classification of the complete solutions in terms of the associated gravitational fields. It is shown that vanishing curvature implies Minkowski space and does not permit other conceivable possibilities such as a Clifford space. Generalizations of the vanishing curvature space are suggested.

PACS numbers: 04.20.Cv, 04.20.Jb

I. INTRODUCTION

Almost nothing is now known about classical solutions of the equations of supergravity. In fact the only exact solution, to our knowledge, is a generalization of the parallel plane waves that solve the vacuum Einstein equations[1]; this is a "true" solution in the sense that it cannot be reduced by a gauge transformation to a solution of the simple gravity equations. We are here interested in searching for a wider class of exact solutions of the supergravity equations that also have the essential property that they cannot be reduced to simple gravity by a gauge transformation.

The additional structure of supergravity stems not only from the coupling of the 3/2 field to the gravitational field but also from the anticommuting character of the fermion field. It turns out that the anticommuting property is a simplifying feature, since an expansion in the related Grassmann algebra reveals that the field equations decompose into a set of coupled equations that may be solved sequentially rather than simultaneously. The algorithm for the solution begins with the usual Einstein equation with no source and a corresponding Rarita–Schwinger equation for the 3/2 field in the same background gravitational field that solves the Einstein equation. The only nonlinear equation that needs to be solved is the Einstein equation itself, while the rest of the coupled set of equations are all linear in the unknown functions. Since the algorithm always starts from a solution of the empty space Einstein equations, the solutions of the supergravity equations may themselves be classified by a classification of the parent empty space gravitational fields.

II. NOTATION AND FIELD EQUATIONS

We introduce the following 1-forms[1]

$$E^a = E^a{}_\mu dx^\mu,$$

$$\omega^a{}_b = \omega_\mu{}^a{}_b dx^\mu,$$

$$\psi = \psi_\mu dx^\mu,$$

$$\gamma = \gamma_\mu dx^\mu,$$

where $E^a{}_\mu$, $\omega_\mu{}^a{}_b$, ψ_μ, and γ_μ are the tetrad field, coefficients of rotation, Rarita–Schwinger field, and Dirac matrices, respectively. In terms of these we define torsion and curvature 2-forms

$$S^a = dE^a + \omega^a{}_b \wedge E^b, \tag{2.1}$$

$$\Theta^a{}_b = d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b. \tag{2.2}$$

These equations imply

$$dS^a + \omega^a{}_b \wedge S^b = \Theta^a{}_b \wedge E^b. \tag{2.3}$$

In this notation the field equations are

$$\gamma \wedge D\psi = 0, \tag{A}$$

$$*\Theta^a{}_b \wedge E^b = -(i/2)\psi^T C\gamma^a\gamma^5 \wedge D\psi, \tag{B}$$

$$S^a = (i/2)\psi^T C\gamma^a \wedge \psi, \tag{C}$$

where

$$D = d + \tfrac{1}{2}\omega_{ab}\sigma^{ab}\wedge, \tag{2.4}$$

and $*\Theta$ is the dual curvature.

In addition, the vector–spinor field is subject to the supplementary conditions

$$\psi_\alpha = C\bar{\psi}_\alpha{}^T, \tag{D}$$

$$(\psi_\alpha, \psi_\beta)_+ = 0. \tag{E}$$

Urrutia has noted that if (C) is substituted into (2.3), one finds

$$\Theta^a{}_b \wedge E^b = -(i/2)\psi^T C\gamma^a \wedge D\psi, \tag{*B}$$

which is dual to (B). In this way supergravity preserves the usual duality between the field equations and the cyclic identities of simple gravity.[2]

The component forms of (A) and (B) are

$$\epsilon^{\mu\lambda\alpha\beta}\gamma_\lambda D_\alpha \gamma^5 \psi_\beta = 0, \tag{A'}$$

$$(R^\mu{}_a - \tfrac{1}{2}RE_a{}^\mu)e = -(i/2)\epsilon^{\mu\lambda\alpha\beta}\psi_\lambda{}^T C\gamma_a D_\alpha \gamma^5\psi_\beta, \tag{B'}$$

where $R^\mu{}_a$ is the Ricci contraction of the curvature and e is the determinant $|E^a{}_\mu\cdot|$. It follows from (A') and (B') that

$$R = 0 \tag{2.5}$$

III. EXPANSION IN GRASSMANN ALGEBRA

Let us make the preceding formulation more precise by assuming that all the above forms lie in a Grassmann alge-

 0022-2488/81/102228-07$01.00

bra. Then for the vector–spinor form we assume

$$\psi = \sum \overset{1}{\psi}_i \epsilon_i + \sum \overset{3}{\psi}_{ijk} \epsilon_i \epsilon_j \epsilon_k + \cdots,$$
$$= \sum \overset{1}{\psi}_{\mu i} \epsilon_i dx^\mu + \sum \overset{3}{\psi}_{\mu ijk} \epsilon_i \epsilon_j \epsilon_k dx^\mu + \cdots, \qquad (3.1)$$

where

$$\epsilon_i \epsilon_j + \epsilon_j \epsilon_i = 0. \qquad (3.2)$$

Then the commuting forms E^a, $\omega^a{}_b$, S^a, and $\Theta^a{}_b$ may be expanded in even elements of the algebra. Let us first consider the simplest case:

$$\psi_\mu = \psi_{\mu 1} \epsilon_1 + \psi_{\mu 2} \epsilon_2. \qquad (3.3)$$

Then

$$E^a = \overset{0}{E}{}^a + \overset{2}{E}{}^a (i\epsilon_1 \epsilon_2), \qquad (3.4)$$

$$\omega^a{}_b = \overset{0}{\omega}{}^a{}_b + \overset{2}{\omega}{}^a{}_b (i\epsilon_1 \epsilon_2), \qquad (3.5)$$

$$S^a = \overset{2}{S}{}^a (i\epsilon_1 \epsilon_2), \qquad (3.6)$$

$$\Theta^a{}_b = \overset{0}{\Theta}{}^a{}_b + \overset{2}{\Theta}{}^a{}_b (i\epsilon_1 \epsilon_2). \qquad (3.7)$$

Equations (A), (B), and $\star$(B) now imply in lowest order of the Grassmann algebra

$$\overset{0}{\gamma} \wedge \overset{0}{D} \psi = 0, \qquad (3.8)$$

$$\star \overset{0}{\Theta}{}^a{}_b \wedge \overset{0}{E}{}^b = 0, \qquad (3.9)$$

$$\overset{0}{\Theta}{}^a{}_b \wedge \overset{0}{E}{}^b = 0, \qquad (3.10)$$

where

$$\overset{0}{\gamma} = \gamma_a \overset{0}{E}{}^a, \qquad (3.11)$$

$$\overset{0}{D} = d + \tfrac{1}{2} \overset{0}{\omega}_{ab} \sigma^{ab} \wedge . \qquad (3.12)$$

Equations (3.9) and (3.10) may be expressed in the familiar forms

$$\overset{0}{R}{}^\alpha{}_\beta - \tfrac{1}{2} \overset{0}{R} \delta^\alpha{}_\beta = \overset{0}{R}{}^\alpha{}_\beta = 0 \qquad (3.9')$$

by (2.5), and

$$\overset{0}{R}{}^\lambda{}_{\alpha\beta\gamma} + \overset{0}{R}{}^\lambda{}_{\beta\gamma\alpha} + \overset{0}{R}{}^\lambda{}_{\gamma\alpha\beta} = 0, \qquad (3.10')$$

where $\overset{0}{R}{}^\alpha{}_\beta$ is the Ricci contraction of the curvature

$$\overset{0}{R}{}^\alpha{}_\beta = \overset{0}{\Theta}{}^{\lambda\alpha}{}_{\lambda\beta}.$$

The equations (3.9′) are differential equations for the $\overset{0}{\omega}_\mu{}^a{}_b$ and (3.10′) are identities. By (2.1) the forms $\overset{0}{\omega}{}^a{}_b$ may be related to the $\overset{0}{E}{}^a$:

$$d\overset{0}{E}{}^a + \overset{0}{\omega}{}^a{}_b \wedge \overset{0}{E}{}^b = 0. \qquad (3.13)$$

The connection between $\overset{0}{E}$ and $\overset{0}{\omega}$ is thus Riemannian. These equations may be solved for the $\overset{0}{\omega}_{\lambda\rho\mu}$:

$$\overset{0}{\omega}_{\lambda\mu\rho} = \tfrac{1}{2}\Big[-\overset{0}{E}_{a\lambda}\Big(\partial_\mu \overset{0}{E}{}^a{}_\rho - \partial_\rho \overset{0}{E}{}^a{}_\mu\Big) + \overset{0}{E}_{a\mu}\Big(\partial_\rho \overset{0}{E}{}^a{}_\lambda - \partial_\lambda \overset{0}{E}{}^a{}_\rho\Big) + \overset{0}{E}_{a\rho}\Big(\partial_\lambda \overset{0}{E}{}^a{}_\mu - \partial_\mu \overset{0}{E}{}^a{}_\lambda\Big)\Big]. \qquad (3.14)$$

After substitution of (3.14) into (3.9′) the set (3.9′) contains 10 equations for the $16\overset{0}{E}{}^a{}_\alpha$. The remaining six degrees of freedom correspond to the arbitrariness in orienting the tetrad.

The important simplification in (3.8) is that the torsion does not appear in $\overset{0}{D}$ and therefore the equation for ψ is linear. That fact not only makes the solution of (3.8) tractable but also permits one to build up linear superpositions to satisfy additional constraints. Equation (3.8) is, of course, the usual Rarita–Schwinger equation in a given gravitational field $\overset{0}{\omega}_{\mu ab}$, where $\overset{0}{\omega}_{\mu ab}$ is a solution of the field equations (3.9′). Therefore, the general problem defined by (3.8) and (3.9) is to determine ψ in a given gravitational background. This has in fact already been carried out for a class of solutions in the Kerr–Newman background.[3]

The fact that the lowest order equations (3.8)–(3.10) split off from the remaining equations implies also that the supergravity solutions may be partially classified with the aid of the associated solution $\overset{0}{\omega}$ of the simple gravity equation. For example, one might classify the gravitational components according to the Pirani–Petrov scheme, or in some other way such as the Bianchi classification. The supergravity solutions would then divide into classes distinguished by the label of the parent gravitational class.

IV. THE SECOND ORDER EQUATIONS

Equations (A), (B), and (B′) then imply in the next order

$$\star \overset{0}{\Theta}{}^a{}_b \wedge \overset{2}{E}{}^b + \star \overset{2}{\Theta}{}^a{}_b \wedge \overset{0}{E}{}^b = -(i/2)\psi^T C \gamma^a \overset{0}{D} \gamma^5 \psi, \qquad (4.1)$$

$$\overset{0}{\Theta}{}^a{}_b \wedge \overset{2}{E}{}^b + \overset{2}{\Theta}{}^a{}_b \wedge \overset{0}{E}{}^b = -(i/2)\psi^T C \gamma^a \wedge \overset{0}{D} \psi. \qquad (4.2)$$

Let

$$\star \overset{0}{\Theta}{}^a{}_b \wedge \overset{2}{E}{}^b = \Psi^a; \qquad (4.3)$$

then

$$\Big(\overset{0}{R}{}^\mu{}_a - \tfrac{1}{2} \overset{0}{R}\, \overset{0}{E}{}_a{}^\mu\Big) \overset{2}{e} = -\tfrac{1}{2} \epsilon^{\mu\alpha\beta\gamma} \Psi_{a\alpha\beta\gamma}, \qquad (4.4)$$

where $\overset{2}{e}$ is the following determinant:

$$\overset{2}{e} \epsilon_{\alpha\beta\lambda\mu} = \epsilon_{abcd} \overset{0}{E}{}^a{}_\alpha \overset{0}{E}{}^b{}_\beta \overset{0}{E}{}^c{}_\lambda \overset{2}{E}{}^d{}_\mu. \qquad (4.5)$$

By (3.9′) however the left side of (4.4) vanishes and therefore Ψ^a also vanishes. Hence the first term of (4.1) drops out and (4.1) becomes

$$\star\overset{2}{\Theta}{}^a{}_b \wedge \overset{0}{E}{}^b = -(i/2)\psi^T C\gamma^a \wedge \overset{0}{D}\gamma^5\psi \tag{4.6}$$

or

$$\overset{2}{R}{}^\mu{}_a = -(i/2)\overset{0}{e}{}^{-1}\epsilon^{\mu\lambda\alpha\beta}(\psi_\lambda{}^T C\gamma_a \overset{0}{D}_\alpha\gamma^5\psi_\beta), \tag{4.6'}$$

where $\overset{2}{R}$ is the Ricci contraction of $\overset{2}{\Theta}$ and (2.5) has again been used.

In contrast to the first term of (4.1) the first term of (4.2) is not required to vanish. This term then brings the second order tetrad field $\overset{2}{E}{}^a$ into the problem. That is because (4.2) incorporates the second order part of (2.1).

One now has to solve (4.6) for the second order tetrad $\overset{2}{E}{}^a$ where

$$\overset{2}{\Theta}{}^a{}_b = d\overset{2}{\omega}{}^a{}_b + \overset{2}{\omega}{}^a{}_c \wedge \overset{0}{\omega}{}^c{}_b + \overset{0}{\omega}{}^a{}_c \wedge \overset{2}{\omega}{}^c{}_b \tag{4.7}$$

and $\overset{2}{\omega}$ is related to $\overset{2}{E}$ by

$$d\overset{2}{E}{}^a + \overset{0}{\omega}{}^a{}_b \wedge \overset{2}{E}{}^b + \overset{2}{\omega}{}^a{}_b \wedge \overset{0}{E}{}^b = \overset{2}{S}{}^a. \tag{4.8}$$

This last equation may be solved for the components of $\overset{2}{\omega}$:

$$\overset{2}{\omega}_{\mu\alpha\lambda} = \tfrac{1}{2}[\{\mu\alpha\lambda\} + \{\lambda\alpha\mu\} + \{\alpha\mu\lambda\}], \tag{4.9}$$

where

$$\{\alpha\mu\lambda\} = \overset{2}{S}_{\alpha\mu\lambda} - \overset{0}{E}_{a\alpha}\left(\partial_\mu \overset{2}{E}{}^a{}_\lambda - \partial_\lambda \overset{2}{E}{}^a{}_\mu\right) - \left(\overset{0}{\omega}_{\mu ab}\overset{2}{E}{}^b{}_\lambda - \overset{0}{\omega}_{\lambda ab}\overset{2}{E}{}^b{}_\mu\right). \tag{4.9a}$$

After (4.7) is substituted into (4.6) the resulting equations are linear in $\overset{2}{\omega}$, while all other functions (ψ, $\overset{0}{\omega}$, $\overset{0}{E}$) in these equations are already determined by the lowest order problem. In addition according to (4.9) the $\overset{2}{\omega}_{\alpha\mu\lambda}$ are themselves linear functions of the $\overset{2}{E}{}^a{}_\mu$ and its first derivatives. Therefore, after substitution of (4.9) into (4.6), (4.6) becomes a set of linear second order differential equations for the $\overset{2}{E}{}^a{}_\lambda$. There are 16 equations for the 16 unknown functions $\overset{2}{E}{}^a{}_\lambda$. All other functions in these equations are again known from the solution to the lowest order problem.

The problem always divides into two tractable parts. In the first part one finds a ψ_μ field in a background gravitational field, which itself is an arbitrary solution of the empty space Einstein equations. Having determined the $\overset{0}{E}$, $\overset{0}{\omega}$, and ψ in this way, one can then go into the second order equation and solve 16 linear partial differential equations for the 16 functions $\overset{2}{E}{}^a{}_\alpha$.

Since this procedure incorporates (2.1) as well as the structure of the torsion (C) via (4.8) or (4.9) it is guaranteed to preserve the identities ($\star$B) and (4.2). The exact solution obtained in this way has the property that ψ_μ and $E^a{}_\alpha$ are mutually codetermined but the expansion in the Grassmann algebra permits a procedure that separates the total problem into two linear problems which can be solved in sequence rather than simultaneously.

V. CONSTRAINTS ON THE VECTOR SPINOR FIELD

In Minkowski space the irreducible representation of the spin-3/2 field is totally symmetric in the spinor indices. The irreducibility condition may also be expressed by imposing

$$\gamma^\mu\psi_\mu = 0. \tag{5.1}$$

Since the γ^μ are independent of position in flat space, (5.1) reduces the Rarita–Schwinger equations to the following set of four Dirac equations:

$$\gamma^\mu\partial_\mu\psi_\lambda = 0. \tag{5.2}$$

Then contraction of (5.2) with γ^λ yields

$$\partial^\mu\psi_\mu = 0 \tag{5.3}$$

so that there are altogether eight conditions, (5.1) and (5.3), exclusive of the Majorana constraints, that limit the solutions of 16 equations (5.2). That leaves eight independent components, the proper number for a massive 3/2 field. In addition the massless equation is invariant under the gauge transformation

$$\psi_\mu{}' = \psi_\mu + \partial_\mu\epsilon. \tag{5.4}$$

In this case the number of independent components may be reduced to four.

Let us now see how these conditions generalize to solutions of (3.8). We begin as before by imposing

$$\overset{0}{\gamma}{}^\mu\psi_\mu = 0. \tag{5.5}$$

To pass from flat to curved space, we need the formula

$$\left(\overset{0}{\gamma}{}^\mu, \overset{0}{D}_\lambda\right) = \{{}^{\mu}_{\sigma\lambda}\}\overset{0}{\gamma}{}^\sigma \tag{5.6}$$

and the Rarita–Schwinger equation in the following form:

$$\overset{0}{\gamma}{}^\mu\left(\overset{0}{D}_\mu\psi_\lambda - \overset{0}{D}_\lambda\psi_\mu\right) = 0. \tag{5.7}$$

It follows from (5.5), (5.6), and (5.7) that

$$\overset{0}{\gamma}{}^\mu\overset{0}{\nabla}_\mu\psi_\lambda = 0, \tag{5.8}$$

where

$$\overset{0}{\nabla}_\mu\psi_\lambda = \overset{0}{D}_\mu\psi_\lambda - \{{}^{\sigma}_{\lambda\mu}\}\psi_\sigma. \tag{5.9}$$

Then $\overset{0}{\nabla}_\mu$ is the "complete" covariant derivative and (5.8) is the analog of (5.2).

Now contract (5.8) with $\overset{0}{\gamma}{}^\lambda$. Then

$$\left[\tfrac{1}{2}\left(\overset{0}{\gamma}{}^\lambda, \overset{0}{\gamma}{}^\mu\right)_+ + \tfrac{1}{2}\left(\overset{0}{\gamma}{}^\lambda, \overset{0}{\gamma}{}^\mu\right)_-\right]\overset{0}{\nabla}_\mu\psi_\lambda = 0,$$

$$\overset{0}{\nabla}{}^{\mu}\psi_{\mu}+\tfrac{1}{2}\overset{0}{\gamma}{}^{\lambda}\overset{0}{\gamma}{}^{\mu}\Big(\overset{0}{\nabla}_{\mu}\psi_{\lambda}-\overset{0}{\nabla}_{\lambda}\psi_{\mu}\Big)=0$$

and by (5.7)

$$\overset{0}{\nabla}{}^{\mu}\psi_{\mu}=0 \tag{5.10}$$

which is the analog of (5.3). Just as in flat space the divergence condition (5.10) is a consequence of the preceding two conditions (5.5) and (5.8). Therefore, one has the same number (eight) of independent components as in the flat case. Moreover, these three constraints are invariant under the following gauge transformation:

$$\psi_{\lambda}{}'=\psi_{\lambda}+\overset{0}{D}_{\lambda}\epsilon, \tag{5.11}$$

where ϵ satisfies the Dirac equation

$$\overset{0}{\gamma}{}^{\lambda}\overset{0}{D}_{\lambda}\epsilon=0. \tag{5.12}$$

Proof: The preceding statement is immediately obvious for the first of these constraints, (5.5) in view of (5.12).

Next consider

$$\begin{aligned}
\overset{0}{\gamma}{}^{\mu}\overset{0}{\nabla}_{\mu}\psi_{\lambda}{}' &= \overset{0}{\gamma}{}^{\mu}\overset{0}{\nabla}_{\mu}\overset{0}{D}_{\lambda}\epsilon\\
&= \overset{0}{\gamma}{}^{\mu}\overset{0}{D}_{\mu}\overset{0}{D}_{\lambda}\epsilon-\overset{0}{\gamma}{}^{\mu}\{{}^{\sigma}_{\lambda\mu}\}\overset{0}{D}_{\sigma}\epsilon\\
&= \overset{0}{\gamma}{}^{\mu}\overset{0}{D}_{\lambda}\overset{0}{D}_{\mu}\epsilon+\overset{0}{\gamma}{}^{\mu}\Big(\overset{0}{D}_{\mu},\overset{0}{D}_{\lambda}\Big)\epsilon-\overset{0}{\gamma}{}^{\mu}\{{}^{\sigma}_{\lambda\mu}\}\overset{0}{D}_{\sigma}\epsilon\\
&= \overset{0}{D}_{\lambda}\overset{0}{\gamma}{}^{\mu}\overset{0}{D}_{\mu}\epsilon+\overset{0}{\gamma}{}^{\mu}\Big(\overset{0}{D}_{\mu},\overset{0}{D}_{\lambda}\Big)\epsilon \qquad \text{by (5.6)}\\
&= \overset{0}{\gamma}{}^{\mu}\Big(\overset{0}{D}_{\mu},\overset{0}{D}_{\lambda}\Big)\epsilon
\end{aligned} \tag{5.13}$$

again by (5.12). Then

$$\overset{0}{\gamma}{}^{\mu}\Big(\overset{0}{D}_{\mu},\overset{0}{D}_{\lambda}\Big)\epsilon=\overset{0}{\gamma}{}^{\mu}\Big(\tfrac{1}{2}\overset{0}{R}_{ab\mu\lambda}\sigma^{ab}\Big)\epsilon \tag{5.14}$$

$$\begin{aligned}
&=\tfrac{1}{4}\overset{0}{R}_{\alpha\beta\mu\lambda}\overset{0}{\gamma}{}^{\mu}\overset{0}{\gamma}{}^{\alpha}\overset{0}{\gamma}{}^{\beta}\epsilon\\
&=\tfrac{1}{4}\overset{0}{R}_{\alpha\beta\mu\lambda}\Big[g^{\mu\alpha}\overset{0}{\gamma}{}^{\beta}-g^{\mu\beta}\overset{0}{\gamma}{}^{\alpha}+g^{\alpha\beta}\overset{0}{\gamma}{}^{\mu}+i\epsilon^{\mu\alpha\beta\sigma}\gamma_{5}\overset{0}{\gamma}_{\sigma}\Big]\epsilon\\
&=0
\end{aligned} \tag{5.15}$$

by (3.9′) and (3.10′). Then

$$\overset{0}{\gamma}{}^{\mu}\overset{0}{\nabla}_{\mu}\psi_{\lambda}{}'=0 \tag{5.16}$$

so that (5.8) is invariant under (5.11). But if (5.5) and (5.8) are invariant so also is (5.10) since it follows from the two earlier equations.

VI. GENERALIZATION TO ARBITRARY GRASSMANN ALGEBRA

The limitation to the simplest Grassmann algebra is inessential, and the preceding procedure goes through exactly the same for a Grassmann algebra of dimension 2^{d}. The lowest order problem is again to determine ψ in a given gravitational background which is in turn a solution of the empty space Einstein equations. One may then proceed to higher order equations as before. In going from the nth order to the $(n+2)$ order the solution of the nth order problem may be assumed to be known. Then there are always only linear equations to be solved for the terms of order $n+2$.

One may illustrate the general procedure by going to the next order. In the third order Eq. (A) is

$$\overset{0}{\gamma}\wedge\overset{0}{D}\overset{3}{\psi}+\overset{2}{\gamma}\wedge\overset{0}{D}\overset{1}{\psi}+\overset{0}{\gamma}\wedge\overset{2}{D}\overset{1}{\psi}=0. \tag{6.1}$$

This is a linear equation for $\overset{3}{\psi}$. Everything else in this equation is known from the second order calculation. After $\overset{3}{\psi}$ is determined from (6.1), one may go on to the fourth order forms of Eq. (B):

$$\begin{aligned}
&\star\overset{0}{\Theta}{}^{a}{}_{b}\wedge\overset{4}{E}{}^{b}+\star\overset{2}{\Theta}{}^{a}{}_{b}\wedge\overset{2}{E}{}^{b}+\star\overset{4}{\Theta}{}^{a}{}_{b}\wedge\overset{0}{E}{}^{b}\\
&=-(i/2)\Big(\overset{1}{\psi}{}^{T}C\overset{2}{\gamma}{}^{a}\wedge\overset{1}{D}\psi+\overset{1}{\psi}{}^{T}C\overset{0}{\gamma}{}^{a}\wedge\overset{3}{D}\psi+\overset{3}{\psi}{}^{T}C\overset{0}{\gamma}{}^{a}\wedge\overset{1}{D}\psi\Big).
\end{aligned} \tag{6.2}$$

The right side of (6.2) is now known from the solution to (6.1). The first term on the left will vanish as it did in second order. The only unknown is then $\overset{4}{\Theta}$ which is

$$\overset{4}{\Theta}{}^{a}{}_{b}=d\overset{4}{\omega}{}^{a}{}_{b}+\overset{4}{\omega}{}^{a}{}_{c}\wedge\overset{0}{\omega}{}^{c}{}_{b}+\overset{2}{\omega}{}^{a}{}_{c}\wedge\overset{2}{\omega}{}^{c}{}_{b}+\overset{0}{\omega}{}^{a}{}_{c}\wedge\overset{4}{\omega}{}^{c}{}_{b}. \tag{6.3}$$

After (6.3) is incorporated into (6.2), the latter becomes a first order linear differential equation for $\overset{4}{\omega}$, where again everything else in (6.2) is known. Instead of a first order equation for $\overset{4}{\omega}$, however, we may write a second order linear differential equation for $\overset{4}{E}$ if we solve

$$d\overset{4}{E}{}^{a}+\overset{4}{\omega}{}^{a}{}_{b}\wedge\overset{0}{E}{}^{b}+\overset{2}{\omega}{}^{a}{}_{b}\wedge\overset{2}{E}{}^{b}+\overset{0}{\omega}{}^{a}{}_{b}\wedge\overset{4}{E}{}^{b}=\overset{4}{S}{}^{b} \tag{6.4}$$

for $\overset{4}{\omega}$ and substitute in (6.2). [See Eqs. (4.8) and (4.9)]. The procedure is obviously general and may be continued to any order.

All the higher Grassmann algebras introduced in this section are permitted within the bounds of simple supergravity. (Increasing the order of the Grassmann algebra does not, for example, imply generalizations to extended supergravity.)

Although the subject of this paper is classical supergravity, a similar Grassmann expansion may be introduced whenever the total field contains both commuting and anticommuting parts.[4] Since no quantization is attempted, however, there is no relation between our use of the Grassmann algebra and the usual absorption and emission operators of a quantized Fermi field.

VII. SPECIAL GEOMETRIES

By imposing suitable constraints on supergravity one must, of course, recover simple gravity. By imposing weaker constraints one may explore field structures intermediate between simple gravity and the full complexity of supergravity. For example, a simple possible constraint is vanishing total curvature. In general relativity this condition implies Minkowskian space-time; but in a non-Riemannian theory, like supergravity, vanishing total curvature, and therefore distant parallelism, does not exclude a nontrivial metric structure. Physical theories based on geometries exhibiting distant parallelism have had a long history[4] and it is natural to ask whether supergravity has similar properties. In fact, it does not since, in the case of supergravity, vanishing total curvature again implies Minkowskian spacetime. On the other hand, similar but weaker conditions may lead to an interesting class of solutions.

VIII. VANISHING CURVATURE

The condition of vanishing curvature reads

$$\Theta^a{}_b = \overset{0}{\Theta}{}^a{}_b + \overset{2}{\Theta}{}^a{}_b(i\epsilon_1\epsilon_2) = 0 \tag{8.1}$$

or

$$\overset{0}{\Theta}{}^a{}_b = \overset{2}{\Theta}{}^a{}_b = 0. \tag{8.1a}$$

Let $R_{\mu\lambda\alpha\beta}$ be the curvature tensor formed from the symmetric part of the affine connection $\Gamma^\mu{}_{(\alpha\beta)}$ and let

$$\Theta_{\mu\lambda\alpha\beta} = R_{\mu\lambda\alpha\beta} + S_{\mu\lambda\alpha\beta}. \tag{8.2}$$

Then

$$S^\mu{}_{\lambda\alpha\beta} = S^\mu{}_{\lambda\beta|\alpha} - S^\mu{}_{\lambda\alpha|\beta} + S^\mu{}_{\sigma\beta}S^\sigma{}_{\lambda\alpha} - S^\mu{}_{\alpha\beta}S^\sigma{}_{\lambda\beta} + 2S^\mu{}_{\sigma\lambda}S^\sigma{}_{\alpha\beta}. \tag{8.3}$$

Here the solidus denotes the covariant derivative with respect to the total connection $\Gamma^\mu{}_{\alpha\beta}$. Then (8.1) implies

$$R_{\mu\lambda\alpha\beta} = -S_{\mu\lambda\alpha\beta} \tag{8.4}$$

for both $\overset{0}{\Theta}$ and $\overset{2}{\Theta}$.

The full affinity and its symmetric part are related to the Christoffel connection by

$$\Gamma^\mu{}_{\alpha\beta} = \left\{ {\mu \atop \alpha\beta} \right\} + K^\mu{}_{\alpha\beta}, \tag{8.5a}$$

where $K^\mu{}_{\alpha\beta}$ is the contortion, and

$$\Gamma^\mu{}_{(\alpha\beta)} = \left\{ {\mu \atop \alpha\beta} \right\} + S_{\alpha\beta}{}^\mu + S_{\beta\alpha}{}^\mu, \tag{8.5b}$$

where $S_{\alpha\beta\mu}$ is antisymmetric in the second and third indices. By (8.1a), (8.4), and (8.5b)

$$\overset{0}{R}{}_{\mu\lambda\alpha\beta} = -\overset{0}{S}{}_{\mu\lambda\alpha\beta} = 0, \tag{8.6}$$

$$-\overset{2}{R}{}^\mu{}_{\lambda\alpha\beta} = \overset{2}{S}{}^\mu{}_{\lambda\alpha\beta} = S^\mu{}_{\lambda\beta|\alpha} - S^\mu{}_{\lambda\alpha|\beta}, \tag{8.7}$$

and

$$\overset{0}{\Gamma}{}^\mu{}_{(\alpha\beta)} = \left\{ {\mu \atop \alpha\beta} \right\}, \tag{8.8}$$

$$\overset{2}{\Gamma}{}^\mu{}_{(\alpha\beta)} = S_{\alpha\beta}{}^\mu + S_{\beta\alpha}{}^\mu. \tag{8.9}$$

By (8.6) and (8.8)

$$\overset{0}{R}{}_{\mu\lambda\alpha\beta}\left[\left\{ {\mu \atop \alpha\beta} \right\}\right] = 0. \tag{8.10}$$

Therefore, vanishing total curvature as expressed in (8.1) implies vanishing Riemannian curvature. This result differs from the corresponding statement for Clifford space, for example.

One slight difference remains: By (8.1) it follows that up to a gauge transformation both $\overset{0}{\omega}_\mu$ and $\overset{2}{\omega}_\mu$ vanish. Therefore, by (4.8)

$$d\overset{2}{E}{}^a = S^a \tag{8.11}$$

or

$$\partial_\mu \overset{2}{E}{}^a{}_\lambda - \partial_\lambda \overset{2}{E}{}^a{}_\mu = S^a{}_{\mu\lambda}. \tag{8.11a}$$

Since there is no necessity for the ψ_μ to vanish there is in general the possibility of a torsion and therefore a nontrivial second-order tetrad field: $\overset{2}{E}{}^a{}_\mu$.

The remaining field equations to be satisfied are (4.1) which now read

$$\psi^T C\gamma^a\gamma^5 \wedge d\psi = 0 \tag{8.12}$$

or

$$\epsilon^{\mu\lambda\alpha\beta}\psi^T{}_\lambda C\gamma_a\gamma^5\partial_\alpha\psi_\beta = 0. \tag{8.12a}$$

These are nonlinear conditions on the solutions of the linear equations

$$\overset{0}{\gamma} \wedge d\psi = 0 \tag{8.13}$$

or

$$\epsilon^{\mu\lambda\alpha\beta}\gamma_\lambda\gamma^5\partial_\alpha\psi_\beta = 0. \tag{8.13a}$$

Solutions of this set have been found and are described in the Appendix. These solutions lead, however, to vanishing torsion. This may not be surprising since $dS^a = 0$ by (2.3). Therefore if the total curvature vanishes, the resultant structure is entirely Minkowskian.

IX. VANISHING SECOND ORDER CURVATURE

Since (8.1) removes all non-Minkowskian features, let us try the weaker condition

$$\overset{2}{\Theta}{}^a{}_b = 0 \tag{9.1}$$

while

$$\overset{0}{\Theta}{}^a{}_b \neq 0. \tag{9.2}$$

Then

$$\overset{0}{\Theta}{}_{\mu\lambda\alpha\beta} = \overset{0}{R}{}_{\mu\lambda\alpha\beta}\left[\left\{ {\mu \atop \alpha\beta} \right\}\right] \neq 0 \tag{9.3}$$

but

$$\overset{2}{\Theta}{}_{\mu\lambda\alpha\beta} = 0 \tag{9.4}$$

or

$$\overset{2}{R}{}^\mu{}_{\lambda\alpha\beta} = S^\mu{}_{\lambda\alpha;\beta} - S^\mu{}_{\lambda\beta;\alpha}, \tag{9.5}$$

where the covariant derivative is taken with respect to $\left\{ {\mu \atop \alpha\beta} \right\}$. In this case

$$\sum_{[\lambda\alpha\beta]} \overset{2}{R}{}^{\mu}{}_{\lambda\alpha\beta} = 2\sum_{[\lambda\alpha\beta]} S^{\mu}{}_{\lambda\alpha;\beta}. \tag{9.6}$$

By (9.3) this space is not metrically flat. This case is therefore not trivial and at the same time it is greatly simplified by (9.1) or (9.4). We have

$$\overset{2}{\Theta}{}^{a}{}_{b} = d\overset{2}{\omega}{}^{a}{}_{b} + \overset{2}{\omega}{}^{a}{}_{c}\wedge\overset{0}{\omega}{}^{c}{}_{b} + \overset{0}{\omega}{}^{a}{}_{c}\wedge\overset{2}{\omega}{}^{c}{}_{b} = 0. \tag{9.7}$$

Equation (9.7) may also be satisfied by

$$\overset{2}{\omega}{}^{a}{}_{b} = 0 \tag{9.8}$$

while in general

$$\overset{0}{\omega}{}^{a}{}_{b} \neq 0. \tag{9.9}$$

In view of (9.9) there are, of course, solutions other than (9.8).

Let us choose the simplest solution, (9.8). Then

$$S^{a} = d\overset{2}{E}{}^{a} + \overset{0}{\omega}{}^{a}{}_{b}\wedge\overset{2}{E}{}^{b}. \tag{9.10}$$

The $\overset{2}{E}{}^{a}{}_{\mu}$ fields are not trivial since $\overset{0}{\omega}{}^{a}{}_{b}$ no longer vanishes. We also have the identities

$$dS^{a} + \overset{0}{\omega}{}^{a}{}_{b}\wedge S^{b} = \overset{0}{\Theta}{}^{a}{}_{b}\wedge\overset{2}{E}{}^{b}, \tag{9.11}$$

$$0 = \overset{0}{\Theta}{}^{a}{}_{b}\wedge\overset{0}{E}{}^{b}. \tag{9.12}$$

The lowest order field equations (3.8)–(3.10) are unchanged by the restrictions (9.1) and (9.8). From these equations one may choose a solution $(\psi, \overset{0}{E}{}^{a}, \overset{0}{\omega}{}^{a}{}_{b})$. The second order equations are (9.10) and

$$\psi^{T}C\gamma^{a}\gamma^{5}\wedge\overset{0}{D}\psi = 0, \tag{9.13}$$

where

$$\overset{0}{D} = d + \tfrac{1}{2}\overset{0}{\omega}_{ab}\sigma^{ab}\wedge.$$

In component form one must choose solutions of the lowest order problem that satisfy

$$\epsilon^{\mu\lambda\alpha\beta}\psi^{T}{}_{\lambda}C\gamma_{c}\gamma^{5}\Big(\partial_{\alpha} + \tfrac{1}{2}\overset{0}{\omega}_{\alpha ab}\sigma^{ab}\Big)\psi_{\beta} = 0. \tag{9.14}$$

If a ψ_{α} can be found that satisfies these conditions, one may then solve the linear Eqs. (9.10) for $\overset{2}{E}{}^{a}{}_{\lambda}$

$$\partial_{\mu}\overset{2}{E}{}^{a}{}_{\lambda} - \partial_{\lambda}\overset{2}{E}{}^{a}{}_{\mu} + \overset{0}{\omega}{}^{a}{}_{b\mu}\overset{2}{E}{}^{b}{}_{\lambda} - \overset{0}{\omega}{}^{a}{}_{b\lambda}\overset{2}{E}{}^{b}{}_{\mu} = S^{a}{}_{\mu\lambda}. \tag{9.15}$$

Equations (9.14) may be reduced by use of the following expansion

$$\begin{aligned} i\epsilon^{\mu\lambda\alpha\beta}\gamma_{5} &= \overset{0}{\gamma}{}^{\mu}\overset{0}{\gamma}{}^{\lambda}\Big(\overset{0}{\gamma}{}^{\alpha}\overset{0}{\gamma}{}^{\beta} - \overset{0}{g}{}^{\alpha\beta}\Big) + \overset{0}{g}{}^{\mu\lambda}\Big(\overset{0}{g}{}^{\alpha\beta} - \overset{0}{\gamma}{}^{\alpha}\overset{0}{\gamma}{}^{\beta}\Big) \\ &\quad + \Big(\overset{0}{g}{}^{\mu\alpha}\overset{0}{\gamma}{}^{\lambda}\overset{0}{\gamma}{}^{\beta} - \overset{0}{g}{}^{\mu\beta}\overset{0}{\gamma}{}^{\lambda}\overset{0}{\gamma}{}^{\alpha}\Big) + \Big(\overset{0}{g}{}^{\lambda\beta}\overset{0}{\gamma}{}^{\mu}\overset{0}{\gamma}{}^{\alpha} - \overset{0}{g}{}^{\lambda\alpha}\overset{0}{\gamma}{}^{\mu}\overset{0}{\gamma}{}^{\beta}\Big) \\ &\quad + \Big(\overset{0}{g}{}^{\mu\beta}\overset{0}{g}{}^{\lambda\alpha} - \overset{0}{g}{}^{\mu\alpha}\overset{0}{g}{}^{\lambda\beta}\Big). \end{aligned} \tag{9.16}$$

The contracted form of the Rarita–Schwinger equations is

$$\Big(\overset{0}{\gamma}{}^{\alpha}\overset{0}{\gamma}{}^{\beta} - \overset{0}{g}{}^{\alpha\beta}\Big)\overset{0}{D}_{\alpha}\psi_{\beta} = 0. \tag{9.17}$$

Therefore, the first four terms of (9.16), when substituted in (9.14), cancel in pairs, as shown. Likewise the next four terms also cancel in pairs in view of (5.5) and the related equation

$$\overset{0}{\gamma}{}^{\alpha}\overset{0}{D}_{\alpha}\psi_{\beta} = \left\{{}^{\sigma}_{\beta\alpha}\right\}\overset{0}{\gamma}{}^{\alpha}\psi_{\sigma}. \tag{9.18}$$

Hence (9.14) may be reduced to

$$(\psi^{\alpha})^{T}C\gamma_{a}\Big(\overset{0}{D}_{\alpha}\psi_{\mu} - \overset{0}{D}_{\mu}\psi_{\alpha}\Big) = 0. \tag{9.19}$$

By imposing the ansatz (9.1) and in particular (9.8) one replaces linear differential equations for $\overset{2}{E}{}^{a}$ by the nonlinear constraints (9.13) or (9.19). It is possible that Eqs. (3.8) and (9.19) have common solutions that do not reduce to the previous case of totally vanishing curvature.

ACKNOWLEDGMENT

One of us (R. F.) would like to thank Dr. Urrutia for discussions in the early stages of this work.

APPENDIX

We solve the Rarita–Schwinger equation in the flat space with the nonlinear constraint (8.12a). Since ω_{μ} is zero, the Rarita–Schwinger equation reduces to (5.2) with the constraints (5.1) and (5.3). The general solution of these equations is[5]

$$\begin{aligned} \psi_{\mu}(x) &= \int \frac{d^{3}k}{2k^{0}}\,[A(k)U_{\mu}(k,+)e^{-ik\cdot x} + B(k)U_{\mu}(k,-)e^{-ik\cdot x} \\ &\quad + A^{*}(k)V_{\mu}(k,+)e^{ik\cdot x} + B^{*}(k)V_{\mu}(k,-)e^{ik\cdot x}], \end{aligned} \tag{A1}$$

where $A(k)$, $B(k)$, $A^{*}(k)$, and $B^{*}(k)$ are odd elements of Grassmann algebra and

$$U_{\mu}(k,+) = \begin{bmatrix} a_{\mu}(k) \\ -b_{\mu}^{*}(k) \\ 0 \\ 0 \end{bmatrix}, \tag{A2a}$$

$$U_{\mu}(k,-) = \begin{bmatrix} b_{\mu}(k) \\ a_{\mu}^{*}(k) \\ 0 \\ 0 \end{bmatrix}, \tag{A2b}$$

$$V_{\mu}(k,+) = \begin{bmatrix} 0 \\ 0 \\ -b_{\mu}(k) \\ -a_{\mu}^{*}(k) \end{bmatrix}, \tag{A2c}$$

$$V_{\mu}(k,-) = \begin{bmatrix} 0 \\ 0 \\ a_{\mu}(k) \\ -b_{\mu}^{*}(k) \end{bmatrix}, \tag{A2d}$$

$$a_0(k) = 0,$$

$$a_1(k) = [1-(1-\cos\theta)e^{i\phi}\cos\phi] \times [\cos(\theta/2) - i\cos\theta\ \sin(\theta/2)],$$

$$a_2(k) = i[1+i(1-\cos\theta)e^{i\phi}\sin\phi] \times [\cos(\theta/2) - i\cos\theta\ \sin(\theta/2)],$$

$$a_3(k) = -\sin\theta e^{i\phi}[\cos(\theta/2) - i\cos\theta\sin(\theta/2)], \tag{A2e}$$

$$b_0(k) = 0,$$

$$b_1(k) = -i[1-(1-\cos\theta)e^{-i\phi}\cos\phi]\sin\theta\ \sin(\theta/2)e^{-i\phi},$$

$$b_2(k) = -[1-i(1-\cos\theta)e^{-i\phi}\sin\phi]\sin\theta\ \sin(\theta/2)e^{-i\phi},$$

$$b_3(k) = i\sin^2\theta\sin(\theta/2)e^{-2i\phi}, \tag{A2f}$$

$$k^\mu = k^0(1,\sin\theta\cos\phi,\sin\theta\ \sin\phi,\cos\theta). \tag{A2g}$$

Substituting this solution into (8.12a), we obtain six constraints on $(A(k), B(k), A^*(k)$ and $B^*(k)$.

$$A(k')A(k) = 0, \tag{A3a}$$

$$A(k')B(k) = 0, \tag{A3b}$$

$$A(k')A^*(k) = 0, \tag{A3c}$$

$$A(k')B^*(k) = 0, \tag{A3d}$$

$$A^*(k')B(k) = 0, \tag{A3e}$$

$$B^*(k')B(k) = 0. \tag{A3f}$$

Since k and k' are arbitrary, the only non-trivial solution is

$$A(k) = A'(k)\epsilon, \tag{A4a}$$

$$B(k) = B'(k)\epsilon, \tag{A4b}$$

where $A'(k)$ and $B'(k)$ are pure numbers and ϵ is an odd element of Grassmann algebra. Since $\psi_\mu(x)$ contains only one odd element of Grassmann algebra, the torsion tensor vanishes and the second order tetrad field $\overset{2}{E}{}^a{}_\mu$ also vanishes by (8.11a).

[1]P. C. Aichelburg and T. Dereli, Phys. Rev. D **18**, 1754 (1978).

[2]L. F. Urrutia (private communication).

[3]S. Einstein and R. Finkelstein, J. Math. Phys. **20**, 1972 (1979).

[4]A. Einstein, Preuss. Akad. Wiss. Berlin **217**, 224 (1928).

[5]P. A. Carruthers, *Spin and Isospin in Particle Physics* (Gordon and Breach, New York, 1971).

Charged Demianski metric

K. D. Krori, T. Chaudhury, and R. Bhattacharjee
Mathematical Physics Forum, Cotton College, Gauhati–781001, India

(Received 28 May 1980; accepted for publication 30 July 1980)

A solution of Einstein–Maxwell equations is presented. This solution is the charged version of the Demianski metric.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Kerr[1] and Newman *et al.*[2] have solved the Einstein field equations and presented two different metrics which are respectively known as the Kerr metric and the NUT metric. Later Demianski[3] presented a solution of the Einstein field equations from which the Kerr and NUT solutions may be obtained as special cases. On the other hand, Newman *et al.*[4] obtained a solution of Einstein–Maxwell equations which is the charged version of the Kerr solution. We have recently solved Einstein–Maxwell equations and found a solution which is the charged version of the NUT metric.[5]

In this paper we solve the Einstein—Maxwell equations and obtain a solution which is the charged Demianski metric. The interesting feature of our result is the straightforward recovery of the charged Kerr field or the charged NUT–like field by suitable adjustment of the values of the associated parameters.

II. FIELD EQUATIONS AND THEIR SOLUTION

Let us consider the line-element

$$ds^2 = e^{2\alpha}(du - Hd\phi)^2 + 2(du - Hd\phi)(dr + Gd\phi) - e^{2\beta}(d\theta^2 + \sin^2\theta d\phi^2), \tag{1}$$

where α and β are functions of r and θ, and G and H are functions of θ alone. The field equations for a source-free electromagnetic field are

$$R_{ij} = -8\pi T_{ij}, \tag{2}$$

$$F_{ij;k} + F_{jk;i} + F_{ki;j} = 0, \tag{3}$$

$$F^{ij}_{\ ;j} = 0, \tag{4}$$

with

$$T_{ij} = (1/4\pi)[-F_{il}F^l_j + (1/4)g_{ij}F_{lm}F^{lm}], \tag{5}$$

where T_{ij} is the energy momentum tensor due to the source-free electromagnetic field.

The components of the electromagnetic field tensors may be given by[6]

$$F^{\mu\nu} = (-g)^{-1/2}\epsilon^{\mu\nu\lambda}A_{,\lambda}, \tag{6}$$

$$F_{0\mu} = B_{,\mu}, \tag{7}$$

where $\epsilon^{\mu\nu\lambda}$ is the alternating three-index symbol and a comma indicates ordinary differentiation. μ, ν, λ may take the values 1, 2, 3, and A and B are scalar potentials. The nonvanishing components of electromagnetic fields are F_{01}, F_{02}, F^{31} and F^{23}; they are given by

$$F_{01} = B_{,1}, \tag{8}$$

$$F_{02} = B_{,2}, \tag{9}$$

$$F^{31} = e^{-2\beta}\csc\theta A_{,2}, \tag{10}$$

$$F^{23} = e^{-2\beta}\csc\theta A_{,1}, \tag{11}$$

where the suffixes 1 and 2 indicate derivatives with respect to r and θ respectively.

The field equations (2) for the line-element (1) are

$$\begin{aligned}&(e^{4\alpha} + G^2e^{2\alpha-2\beta}\csc^2\theta)(\alpha_{,11} + 2\alpha^2_{,1}) + e^{2\alpha-2\beta}\\&\quad\times(\alpha_{,22} + 2\alpha^2_{,2} + \cot\theta\alpha_{,2}) + 2e^{4\alpha}\alpha_{,1}\beta_{,1}\\&\quad - e^{2\alpha-4\beta}\csc^2\theta H_{,2}G_{,2} + \tfrac{1}{2}e^{-4\beta}\csc^2\theta G^2_{,2}\\&\quad + \tfrac{1}{2}e^{4\alpha-4\beta}\csc^2\theta H^2_{,2}\\&\qquad = (e^{2\alpha} + G^2e^{-2\beta}\csc^2\theta)\\&\qquad\quad\times(A^2_{,1} + B^2_{,1}) + e^{-2\beta}(A^2_{,2} + B^2_{,2}),\end{aligned} \tag{12}$$

$$\begin{aligned}&2\beta_{,11} + 2\beta^2_{,1} - \tfrac{1}{2}e^{-4\beta}\csc^2\theta H^2_{,2}\\&\qquad = 2e^{-4\alpha-2\beta}[2G\csc\theta(A_{,1}B_{,2} - A_{,2}B_{,1})\\&\qquad\quad - G^2\csc^2\theta(A^2_{,1} + B^2_{,1}) - (A^2_{,2} + B^2_{,2})],\end{aligned} \tag{13}$$

$$\beta_{,12} - Ge^{-2\beta}\csc^2\theta\beta_{,1}H_{,2} = 2e^{-2\alpha}(A_{,1}A_{,2} + B_{,1}B_{,2}), \tag{14}$$

$$\begin{aligned}&(e^{2\alpha+2\beta} + G^2\csc^2\theta)\beta_{,11} + \beta_{,22} + 2e^{2\alpha+2\beta}(\alpha_{,1}\beta_{,1} + \beta^2_{,1})\\&\quad + \cot\theta\beta_{,2} + (G_{,2} - \tfrac{1}{2}e^{2\alpha}H_{,2})e^{-2\beta}\csc^2\theta H_{,2} - 1\\&\quad = e^{-2\alpha}(A^2_{,2} + B^2_{,2}) - (e^{2\beta} + G^2e^{-2\alpha}\csc^2\theta)\\&\qquad\times(A^2_{,1} + B^2_{,1}),\end{aligned} \tag{15}$$

$$\begin{aligned}&4GHe^{2\alpha}\alpha_{,1}\beta_{,1} + 2G^2\beta^2_{,1} - 2He^{2\alpha-2\beta}\alpha_{,2}H_{,2} - (He^{2\alpha} - G)\\&\quad\times e^{-2\beta}(H_{,22} - 2\beta_{,2}H_{,2} - \cot\theta H_{,2})\\&\quad + He^{-2\beta}(G_{,22} - 2\beta_{,2}G_{,2} - \cot\theta G_{,2})\\&\quad + GHe^{-4\beta}\csc^2\theta(e^{2\alpha}H^2_{,2} - G_{,2}H_{,2})\\&\quad - \frac{G^2}{2}e^{-4\beta}\csc^2\theta H^2_{,2}\\&\quad = 2(1 + G^2e^{-2\alpha-2\beta}\csc^2\theta)(Ge^{-2\alpha} - H)\\&\qquad\times[2\sin\theta(A_{,1}B_{,2} - A_{,2}B_{,1}) - G(A^2_{,1} + B^2_{,1})]\\&\qquad + 2(GHe^{-2\alpha-2\beta} - e^{-2\alpha}\sin^2\theta - G^2e^{-4\alpha-2\beta})\\&\qquad\times(A^2_{,2} + B^2_{,2}).\end{aligned} \tag{16}$$

Solving the field equations (12)–(16) and the Maxwell equations (3) and (4) one gets

$$e^{2\alpha} = 1 - (2mr + 2bF - e^2)/(r^2 + F^2), \tag{17}$$

$$e^{2\beta} = r^2 + F^2, \tag{18}$$

$$G = a\sin^2\theta - c\cos\theta + c\sin^2\theta\ln\tan(\theta/2), \tag{19}$$

 0022-2488/81/102235-02$01.00

$$H = G - 2b\cos\theta, \tag{20}$$

$$A_{,1} = 2Fer/(r^2 + F^2)^2, \tag{21}$$

$$B_{,1} = (F^2 - r^2)e/(r^2 + F^2)^2, \tag{22}$$

$$A_{,2} = [(r^2 - F^2)/(r^2 + F^2)^2]\,Ge\csc\theta, \tag{23}$$

$$B_{,2} = [2FGer/(r^2 + F^2)^2]\csc\theta, \tag{24}$$

where

$$F = a\cos\theta + b + c + c\cos\theta \ln\tan(\theta/2); \tag{25}$$

and a, b, c, e and m are constants. The constants m and e can be identified as mass and charge of the object under consideration, while a, b, c can be identified as the parameters occurring in the Demianski metric.

Using Eqs. (17)–(25) in Eqs. (8)–(11) we get,

$$F_{01} = -e(r^2 - F^2)/(r^2 + F^2)^2, \tag{26}$$

$$F_{02} = 2FGer/(r^2 + F^2)^2, \tag{27}$$

$$F_{31} = eH(r^2 - F^2)/(r^2 + F^2)^2, \tag{28}$$

and

$$F_{23} = [(r^2 + F^2)\sin\theta + GH\csc\theta\,]2Fer/(r^2 + F^2)^2. \tag{29}$$

III. DISCUSSION

Putting $e = 0$ in our solution, one gets the Demianski metric. If one substitutes $b = c = 0$ in our solution one gets the Kerr–Newman solution. On the other hand, putting a and c equal to zero one gets a charged NUT-like solution. Finally if one puts $c = 0$, one gets the charged version of combined Kerr and NUT-like fields.

ACKNOWLEDGMENT

The authors are grateful to the Government of Assam, Dispur, for providing all facilities at Cotton College, Gauhati–781001, India to carry out this piece of work.

[1]R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).
[2]E. Newman, L. Tomburino, and T. Unti, J. Math. Phys. **4**, 915 (1963).
[3]M. Demianski, Phys. Lett. **42** A 157 (1972).
[4]E. Newman *et al.*, J. Math. Phys. **6**, 918 (1965).
[5]K. D. Krori, T. Chaudhury, B. B. Paul and R. Bhattacharjee, submitted for publication (1980).
[6]B. Kent Harrison, Phys. Rev. **138**, B488 (1965).

Viscous fluid interpretation of electromagnetic fields

A. K. Raychaudhuri and S. K. Saha
Department of Physics, Presidency College, Calcutta-700073, India

(Received 8 March 1980; accepted for publication 17 October 1980)

As observed by Tupper, if the Ricci tensor satisfies the Rainich conditions for a nonnull electromagnetic field, the energy stress tensor may sometimes be interpreted as due to a viscous fluid (apparently Tupper believed this to be possible for all electrovac solutions). The present paper shows that for such an interpretation to be possible, there must be an additional symmetry property. Even then in some cases the interpretation may not work due to unacceptable behavior of the fluid velocity vector.

PACS numbers: 04.20.Jb

1. INTRODUCTION

In the international General Relativity and Gravitation conference, 1977, Tupper[1] reported that many metrics satisfying the Einstein–Maxwell equations could be as well interpreted as due to a viscous fluid. He found that for nonnull electromagnetic fields, all the cases of electrovac solutions he studied admitted such an interpretation although he could not offer any proof for the general validity of this alternative possibility.

In this paper we first make an investigation of the Tupper problem and are led to the conclusion that the viscous fluid interpretation is possible only if the electrovac solution possesses a certain symmetry property. This result allows us to cite a counterexample to Tupper's idea from known electrovac solutions.

However, as noted by Novello,[2] Tupper's interpretation allows one to discover some cosmological solutions having somewhat unusual properties. We here make a study of the viscous fluid universe represented by two already known electrovac metrics. Again we find that in the second case the interpretation cannot be accepted owing to the peculiar velocity vector that is obtained.

2. THE VISCOUS FLUID INTERPRETATION FOR THE ELECTROVAC CASE

In the following we use the signature convention $+ - - -$.

In an electovac universe, the Ricci tensor components satisfy the Rainich algebraic conditions:

$$R^{\mu}_{\ \mu} = 0 \tag{1}$$

$$R^{\mu}_{\ \alpha} R^{\alpha}_{\ \nu} = \tfrac{1}{4}\delta^{\mu}_{\nu}[R_{\alpha\beta}R^{\alpha\beta}], \tag{2}$$

with

$$R_{\alpha\beta}R^{\alpha\beta} \geqslant 0. \tag{3}$$

The equality in (3) occurs for null fields. Tupper's idea is to satisfy the following equation:

$$-R^{\mu}_{\ \nu} = 8\pi[(p+\rho)\vartheta^{\mu}\vartheta_{\nu} - p\delta^{\mu}_{\nu} + 2\eta\sigma^{\mu}_{\nu}], \tag{4}$$

where p and ρ should be positive, ϑ^{μ} a unit timelike vector, and η a scalar (not necessarily constant). The shear tensor $\sigma_{\mu\nu}$ corresponding to the vector ϑ^{μ} is given by

$$\sigma_{\mu\nu} \equiv \vartheta_{(\mu;\nu)} - \tfrac{1}{3}\theta(g_{\mu\nu} - \vartheta_{\mu}\vartheta_{\nu}) - \vartheta_{(\mu}\dot{\vartheta}_{\nu)}. \tag{5}$$

Tupper introduces in (4) a bulk viscosity term as well. However this simply alters the value of pressure and in this purely formal discussion does not play an important part. Also, except under rather stringent conditions; bulk viscosity effect may be neglected in comparison with shear viscosity.

Equation (5) requires $\sigma_{\mu\nu}\vartheta^{\mu} = 0$ and hence from (4)

$$R^{\mu}_{\ \nu}\vartheta^{\nu} = -8\pi\rho\vartheta^{\mu}. \tag{6}$$

From (6) and (2),

$$64\pi^2\rho^2 = \tfrac{1}{4}R_{\alpha\beta}R^{\alpha\beta}. \tag{7}$$

Equation (6) shows that ϑ^{μ} is an eigenvector of $R^{\mu}_{\ \nu}$ while (7) shows that the eigenvalues of $R^{\mu}_{\ \nu}$ are degenerate having the values

$$\pm 8\pi\rho = \pm\tfrac{1}{2}(R_{\alpha\beta}R^{\alpha\beta})^{1/2}.$$

The reality condition for the electromagnetic field

$$R_{\mu\nu}\vartheta^{\mu}\vartheta^{\nu} < 0$$

indicates that for the timelike eigenvector, the eigenvalue is negative, i.e., ρ is positive as is indeed demanded by the Tupper interpretation. The degeneracy of the eigenvalue means that even the normalization condition $\vartheta^{\mu}\vartheta_{\mu} = +1$ would not determine the eigenvector ϑ^{μ} uniquely.

Again Eq. (5) requires

$$\sigma^{\mu}_{\mu} = 0 \tag{8}$$

and hence from (4) and (1),

$$p = \rho/3. \tag{9}$$

As is evident from (4) any eigenvector of $R_{\mu\nu}$ normal to ϑ^{μ} is an eigenvector of $\sigma_{\mu\nu}$ as well. In particular, corresponding to the degenerate eigenvalue $+8\pi\rho$ of $R^{\mu}_{\ \nu}$, the eigenvalue of $\sigma_{\mu\nu}$ is also degenerate and equal to $-\rho/3\eta$. Remembering that ϑ^{μ} is an eigenvector of $\sigma_{\mu\nu}$ with the eigenvalue zero, we have the eigenvalues of $\sigma_{\mu\nu}$ as 0, $-\rho/3\eta$, $-\rho/3\eta$, $2\rho/3\eta$.

To proceed further, we introduce the electric and the magnetic field vectors by the relations

$$E_{\beta} = F_{\alpha\beta}\vartheta^{\alpha}, \quad B_{\beta} = \tfrac{1}{2}\eta_{\alpha\beta\gamma\delta}\vartheta^{\alpha}F^{\gamma\delta}. \tag{10}$$

One has then, for the energy stress tensor of the field,

$$4\pi T_{\alpha\beta} = (\tfrac{1}{2}g_{\alpha\beta} - \vartheta_\alpha\vartheta_\beta)(E_\mu E^\mu + B_\mu B^\mu) - (E_\alpha E_\beta + B_\alpha B_\beta) - (S_\alpha\vartheta_\beta + S_\beta\vartheta_\alpha), \quad (11)$$

with

$$S_\alpha = \eta_{\alpha\rho\lambda\mu}E^\rho B^\lambda\vartheta^\mu. \quad (12)$$

Einstein's equations for the electrovac universe

$$8\pi T_{\alpha\beta} = -R_{\alpha\beta}$$

show that the vector ϑ^μ will be an eigenvector of $R_{\mu\nu}$ iff S_α vanishes and in that case, one can consider the field to be purely electric (or magnetic) so that, in the electrical case (the reasoning is identical in the magnetic case)

$$-R_{\alpha\beta} = 8\pi T_{\alpha\beta} = (g_{\alpha\beta} - 2\vartheta_\alpha\vartheta_\beta)E_\mu E^\mu - 2E_\alpha E_\beta. \quad (13)$$

If the viscous fluid interpretation is to work, we must have, comparing (13) and (4)

$$8\pi\rho = -E^\alpha E_\alpha = E^2, \quad (14)$$

$$2\eta\sigma_{\mu\nu} = (2\rho/3)(\vartheta_\mu\vartheta_\nu - g_{\mu\nu}) - E_\mu E_\nu/(4\pi). \quad (15)$$

Equation (13) shows that E^μ and ϑ^μ are eigenvectors of $R_{\alpha\beta}$ belonging to the same eigenvalue and the corresponding eigenvalues of $\sigma_{\mu\nu}$ are $2\rho/3\eta$ and 0.

The situation thus appears to be as follows: for a given electrovac solution, ϑ^μ and E^μ are arbitrary to the extent that we can change to $\bar{\vartheta}^\mu$ and $\bar{E}^\mu$ defined by

$$\bar{\vartheta}^\mu = \alpha\vartheta^\mu + \beta E^\mu,$$

$$\bar{E}^\mu = F^\mu_\nu\bar{\vartheta}^\nu = \alpha E^\mu + \beta E^2\vartheta^\mu,$$

with

$$\alpha^2 - \beta^2E^2 = 1,$$

$$E^2 = -E^\mu E_\mu > 0.$$

The Tupper type of interpretation is then possible if at least for some α,β (i) $\bar{E}^\mu$ is an eigenvector of the corresponding shear tensor $\bar{\sigma}_{\mu\nu}$ (cases may occur where this condition is satisfied for any α,β), (ii) the section of the $\bar{\sigma}_{\mu\nu}$ ellipsoid by a plane orthogonal to $\bar{\vartheta}^\mu,\bar{E}^\mu$ is a circle. Of these, the first condition seems to be not very critical owing to the freedom in our choice of α,β while the second condition demands some symmetry of the solution—in particular a local rotational symmetry about E^μ as obtained in the cases investigated by Tupper is sufficient for (ii). We shall cite an example where this symmetry is lacking and the viscous fluid interpretation cannot be given.

However, before that we deduce a formal relation which would be useful in our later discussion.

From Eq. (6), we get

$$R^{\mu\nu}\vartheta_{\nu;\mu} = -8\pi\rho_{,\mu}\vartheta^\mu - 8\pi\rho\theta.$$

Also from (5) and (6)

$$R^{\mu\nu}\sigma_{\mu\nu} = R^{\mu\nu}\vartheta_{\nu;\mu} - (8\pi\rho\theta/3)$$
$$= -8\pi\rho_{,\mu}\vartheta^\mu - (32\pi\rho\theta/3).$$

From Eq. (4) we then get, remembering (9) and (2),

$$\rho_{,\mu}\vartheta^\mu = 4\rho^2/(3\eta) - 4\rho\theta/3. \quad (16)$$

3. A COUNTEREXAMPLE

As an example of an electrovac universe where the viscous fluid interpretation fails, consider the following metric[3,4]:

$$ds^2 = -A(d\kappa^1)^2 - B(d\kappa^2)^2 - C(d\kappa^3)^2 + D(d\kappa^0)^2, \quad (17)$$

with

$$C = A = D^{-1} = \ln^2\left(\frac{(\kappa^1)^l}{a}\right), \quad (18)$$

$$B = (\kappa^1)^2\ln^2\left(\frac{(\kappa^1)^l}{a}\right),$$

with l and a arbitrary constants. In this case,

$$R^1_1 = R^0_0 = -R^2_2 = -R^3_3 \quad (19)$$

so that the vectors ϑ^μ and E^μ (i.e. the eigenvectors of R^μ_ν for the eigenvalue $-8\pi\rho$) are in the κ^0,κ^1 space. As however $B \neq C$, there is no rotational symmetry in the orthogonal 2-space.

With

$$\vartheta^\mu = \alpha\delta^\mu_0 + B\delta^\mu_1,$$

where

$$D\alpha^2 - A\beta^2 = 1,$$

we get for the shear tensor components

$$\sigma^2_2 = \tfrac{1}{2}\vartheta^1\frac{\partial}{\partial\kappa^1}(\ln B) - \tfrac{1}{3}\theta,$$

$$\sigma^3_3 = \tfrac{1}{2}\vartheta^1\frac{\partial}{\partial\kappa^1}(\ln C) - \tfrac{1}{3}\theta,$$

$$\sigma^2_1 = \sigma^1_2 = \sigma^2_3 = \sigma^3_2 = \sigma^2_0 = \sigma^0_2 = \sigma^3_0 = \sigma^0_3 = 0. \quad (20)$$

Thus $\sigma^2_2 \neq \sigma^3_3$ and the eigenvalues are not degenerate as required for the viscous fluid interpretation.

4. A DISCUSSION OF TWO CASES

We shall work out the details of the viscous fluid motion in two particular cases. In the first case, we have the metric with plane symmetry[5]:

$$ds^2 = e^{\alpha\kappa}dt^2 - e^{-2\alpha\kappa}d\kappa^2 - e^{-\alpha\kappa}(dy^2 + dz^2), \quad (21)$$

where α is a constant.

The vectors ϑ^μ,E^μ are again in the (κ,t) space and the isotropy of the orthogonal 2-space is apparent. Taking ϑ^0 and ϑ^1 functions of κ alone, we find that the field equations are all satisfied if we take,

$$\eta\vartheta^1{}_{,1} = \frac{\alpha^2}{32\pi}e^{2\alpha\kappa}$$

with ϑ^1,ϑ^0 functions of κ alone. If η is assumed to be constant

$$\vartheta^1 = (\alpha/64\pi\eta)e^{2\alpha\kappa} \quad (22)$$

and

$$\rho = 3p = (\alpha^2/32\pi)e^{2\alpha\kappa}. \quad (23)$$

One may describe the situation as a laminar, irrotational flow of a viscous fluid with the fluid velocity increasing as κ increases, all the physical variables blowing up as $\kappa\to\infty$. As the metric (21) admits a timelike Killing vector, the flow may be called stationary, but the expansion and the shear do not vanish. The situation seems rather remarkable and the authors have not come across any example in the published literature where a stationary metric is associated with an expanding velocity field.

Again here the shear tensor components

$$\sigma_2^2 = \sigma_3^3 = -\tfrac{1}{3}\vartheta^1{}_{,1},$$
$$\sigma_1^2 = \sigma_2^1 = \sigma_3^2 = \sigma_2^3 = \sigma_0^2 = \sigma_2^0 = \sigma_0^3 = \sigma_3^0 = 0$$

indicates that the eigenvalues are degenerate as needed for a viscous fluid interpretation.

The second case that we take up is rather complicated: The cylindrically symmetric metric

$$ds^2 = dt^2 - e^{a^2\gamma^2}(d\gamma^2 + dz^2) - (\gamma^2 - a^2\gamma^4)d\phi^2 + 2a\gamma^2 d\phi dt \tag{24}$$

represents a distribution of massless charges[6] with the nonvanishing Ricci tensor components

$$R^1_1 = -R^2_2 = R^3_3 = -R^0_0 = 2a^2e^{-a^2\gamma^2}, \tag{25}$$
$$R^0_3 = -4a^3\gamma^2e^{-a^2\gamma^2}$$

and a charge density σ given by

$$4\pi|\sigma| = 2a^2e^{-a^2\gamma^2}. \tag{26}$$

The solution is singularity free but has closed timelike lines for $\gamma^2 > a^{-2}$. The electromagnetic field was described as a simple magnetic field in the z direction with

$$|H|^2 = -H^Z H_Z = 2a^2e^{-a^2\gamma^2}. \tag{27}$$

From Eq. (25), the eigenvectors ϑ^μ and H^μ can have only t, z components. The requirement that H^μ must be an eigenvector of the shear tensor $\sigma_{\mu\nu}$ gives,

$$(\vartheta^2)^2 = 1/(A - e^{a^2\gamma^2}), \tag{28}$$

$$(\vartheta^0)^2 = A/(A - e^{a^2\gamma^2}), \tag{29}$$

where A is a function of z alone. In the above we have numbered the coordinates t, γ, z, ϕ by the indices 0, 1, 2, 3, respectively. We might have imagined that as the metric tensor components are independent of z and t, the velocity components would also be independent of these coordinates—but then the expansion θ would vanish and as ρ is independent of z and t, Eq. (16) would make ρ vanish. The field equations are all satisfied if

$$\frac{a^2}{4\pi}e^{-a^2\gamma^2} = \frac{-\eta}{2}\frac{A_{,2}}{(A - e^{a^2\gamma^2})^{3/2}}. \tag{30}$$

Obviously in this case we can not have η a constant. Although the situation is stationary in the sense that the metric as well as the fluid velocity field are independent of the time (the killing vector field is in this case not hypersurface orthogonal), the expansion and the shear do not vanish.

The shear tensor σ^μ_ν has the form

$$\begin{array}{c|cccc|}
 & 0 & 1 & 2 & 3 \\
0 & \frac{2}{3}(1-\vartheta_0^2) & 0 & -\frac{2}{3}\vartheta_2\vartheta_0 & \frac{1}{3}a\gamma^2 + \frac{2}{3}a\gamma^2(1-\vartheta_0^2) \\
1 & 0 & -\frac{1}{3} & 0 & 0 \\
2 & -\frac{2}{3}\vartheta^2\vartheta_0 & 0 & \frac{2}{3}\vartheta_0^2 & -\frac{2}{3}a\gamma^2\vartheta_0\vartheta^2 \\
3 & 0 & 0 & 0 & -\frac{1}{3}
\end{array} \times\theta \tag{31}$$

An intriguing point about the solution under discussion is that while there is no geometric singularity [the metric (24) gives an asymptotically Euclidean geometry as $\gamma \to \infty$], the velocity vector ϑ^μ has a singularity at a finite γ given by $\gamma = a^{-1}\ln^{1/2}A$ and has imaginary components beyond that, however the coordinate velocity dz/dt remains regular and nonvanishing. One should therefore apparently reject the viscous field interpretations in the present case—alternatively one can consider the fluid distribution to be limited within the surface where the velocity becomes singular.

ACKNOWLEDGMENTS

One of the authors (SKS) is thankful to the U.G.C. (India) for the award of a Teacher Fellowship under the Faculty Improvement Programme and to Stewart College, Cuttack for granting him leave. Thanks are also due to the authorities of Presidency College, Calcutta for their hospitality.

[1] B. O. J. Tupper, 8th International Conference on General Relativity and Gravitation, Canada (1977).

[2] M. Novello, CBPF, Preprint A0006/79, Jan. 1979.

[3] W. B. Bonnor, Proc. Phys. Soc. A **66**, 145 (1953).

[4] A. K. Raychaudhuri, Ann. Phys. **11**, 501, (1960).

[5] G. C. McVittie, Proc. Roy. Soc. **124**, 366 (1929).

[6] M. M. Som, and A. K. Raychaudhuri, Proc. Roy. Soc. A **304**, 81 (1968).

Space–times with geodesic, shear-free, twist-free, nonexpanding rays. II

Garry Ludwig
Department of Mathematics, University of Alberta, Edmonton, Alberta, T6G2G1, Canada

(Received 30 January 1981; accepted for publication 18 March 1981)

Reduced gravitational field equations are obtained for space–times with geodesic, shear-free, twist-free, and nonexpanding rays. No restrictions are imposed on either the Weyl tensor or the Ricci tensor, and a further restriction made previously that the spin coefficient τ vanish is also dropped. Special cases of Einstein–Maxwell space–times are investigated further.

PACS numbers: 04.50. + h, 04.20.Cv

I. INTRODUCTION

In a previous paper[1] the present author derived reduced gravitational field equations for space–times with geodesic, shear-free, twist-free, nonexpanding rays. The (implicit) solutions obtained generalized the Kundt[2] class of space–times to arbitrary trace-free Ricci tensors. They were solutions in the sense that the radial dependence had been found but the reduced gravitational field equations together with the equations governing the source still had to be solved before explicit solutions would be obtained. In general, such equations are quite difficult to deal with.

In deriving these implicit solutions the following additional assumptions were made. First, space–time was assumed to be algebraically special with the repeated principal null vector of the Weyl tensor being tangential to the ray congruence. Second, the Ricci scalar was assumed to be zero, and third, the spin-coefficient τ also was to vanish. In this paper we remove all three of these conditions, assuming only that the space–times under investigation possess a geodesic null congruence free of shear, twist, and expansion. In Sec. II, we use Penrose's method of conformal rescaling[3] to find the spin-coefficients, the metric coefficients, the metric, and the tetrad components (relative to a suitable tetrad) of the Weyl and Ricci tensors for such space–times. The advantages of the conformal method have been discussed previously.[1] In Sec. III, we specialize to the Einstein–Maxwell case, obtaining more explicit solutions for some special cases. The notation of this paper is that of Ref. 1.

II. THE SOLUTION

Again[1] we consider a space–time which has as (part of) its conformal boundary a line N on which the conformal factor Ω vanishes, and on which $\hat{\nabla}_a \Omega \neq 0$. Corresponding to each geodesic arriving at a point S on N, we choose a null tetrad $\{\hat{k}_a, \hat{m}_a, \hat{\bar{m}}_a, \hat{n}_a\}$, $\hat{k}_a$ tangent to the geodesic, such that

$$\hat{\nabla}_a \Omega|_{\Omega=0} = K^0 \hat{k}_a - \hat{n}_a \tag{2.1}$$

for some function K^0 also depending on the geodesic. This time, however, for reasons to become clear later, rather than propagating the null tetrad parallelly along the geodesic, we arrange only that

$$\kappa = \epsilon = 0. \tag{2.2}$$

Since the geodesics are to be hypersurface-orthogonal, we know that

$$\rho = \bar{\rho}, \tag{2.3}$$

and $\hat{k}_a$ is proportional to the gradient of some function u. Using the (partial) arbitrariness in the choice of tetrad and conformal factor we choose the proportionality constant to be one, so that

$$\hat{k}_a = \hat{\nabla}_a u$$

and

$$\hat{\tau} = \hat{\bar{\alpha}} + \hat{\beta}.$$

The parameter u labeling the null hypersurfaces and the conformal factor Ω will again be our first two coordinates. On a two-surface of intersection of a hypersurface $\Omega = \text{const} \neq 0$ and a hypersurface $u = \text{const}$, we select two coordinates x and y, or $\zeta = -x + iy$, such that $\hat{\bar{\delta}}\zeta = 0$, and propagate these along the geodesic. Therefore, our coordinates are

$$(\hat{x}^a) = (u, \Omega, x, y) \quad (a = 1,2,3,4).$$

Again,

$$\hat{D} = \hat{f}\,\frac{\partial}{\partial\Omega}, \hat{\delta} = \hat{\omega}\,\frac{\partial}{\partial\Omega} + \hat{\xi}^i\,\frac{\partial}{\partial\hat{x}^i}$$

$$\hat{\Delta} = \frac{\partial}{\partial u} + \hat{U}\,\frac{\partial}{\partial\Omega} + \hat{X}^i\,\frac{\partial}{\partial\hat{x}^i} \quad (i = 3,4),$$

with

$$\hat{f} \to -1,\ \hat{\omega} \to 0,\ \hat{U} \to K^0 \text{ as } \Omega \to 0.$$

Having kept track, at each stage, of the remaining freedom in the choice of frame (i.e., of tetrad, conformal factor, and coordinate system) we find that we can

(i) relabel the null hypersurfaces $u = \text{const}$ by means of a transformation

$$u' = \gamma(u),$$

provided we follow this by a conformal rescaling with conformal factor $\theta = \dot{\gamma}$, and by a scale change with parameter $a = \theta^{1/2}$;

(ii) make spatial rotations with parameter ϕ depending on u, x, and y;

(iii) make a conformal change and a null rotation about $\hat{k}_a$ such that parameters θ and c approach 1 and 0, respectively, as $\Omega \to 0$;

(iv) change the x and y coordinates by a transformation

$$\zeta' = \zeta'(u, \zeta).$$

We shall assume that the null geodesics are not only geodesic and hypersurface orthogonal but also shear-free and expansion-free, i.e.,

$$\hat\sigma = 0,$$

$$\hat\rho = -\hat f \Omega^{-1}.$$

Note that Eq. (NP4.2b) [i.e., Eq. (4.2b) of Ref. 4] then implies that

$$\hat\Psi_0 = 0.$$

These conditions are invariant under freedoms (i)–(iv). However, under freedom (iii) the metric variable $\hat f$ changes and we can arrange that

$$\hat f = -1,$$

and hence,

$$\hat\rho = \Omega^{-1}.$$

This imposes the condition that in (iii) the conformal factor θ must have the form

$$\theta = (1 + R\Omega)^{-1},$$

where $R = R(u,x,y)$ is arbitrary. If $\hat\tau$ is finite this can now be used to make

$$\hat\tau = \hat{\bar\pi},$$

imposing the additional restriction on freedom (iii) that the parameter c for the null rotation satisfy

$$\frac{\partial \bar c}{\partial \Omega} = 2\hat\delta\theta - \bar c \Omega^{-1}\theta. \tag{2.4}$$

Proceeding as in Ref. 1 we can show that

$$\hat\omega = 0,$$

$$\hat\xi^i \frac{\partial}{\partial \hat x^i} = \Omega^{-1} P \nabla,$$

where P defined by

$$P(u,x,y) = \hat\xi^{30} = -i\hat\xi^{40},$$

can be made real by means of freedom (iii), and where

$$\nabla = -2\frac{\partial}{\partial \zeta} = \frac{\partial}{\partial \hat x^3} + i\frac{\partial}{\partial \hat x^4}.$$

Equation (2.4) can be solved for the parameter c; we find

$$c = -P(\bar\nabla R)\Omega(1 + \Omega R)^{-1}. \tag{2.5}$$

In summary, we have

$$\hat\kappa = \hat\epsilon = \hat\sigma = \hat\omega = \hat\Psi_0 = 0, \quad \hat\rho = \Omega^{-1}, \quad \hat{\bar\pi} = \hat\tau = \bar{\hat\alpha} + \hat\beta,$$

$$\hat f = -1, \hat\xi^i \frac{\partial}{\partial \hat x^i} = \Omega^{-1} P \nabla, U \to K^0 \text{ as } \Omega \to 0.$$

The remaining freedom in the choice of frame is

(i) a relabeling $u' = \gamma(u)$ of the null hypersurfaces, accompanied by a conformal change with $\theta = \dot\gamma$ and a scale change with $a^2 = \theta$;

(ii) a change of coordinates $\zeta' = \zeta'(u,\zeta)$, and a phase change with parameter $\phi(u,\zeta)$ subject to

$$e^{2i\phi}\frac{\partial \zeta'}{\partial \zeta} = \text{real};$$

(iii) a conformal change with conformal factor $\theta = (1 + R\Omega)^{-1}$ followed by a null rotation with parameter c given by Eq. (2.5), where $R(u,\hat x^3,\hat x^4)$ is arbitrary.

The remaining metric equations

$$\hat D\hat U = \hat\gamma + \bar{\hat\gamma}, \quad \hat D\hat X^i = 2\hat\tau\bar{\hat\xi}^i + 2\bar{\hat\tau}\hat\xi^i,$$

$$\hat\delta\hat U = \bar{\hat\nu}, \quad \hat\mu = \bar{\hat\mu},$$

$$\hat\delta\hat X^i - \hat\Delta\hat\xi^i = \bar{\hat\lambda}\bar{\hat\xi}^i + (\hat\mu - \hat\gamma + \bar{\hat\gamma})\hat\xi^i,$$

$$\bar{\hat\delta}\hat\xi^i - \hat\delta\bar{\hat\xi}^i = (\hat\beta - \bar{\hat\alpha})\bar{\hat\xi}^i + (\hat\alpha - \bar{\hat\beta})\hat\xi^i,$$

are now solved together with the transformation equation[1] for the Ricci scalar and the Ricci identities[4] [Eqs. (NP4.2)]. The calculation is tedious but straightforward. The results are:

$$
\begin{aligned}
&\hat\kappa = \hat\epsilon = \hat\sigma = \hat\omega = \hat\Psi_0 = \hat\Phi_{00} = 0,\\
&\hat\rho = \Omega^{-1}, \quad \hat{\bar\pi} = \hat\tau = \Omega^{-1}(\hat\tau^0 - 2\hat A),\\
&\hat\alpha = \Omega^{-1}(\hat\alpha^0 - \hat{\bar A}), \quad \hat\beta = \Omega^{-1}(\hat\beta^0 - \hat A),\\
&\hat\lambda = -\tfrac12\bar\nabla\hat X^0 + 2\Omega^{-1}\bar\nabla(P\bar{\hat\tau}^0) + 4\bar\nabla(P\bar B),\\
&\hat\nu = P\Omega^{-1}\bar\nabla K^0 - P\bar\nabla a^0 - \Omega P\bar\nabla U^0 - P\bar\nabla\hat H + \tfrac12 P\Big(\Omega\frac{\partial}{\partial\Omega}\bar\nabla\hat H + \Omega^{-1}\bar\nabla\hat F\Big),\\
&\hat\gamma = \tfrac12 a^0 + U^0\Omega + \tfrac12\hat H + \tfrac18(\nabla\hat X^0 - \bar\nabla\bar{\hat X}^0)\\
&\qquad + \tfrac12\Omega^{-1}[\bar\nabla(P\hat\tau^0) - \nabla(P\bar{\hat\tau}^0)] + \bar\nabla(P\hat B) - \nabla(P\bar B),\\
&\hat\mu = K^0\Omega^{-1} - U^0\Omega - \hat H + \tfrac12\Omega\frac{\partial\hat H}{\partial\Omega} + \tfrac12\Omega^{-1}\hat F + \hat\mu^*,\\
&\hat\xi^i\frac{\partial}{\partial\hat x^i} = P\Omega^{-1}\nabla, \quad \hat U = K^0 - a^0\Omega - U^0\Omega^2 - \hat H\Omega + \tfrac12\Omega^2\frac{\partial\hat H}{\partial\Omega} + \tfrac12\hat F,\\
&\hat X^i\frac{\partial}{\partial\hat x^i} = (-\tfrac12\hat X^0 + 2P\Omega^{-1}\bar{\hat\tau}^0 + 4P\bar B)\nabla + (-\tfrac12\bar X^0 + 2P\Omega^{-1}\hat\tau^0 + 4P\hat B)\bar\nabla, \qquad (2.6)\\
&\hat\Phi_{01} = \hat\Psi_1 - \Omega^{-2}\hat\tau^0 + 2\Omega^{-2}\hat A,\\
&\hat\Phi_{02} = \Omega^{-2}[-2P\nabla\hat A - \nabla(P\hat\tau^0) - (\hat\tau^0)^2 + 2\hat A(2\hat\tau^0 - 2\hat A - \nabla P)] + \Omega^{-1}[\tfrac12\nabla\bar X^0 - 4\nabla(P\hat B)],
\end{aligned}
$$

$$\hat\Phi_{11} = -\tfrac12\Omega^{-1}\hat\mu^* + 3\Lambda\Omega^{-2} + \tfrac12\Omega^{-1}\hat H - \tfrac12\frac{\partial\hat H}{\partial\Omega} + P^2\Omega^{-2}[\nabla(P^{-1}\bar{\hat A}) + \bar\nabla(P^{-1}\hat A)]$$
$$- \tfrac12 P^2\Omega^{-2}[\nabla(P^{-1}\bar{\hat\tau}^0) + \bar\nabla(P^{-1}\hat\tau^0)] + \Omega^{-2}[4\bar{\hat\tau}^0\hat A + 4\hat\tau^0\bar{\hat A} - 8\hat A\bar{\hat A} - 2\hat\tau^0\bar{\hat\tau}^0],$$
$$\hat\Phi_{12} = \Omega^{-1}P\nabla\hat\gamma - \hat\beta - \hat U\frac{\partial\hat\beta}{\partial\Omega} - \hat X^i\frac{\partial\hat\beta}{\partial\hat x^i} - \hat\mu\hat\tau + \hat\beta(\hat\gamma - \bar{\hat\gamma} - \hat\mu) - \hat\alpha\bar{\hat\lambda},$$
$$\hat\Phi_{22} = \Omega^{-1}P\nabla\hat\nu - \hat\mu - \hat U\frac{\partial\hat\mu}{\partial\Omega} - \hat X^i\frac{\partial\hat\mu}{\partial\hat x^i} - \hat\mu^2 - \hat\lambda\bar{\hat\lambda} - \hat\mu(\hat\gamma + \bar{\hat\gamma}) + \hat\tau\bar{\hat\nu} + 2\hat\nu\hat\beta,$$
$$\hat\Lambda = \Omega^{-2}\Lambda + U^0 + \tfrac12\Omega^{-1}(\hat H - \hat\mu^*),$$
$$\hat\Psi_2 = \Omega^{-2}[-2\Lambda - 2P^2\bar\nabla(P^{-1}\hat A) + 2\bar{\hat\tau}^0\hat A + 2\hat\tau^0\bar{\hat A} - 4\hat A\bar{\hat A} - \hat\tau^0\bar{\hat\tau}^0 + P^2\bar\nabla(P^{-1}\hat\tau^0)],$$
$$\hat\Psi_3 = \Omega^{-1}P\bar\nabla\hat\gamma - \hat\alpha - \hat U\frac{\partial\hat\alpha}{\partial\Omega} - \hat X^i\frac{\partial\hat\alpha}{\partial\hat x^i} + \Omega^{-1}\hat\nu - \hat\lambda(\hat\tau + \hat\beta) + \hat\alpha(\bar{\hat\gamma} - \bar{\hat\mu}) - \hat\gamma\hat\alpha,$$
$$\hat\Psi_4 = \Omega^{-1}P\bar\nabla\hat\nu - \hat\lambda - \hat U\frac{\partial\hat\lambda}{\partial\Omega} - \hat X^i\frac{\partial\hat\lambda}{\partial\hat x^i} - 2\hat\lambda\hat\mu + \hat\lambda(\bar{\hat\gamma} - 3\hat\gamma) + \hat\nu(2\hat\alpha + \bar{\hat\tau}),$$

where

$$\hat\mu^* = -a^0 - \dot P P^{-1} - \tfrac14(\nabla\hat X^0 + \bar\nabla\bar{\hat X}^0) + \tfrac12\hat X^0\nabla\ln P + \tfrac12\bar{\hat X}^0\bar\nabla\ln P + 2P^2[\nabla(P^{-1}\bar{\hat B}) + \bar\nabla(P^{-1}\hat B)] + \Omega^{-1}P^2[\nabla(P^{-1}\bar{\hat\tau}^0) + \bar\nabla(P^{-1}\hat\tau^0)],$$
$$K^0 = P^2[\bar\nabla(P^{-1}\hat\tau^0) + \nabla(P^{-1}\bar{\hat\tau}^0) + \nabla\bar\nabla\ln P] + \hat\tau^0\bar{\hat\tau}^0,$$
$$\hat\alpha^0 = \tfrac12(\bar{\hat\tau}^0 + \bar\nabla P),\ \hat\beta^0 = \tfrac12(\hat\tau^0 - \nabla P),$$
$$\hat X^0 = -\hat X^{30} + i\hat X^{40},$$
$$\hat A(u,\Omega,\hat x^i) = \int\Omega\hat\Psi_1\,d\Omega, \quad \hat B(u,\Omega,\hat x^i) = \int\Omega^{-2}\hat A\,d\Omega,$$
$$\hat F(u,\Omega,\hat x^i) = -4P^2[\bar\nabla(P^{-1}\hat A) + \nabla(P^{-1}\bar{\hat A})] + 8\hat A\bar{\hat A} - 4\bar{\hat\tau}^0\hat A - 4\hat\tau^0\bar{\hat A} - 12\Lambda,$$
$$\frac{\partial^2\hat H}{\partial\Omega^2} = -\Omega^{-2}\frac{\partial\hat F}{\partial\Omega}.$$

The integration constants are so chosen that, for example $\hat A$ vanishes when $\hat\Psi_1$ does.

The variables $P, \hat X^0, a^0, \hat\tau^0, U^0$ are functions of u, x, and y, whereas Λ and $\hat\Psi_1$ are functions of all four coordinates. The quantities $\hat\Phi_{12}$, $\hat\Phi_{22}$, $\hat\Psi_3$, and $\hat\Psi_4$ are not given explicitly since the resultant expressions are rather lengthy.

The nonphysical metric is given by

$$d\hat s^2 = [2\hat U - \tfrac12 P^{-2}\Omega^2((\hat X^3)^2 + (\hat X^4)^2)]\,du^2 - 2\,du\,d\Omega + P^{-2}\Omega^2[\hat X^3 du\,d\hat x^3 + \hat X^4 du\,d\hat x^4 - \tfrac12((d\hat x^3)^2 + (d\hat x^4)^2)], \tag{2.7}$$

with the metric variables $\hat U$ and $\hat X^i$ given by Eq. (2.6).

The solution in physical space-time is now readily obtained. We adopt coordinates $(x^a) = (u, r, x, y)$, where $x^a = \hat x^a$ for $a = 1,3,4$ and $r = \Omega^{-1}$. The results are as follows.

$$\kappa = \sigma = \rho = \epsilon = \omega = \Psi_0 = 0, \quad \alpha = \hat\alpha^0 - \bar A,$$
$$\beta = \hat\beta^0 - A, \quad \bar\pi = \tau = \hat\tau^0 - 2A,$$
$$\lambda = -\tfrac12\bar\nabla\bar{\hat X}^0 + 2r\bar\nabla(P\bar{\hat\tau}^0) + 4\bar\nabla(P\bar B),$$
$$\nu = Pr^2\bar\nabla K^0 - Pr\bar\nabla a^0 - P\bar\nabla U^0 - Pr^2\bar\nabla\int Hr^{-2}\,dr,$$
$$\mu = -P^{-1}\dot P - \tfrac14(\nabla\hat X^0 + \bar\nabla\bar{\hat X}^0) + \tfrac12\hat X^0\nabla\ln P + \tfrac12\bar{\hat X}^0\bar\nabla\ln P + 2P^2[\nabla(P^{-1}\bar B) + \bar\nabla(P^{-1}B)] + rP^2[\nabla(P^{-1}\bar{\hat\tau}^0) + \bar\nabla(P^{-1}\hat\tau^0)],$$
$$\gamma + \bar\gamma = 2K^0 r - a^0 + \int F\,dr,$$
$$\gamma - \bar\gamma = \tfrac14(\nabla\hat X^0 - \bar\nabla\bar{\hat X}^0) + r[\bar\nabla(P\hat\tau^0) - \nabla(P\bar{\hat\tau}^0)] + 2[\bar\nabla(PB) - \nabla(P\bar B)],$$
$$\xi^i\frac{\partial}{\partial x^i} = P\nabla,$$
$$U = -K^0r^2 + a^0r + U^0 + r^2\int Hr^{-2}\,dr,$$
$$X^i\frac{\partial}{\partial x^i} = (-\tfrac12\hat X^0 + 2rP\bar{\hat\tau}^0 + 4P\bar B)\nabla + (-\tfrac12\bar{\hat X}^0 + 2rP\hat\tau^0 + 4PB)\bar\nabla,$$
$$\Psi_2 = -2\Lambda - 2P^2\bar\nabla(P^{-1}A) + P^2\bar\nabla(P^{-1}\hat\tau^0) + 2\bar{\hat\tau}^0A + 2\hat\tau^0\bar A - 4A\bar A - \hat\tau^0\bar{\hat\tau}^0,$$
$$\Psi_3 = P\bar\nabla\gamma - \dot\alpha - U\frac{\partial\alpha}{\partial r} - X^i\frac{\partial\alpha}{\partial x^i} - \lambda(\tau + \beta) + \alpha(\bar\gamma - \gamma - \bar\mu),$$
$$\Psi_4 = P\bar\nabla\nu - \dot\lambda - U\frac{\partial\lambda}{\partial r} - X^i\frac{\partial\lambda}{\partial x^i} - 2\lambda\mu + \lambda(\bar\gamma - 3\gamma) + \nu(2\alpha + \bar\tau), \tag{2.8}$$

and

$$\Phi_{00} = 0, \quad \Phi_{01} = \Psi_1,$$
$$\Phi_{02} = \nabla(P\hat\tau^0) - (\hat\tau^0)^2 - 2\nabla(PA) + 4\hat\tau^0A - 4A^2,$$
$$\Phi_{11} = P^2\nabla\bar\nabla\ln P + \tfrac12P^2[\nabla(P^{-1}\bar{\hat\tau}^0) + \bar\nabla(P^{-1}\hat\tau^0)] - \hat\tau^0\bar{\hat\tau}^0 + 2\hat\tau^0\bar A + 2\bar{\hat\tau}^0A - 4A\bar A,$$
$$\Phi_{12} = P\nabla\gamma - \dot\beta - U\frac{\partial\beta}{\partial r} - X^i\frac{\partial\beta}{\partial X^i} - \tau\mu + \beta(\gamma - \bar\gamma - \mu) - \alpha\bar\lambda,$$
$$\Phi_{22} = P\nabla\nu - \dot\mu - U\frac{\partial\mu}{\partial r} - X^i\frac{\partial\mu}{\partial x^i} - \mu^2 - \lambda\bar\lambda - \mu(\gamma + \bar\gamma) + \tau\bar\nu + 2\nu\beta, \tag{2.9}$$

where

$$\hat\alpha^0 = \tfrac12(\bar{\hat\tau}^0 + \bar\nabla P), \quad \hat\beta^0 = \tfrac12(\hat\tau^0 - \nabla P),$$
$$A(u,r,x^i) = -\int\Psi_1\,dr, \ B(u,r,x^i) = -\int A\,dr,$$

$$K^0 = \hat{\tau}^0\hat{\bar{\tau}}^0 + P^2[\bar{\nabla}(P^{-1}\hat{\tau}^0) + \nabla(P^{-1}\hat{\bar{\tau}}^0) + \nabla\bar{\nabla}\ln P],$$

$$F(u,r,x^i) = -4P^2[\bar{\nabla}(P^{-1}A) + \nabla(P^{-1}\bar{A})] + 8A\bar{A} - 4\hat{\bar{\tau}}^0 A - 4\hat{\tau}^0\bar{A} - 12\Lambda,$$

$$H(u,r,x^i) = \int C\,dr, \quad C(u,r,x^i) = r^{-2}\int r^2\frac{\partial F}{\partial r}\,dr.$$

Again, the integration constants are chosen so that, for example, A vanishes when Ψ_1 does; the variables $P, \hat{X}^0, a^0, \hat{\tau}^0, U^0$ are functions of u, x, and y, whereas the variables Λ and Ψ_1 are arbitrary functions of all four coordinates. The frame is not parallelly propagated since $\pi \neq 0$. Parallel propagation could be achieved by means of a null rotation with $c = -r\hat{\bar{\tau}}^0 - 2\bar{B}$, but most expressions would then become more complicated.

Since $\hat{g}_{ab} = \Omega^2 g_{ab}$ the metric is easily derived from Eq. (2.7). It is

$$-ds^2 = [2U + \tfrac{1}{2}P^{-2}((X^3)^2 + (X^4)^2)]\,du^2 - 2\,du\,dr - P^{-2}[X^3\,du\,dx^3 + X^4\,dy\,dx^4 - \tfrac{1}{2}((dx^3)^2 + (dx^4)^2)],$$

where U and X^i are given in Eq. (2.8).

The tetrad-coordinate freedom under which the above solution is invariant is given by

(i) $u' = \gamma(u)$, $r' = \dot{\gamma}^{-1}r$

accompanied by a scale change with parameter $a = \dot{\gamma}^{\frac{1}{2}}$;

(ii) $\zeta' = \zeta'(u,\zeta)$

accompanied by a phase change with parameter $\phi(u,\zeta,\bar{\zeta})$ subject to $e^{2i\phi}\,\partial\zeta'/\!\!-\!\!\zeta$ being real;

(iii) $r' = r + R(u, x^3, x^4)$

followed by a null rotation with parameter $c = -P\bar{\nabla}R$.

III. EINSTEIN-MAXWELL FIELDS

To get explicit solutions the reduced gravitational field equations must be solved together with the equations of the source. The former are obtained by equating the components of the Ricci tensor, as given by Λ and Eqs. (2.9), to the appropriate components of the energy-momentum tensor for the source. In the case of Einstein-Maxwell fields these reduced equations are

$$\Lambda = 0, \quad \Phi_{mn} = \phi_m\bar{\phi}_n \quad (m, n = 0,1,2), \tag{3.1}$$

where the ϕ_m and Φ_{mn} are the tetrad components of the electromagnetic field and the trace-free Ricci tensor, respectively. The source equations are Maxwell's equations, namely

$$\begin{aligned} &D\phi_1 = 0,\\ &D\phi_2 = P\bar{\phi}_1 + 2\hat{\tau}^0\phi_1,\\ &P\nabla\phi_1 = 2\hat{\tau}^0\phi_1,\\ &P\nabla\phi_2 = \dot{\phi}_1 + X^i\phi_{1,i} + 2\mu\phi_1 + (\nabla P)\,\phi_2. \end{aligned} \tag{3.2}$$

From Eqs. (2.9) and (3.1) it can be seen that ϕ_0 and Ψ_1 vanish. This means that the Maxwell tensor is "aligned" with the Weyl tensor and that the latter is algebraically special. The radial Maxwell equations, i.e., the first two of Eqs. (3.2), are readily integrated and give

$$\phi_1 = \phi_1^0(u,\zeta,\bar{\zeta}),$$
$$\phi_2 = \phi_2^0(u,\zeta,\bar{\zeta}) + r[P\bar{\nabla}\phi_1^0 + 2\hat{\bar{\tau}}^0\phi_1^0].$$

Apart from a redundant equation, the last two of Eqs. (3.2) become

$$\begin{aligned} &P\nabla\phi_1^0 = \hat{\tau}^0\phi_1^0,\\ &P^2\nabla(P^{-1}\phi_2^0) = \dot{\phi}_1^0 - \tfrac{1}{2}\hat{X}^0\bar{\nabla}\phi_1^0 - P^{-1}\hat{\tau}^0\hat{X}^0\phi_1^0. \end{aligned} \tag{3.3}$$

These two equations, together with the reduced gravitational field equations

$$\nabla(P\hat{\tau}^0) - (\hat{\tau}^0)^2 = 0,$$
$$P^2\nabla\bar{\nabla}\ln P + \tfrac{1}{2}P^2[\nabla(P^{-1}\hat{\bar{\tau}}^0) + \bar{\nabla}(P^{-1}\hat{\tau}^0)] - \hat{\tau}^0\hat{\bar{\tau}}^0 = \phi_1^0\bar{\phi}_1^0, \tag{3.4}$$

$$\Phi_{12}^0 = \phi_1^0\bar{\phi}_2^0, \quad \Phi_{12}^{(1)} = \phi_1^0(P\nabla\bar{\phi}_1^0 + 2\hat{\tau}^0\bar{\phi}_1^0),$$
$$\Phi_{22}^0 = \phi_2^0\bar{\phi}_2^0$$
$$\Phi_{22}^{(1)} = 2\mathrm{Re}[\phi_2^0(P\nabla\bar{\phi}_1^0 + 2\hat{\tau}^0\bar{\phi}_1^0)],$$
$$\Phi_{22}^{(2)} = |P\bar{\nabla}\phi_1^0 + 2\hat{\bar{\tau}}^0\phi_1^0|^2,$$

where Φ_{12} and Φ_{22} are given by Eqs. (2.9), just be solved for the variables $P, \hat{X}^0, \hat{a}^0, \hat{\tau}^0$, and U^0.

We can reduce the complexity of these equations considerably if we assume that $\hat{\tau}^0$ vanishes. In this case we have seen[1] that by means of freedom (iii) we can arrange for the spin-coefficient μ to vanish. Freedom (iii) is now restricted by $\nabla\bar{\nabla}R = 0$, and there is an additional equation that must be satisfied, namely

$$\dot{P} = -\tfrac{1}{4}P(\nabla\hat{X}^0 + \bar{\nabla}\hat{X}^0) + \tfrac{1}{2}\hat{X}^0\bar{\nabla}P + \tfrac{1}{2}\hat{\bar{X}}^0\nabla P. \tag{3.5}$$

The Weyl tensor is now of type III (or N or 0) since Ψ_2 vanishes as well Ψ_0 and Ψ_1. From Eqs. (3.3) and (3.4) we find that $\phi_1 = \phi_1(u,\bar{\zeta})$ is independent of ζ and that the following equations hold:

$$\begin{aligned} &P^2\nabla(P^{-1}\phi_2^0) = \dot{\phi}_1^0 - \tfrac{1}{2}\hat{X}^0\bar{\nabla}\phi_1^0,\\ &P^2\bar{\nabla}\nabla\ln P = \phi_1^0\bar{\phi}_1^0,\\ &-\tfrac{1}{2}P\nabla a^0 - \tfrac{1}{4}P\bar{\nabla}\nabla\hat{X}^0 + \tfrac{1}{2}(\bar{\nabla}P)\nabla\hat{X}^0 = \phi_1^0\bar{\phi}_2^0,\\ &P^2\nabla\bar{\nabla}U^0 + \tfrac{1}{4}|\bar{\nabla}\hat{X}^0|^2 = -\phi_2^0\bar{\phi}_2^0,\\ &-P^2\nabla\bar{\nabla}a^0 = \bar{\phi}_2^0 P\bar{\nabla}\phi_1^0 + \phi_2^0 P\nabla\bar{\phi}_1^0. \end{aligned} \tag{3.6}$$

Redundant equations have been discarded.

If we assume further that the Maxwell field is null, i.e., that $\phi_1^0 = 0$, Eqs. (3.6) become

$$\begin{aligned} &\nabla(P^{-1}\phi_2^0) = 0, \quad \nabla\bar{\nabla}\ln P = 0, \quad \nabla\bar{\nabla}a^0 = 0,\\ &P\nabla a^0 + \tfrac{1}{2}P^3\bar{\nabla}(P^{-2}\nabla\hat{X}^0) = 0,\\ &P^2\nabla\bar{\nabla}U^0 + \tfrac{1}{4}|\bar{\nabla}\hat{X}^0|^2 = -\phi_2^0\bar{\phi}_2^0. \end{aligned} \tag{3.7}$$

The solution to Eqs. (3.5) and (3.7) is given by

$$P = \text{const}, \quad X = -\bar{h}(u,\zeta) + \zeta\frac{\partial h(u,\bar{\zeta})}{\partial\bar{\zeta}} + \frac{\partial k(u,\bar{\zeta})}{\partial\bar{\zeta}},$$

$$a^0 = \frac{\partial h(u,\bar{\zeta})}{\partial\bar{\zeta}} + \frac{\partial\bar{h}(u,\zeta)}{\partial\zeta}$$

$$-4P^2U^0 = |\Phi_2(u,\bar{\zeta})|^2 + \left|\bar{h} - \zeta\frac{\partial\bar{h}_i}{\partial\zeta}\right|^2 + \left|\frac{\partial k}{\partial\bar{\zeta}}\right|^2 + \frac{\partial\bar{h}}{\partial\zeta}\left(\bar{\zeta}\frac{\partial k}{\partial\bar{\zeta}} - k\right) + \frac{\partial h}{\partial\bar{\zeta}}\left(\zeta\frac{\partial\bar{k}}{\partial\zeta} - \bar{k}\right),$$

where

$$\phi_2^0(u,\bar{\zeta}) = \frac{\partial\Phi_2(u,\bar{\zeta})}{\partial\bar{\zeta}}.$$

The functions h, k, and Φ_2 are arbitrary functions of the coordinates u and $\bar{\zeta}$. The Weyl tensor component Ψ_3 is given by

$$\Psi_3 = -2P\frac{\partial^2 h}{\partial \bar{\zeta}^2}.$$

Therefore, if we are looking for a type N solution, the variable $h(u,\bar{\zeta})$ must have the form $h(u,\bar{\zeta}) = d(u)\bar{\zeta} + e(u)$. We can use the remaining freedom to make $d(u)$ and X vanish.

Hence, the type N solution is determined by

$$P = \text{const}, X = a^0 = 0,$$
$$-4P^2U^0 = |\Phi_2(u,\bar{\zeta})|^2 + |h(u)|^2.$$

Since the remaining Weyl tensor component is

$$\Psi_4 = \bar{\Phi}_2(u,\zeta)\frac{\partial \phi_2(u,\bar{\zeta})}{\partial \bar{\zeta}},$$

we see that for a vacuum space–time the solution becomes flat.

In conclusion we note the following Goldberg–Sachs[5] type

Theorem: Assume that an Einstein–Maxwell space–time has a geodesic, expansion-free, and twist-free null congruence. This congruence is shear-free if and only if the Weyl tensor is algebraically special.

The proof is simple. We have already seen above that the vanishing of the shear σ implies that the space–time is algebraically special. Conversely, if $\Psi_0 = \Psi_1 = 0$ (and $\rho = \epsilon = \kappa = 0$) it follows from Eq. (NP4. 2a) that

$$\sigma\bar{\sigma} + \Phi_{00} = 0.$$

Since the second term is nonnegative for an Einstein–Maxwell field, the shear σ must vanish.

ACKNOWLEDGMENT

This work was supported in part by a grant from the Natural Sciences/Engineering Research Council of Canada.

[1] G. Ludwig, J. Math. Phys. **21**, 2543 (1980).
[2] W. Kundt, Z. Phys. **163**, 77 (1961).
[3] R. Penrose, in *Battelle Rencontres* (Benjamin, New York, 1968).
[4] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[5] J. N. Goldberg and R. K. Sachs, Acta Phys. Pol. Suppl. **22**, 13 (1962).

The Palatini variational principle for the general Bergmann–Wagoner–Nordtvedt theory of gravitation

N. Van den Bergh
Department of Physics, Universitaire Instelling Antwerpen, Antwerpen, Belgium

(Received 1 April 1981; accepted for publication 1 May 1981)

The Hilbert variational principle for the Bergmann–Wagoner–Nordtvedt scalar–tensor theory of gravitation is presented in a general form, yielding the field equations in arbitrary units. Applying then the Palatini method of variation to this action integral, the connection can be shown to be an integrable Weyl connection, even if the matter part contains explicitly the connection coefficients. However, the resulting field equations always appear in the wrong unit system, or turn out to be wrong in a general gauge.

PACS numbers: 04.50. + h, 04.20.Cv

I. INTRODUCTION

It is well known that the field equations of the general theory of relativity may be obtained in two different ways. According to the Hilbert method of variation, one requires that the action integral

$$I=\int(-g^{ij}R_{ij}+16\pi GL_M)\sqrt{-g}\,d^4x \tag{1}$$

is stationary under arbitrary variations of the symmetric metric g_{ij}. Hereby is R_{ij} the Ricci tensor of the corresponding Levi-Civita connection, i. e., one assumes *a priori* that the connection is symmetric and that

$$g_{ij;k}=0. \tag{2}$$

In the Palatini method of variation, on the other hand, one varies, in (1), g_{ij} and Γ^k_{ij} independently, with R_{ij} the Ricci tensor of Γ^k_{ij}. Taking the connection to be a symmetric one, one deduces the metricity condition (2), as well as the correct set of field equations if, and only if,[1–3] the connection coefficients Γ^k_{ij} do not appear explicitly in L_M.

Both variational principles have also been analyzed[4] for the Brans–Dicke scalar–tensor theory of gravitation, which has been for a long time the standard alternative for general relativity. Applying the Palatini method to the Brans–Dicke action integral[5] (with the Brans–Dicke φ replaced by ϕ^2)

$$I=\int(-\phi^2g^{ij}R_{ij}-4\omega g^{ij}\phi_i\phi_j+16\pi L_M)\sqrt{-g}\,d^4x \tag{3}$$

it was found that

$$\Gamma^k_{ij}=\{{}^{\,k}_{ij}\}_g+2\delta^k_{(i}(\log\phi)_{,j)}-g_{ij}g^{kl}(\log\phi)_{,l} \tag{4}$$

and that the correct set of field equations (up to another coupling constant ω) only resulted when all covariant derivatives were written with respect to the Levi-Civita connection $\{{}^{\,k}_{ij}\}_g$. The same result was obtained for slightly more general scalar–tensor theories with a particular power-law coupling.[6,7]

Now we observe that (4) precisely states that the Γ^k_{ij} are the Christoffel symbols $\{{}^{\,k}_{ij}\}_u$ with respect to the metric

$$u_{ij}=\phi^2g_{ij}. \tag{5}$$

Hence, taking into account that, in (3), $R_{ij}=R_{ij}{}^{(u)}$ is the Ricci tensor for Γ^k_{ij}, one can rewrite (3) as

$$I=\int(-u^{ij}R_{ij}{}^{(u)}-4\omega u^{ij}\phi_i\phi_j/\phi^2+16\pi\phi^{-4}L_M)\times\sqrt{-u}\,d^4x. \tag{6}$$

But this is precisely the action integral in Brans and Dicke's[5] "unit-transformed version" of the theory, in which the gravitational constant is effectively constant and rest masses are varying. We shall refer to these units henceforth as Planck units. For this version it is no wonder that the Hilbert–Palatini methods both yield the same answer! In fact, (6) can be looked at as the standard general relativity action (1) with interacting matterfields ϕ and M, and in which—for the case of Palatini variation—the matter part does not explicitly depends upon the connection coefficients.[3] One is left now with the problem that Hilbert-variation of (3) gives the correct field equations, with R_{ij} the Ricci tensor for the connection in particle units, whereas Palatini variation requires R_{ij} to be the Ricci tensor for the connection in Planck units. Both methods are thus clearly inequivalent for the Brans–Dicke theory. Apart from this, it looks strange that the Planck unit system acquires here a peculiar status (being the unit system in which the two methods yield the same answer), whereas at first sight there is nothing fundamental about the connection in Planck units. On the contrary, in the axiomatic approach towards the structure of space–time,[8,9] the fundamental object one first constructs, is the Weyl connection $\bar\nabla$ (with coefficients $\bar\Gamma^k_{ij}$) whose geodesics are the paths of freely falling particles and light rays. If $\bar\nabla$ is assumed to be an integrable connection, there is a unique metric $\overline{g_{ij}}$ (up to a constant factor) for which $\bar\nabla\,\overline{g_{ij}}=0$; this metric determines the geometry in geodesic units or in particle units. When the gravitational constant is observed to depend on the space–time coordinates in this geometry, according to

$$\bar G\sim\phi^{-2} \tag{7}$$

with $\phi=\phi(x^\alpha)$, one defines the geometry in Planck units $g_{ij}{}^{(P)}$ through the conformal transformation

$$g_{ij}{}^{(P)}=\phi^2\,\overline{g_{ij}}. \tag{8}$$

As the gravitational "constant" is a coscalar of power $+2$ in the Dirac definition of a cotensor,[10] it follows from (7) and (8) that $G^{(P)}$ is effectively constant. One has then $\bar{\nabla}_k g_{ij}{}^{(P)} = 2g_{ij}{}^{(P)}(\log\phi)_k$ and $\nabla_k{}^{(P)}g_{ij}{}^{(P)} = 0$.

It would be preferable then to derive from a single variational principle the field equations of the corresponding scalar–tensor theory, as well as the integrable Weyl character of the $\bar{\Gamma}^k_{ij}$—rather than that of the $\Gamma^{k(P)}_{ij}$, which can only be constructed if the properties of the former are already known. Our plan is now as follows: In (2), we present the action integral for the general Bergmann–Wagoner–Nordtvedt scalar–tensor theory in a unit-independent (conformally invariant) form. Applying the Hilbert method of variation gives conformally invariant field equations, which reduce to the standard equations in particle units or to the unit-transformed equations in Planck units, according to the choices $\lambda = \text{const}$ or $\beta = \text{const}$ (λ and β being two coscalars of power -1). In (3) we generalize this action integral to include explicitly $\bar{\Gamma}^k_{ij}$-dependent terms in the matter part, but such that one reobtains the Hilbert integral when $\bar{\Gamma}^k_{ij} = \{^{\,k}_{ij}\}_{\bar{g}}$. It is shown then that, even with this generalization, the Palatini method always gives the field equations in the *wrong* unit system or gives the wrong equations, when a general gauge is used.

II. THE HILBERT METHOD

In particle units the Lagrangian density for the Bergmann–Wagoner–Nordtvedt scalar–tensor theory is given by[11–13]

$$\mathscr{L}_{\text{tot}} = \mathscr{L}_0 + 16\pi\mathscr{L}_M = -(\overline{-g})^{1/2}\phi^2\left[\bar{R} + 4\omega(\phi)\,\overline{g^{ij}}\phi_{,i}\phi_{,j}/\phi^2\right] + 16\pi\mathscr{L}_M. \tag{9}$$

Here ϕ is an in-scalar,[10] as it describes the behavior of the gravitational constant in a fixed unit system. We have put the cosmological term equal to zero, as it is of no relevance in the present investigations. A conformally invariant generalization of (9) is obtained by replacing all covariant derivatives by cocovariant derivatives.[10] These are constructed through the introduction of a coscalar λ of power -1 and the definition of the Weyl vector

$$\kappa_i = -\lambda_{,i}/\lambda = \tfrac{1}{4}\{^{\,n}_{ni}\}_g - \tfrac{1}{4}\{^{\,n}_{ni}\}_{\bar{g}}. \tag{10}$$

The Weyl geometry (g_{ij},κ_i) is then related to the Riemann geometry in particle units $\overline{g_{ij}}$ by

$$\overline{g_{ij}} = \lambda^2 g_{ij}. \tag{11}$$

Hence, from (9) one obtains

$$\mathscr{L}_{\text{tot}} = -\sqrt{-g}\,\phi^2\lambda^2[R^* + 4\omega(\phi)g^{ij}\phi_{,i}\phi_{,j}/\phi^2] + 16\pi\mathscr{L}_M, \tag{12}$$

in which now $\mathscr{L}_M$ depends on λ through its coupling with the metric.

As[10]

$$R^* = R - 6g^{ij}\kappa_{i;j} + 6g^{ij}\kappa_i\kappa_j, \tag{13}$$

we can rewrite (12) as

$$\begin{aligned}\mathscr{L}_{\text{tot}} = &-\sqrt{-g}\,\phi^2\lambda^2[R + 4\omega(\phi)g^{ij}\phi_{,i}\phi_{,j}/\phi^2 \\ &- 6g^{ij}\lambda_{,i}\lambda_{,j}/\lambda^2 - 12g^{ij}\lambda_{,i}\phi_{,j}/\lambda\phi\,] \\ &+ 16\pi\mathscr{L}_M\ \ (+\ \ \text{divergence}).\end{aligned} \tag{14}$$

Varying independently the variables g_{ij}, ϕ, and λ, one obtains the conformally invariant field equations[14]

$$\begin{aligned}R_{ij} - \tfrac{1}{2}Rg_{ij} = &-8\pi GT_{ij} - 2\left(\frac{\beta_{;ij}}{\beta} - g_{ij}\frac{\Box\beta}{\beta}\right) \\ &+ 4\left(\frac{\beta_{,i}\beta_{,j}}{\beta^2} - \frac{1}{4}g_{ij}\frac{\beta_{,n}\beta^{,n}}{\beta^2}\right) \\ &- [4\omega(\phi) + 6]\left(\frac{\phi_{,i}\phi_{,j}}{\phi^2} - \frac{1}{2}g_{ij}\frac{\phi_{,n}\phi^{,n}}{\phi^2}\right)\end{aligned} \tag{15}$$

and

$$\frac{\Box\phi}{\phi} - \left(1 - \frac{\phi}{2\omega(\phi) + 3}\frac{d\omega}{d\phi}\right)\frac{\phi_{,n}\phi^{,n}}{\phi^2} + 2\frac{\phi_{,n}\beta^{,n}}{\phi\beta} = \frac{4\pi GT}{2\omega(\phi) + 3}, \tag{16}$$

where we have written

$$G = \text{const}\cdot\beta^{-2} \quad \text{and} \quad \beta = \lambda\cdot\phi. \tag{17}$$

These equations are easily seen to reduce to the standard Bergmann–Wagoner–Nordtvedt equations in the gauge $\lambda = \text{const}$ (particle units), whereas in the gauge $\beta = \text{const}$ one obtains the Planck unit version of the same theory.[5]

III. THE PALATINI METHOD

Introducing the metric $g_{ij}{}^{(P)}$ in Planck units and the corresponding Christoffel symbols $\Gamma^{k(P)}_{ij}$, one rewrites (14) as

$$\begin{aligned}\mathscr{L}_{\text{tot}} = &-\sqrt{-g(P)}\,G^{-1} \\ &\times\{g^{ij(P)}R_{ij}{}^{(P)} + [4\omega(\phi) + 6]g^{ij(P)}\phi_{,i}\phi_{,j}/\phi^2\} \\ &+ 16\pi\mathscr{L}_M\ \ (+\ \ \text{divergence})\end{aligned} \tag{18}$$

with G^{-1} a constant and

$$\tfrac{1}{2}R_{ij}{}^{(P)} = \Gamma^{n}_{i[n,j]}{}^{(P)} + \Gamma^{m}_{i[n}{}^{(P)}\Gamma^{n}_{j]m}{}^{(P)}. \tag{19}$$

When $\mathscr{L}_M$ does not depend on the $\Gamma^{k(P)}_{ij}$, it is obvious[3] that Palatini variation of (18)—with $g_{ij}{}^{(P)}$ and $\Gamma^{k(P)}_{ij}$ as independent field variables—yields the correct set of field equations in Planck units and the metricity condition

$$\Gamma^{k(P)}_{ij} = \{^{\,k}_{ij}\}_{g^{(P)}}. \tag{20}$$

Let us see now whether it is possible to use as independent field variables other connection coefficients, such as $\bar{\Gamma}^k_{ij}$, which are of a more fundamental significance.

Using the fact[10] that R^*_{ij} is the Ricci tensor $\overline{R_{ij}}$ for the $\bar{\Gamma}^k_{ij}$, we will take the action density (12)

$$\mathscr{L}_{\text{tot}} = -\sqrt{-g}\,\phi^2\lambda^2[g^{ij}\,\overline{R_{ij}} + 4\omega(\phi)g^{ij}\phi_{,i}\phi_{,j}/\phi^2] + 16\pi\mathscr{L}_M \tag{21}$$

with

$$\tfrac{1}{2}\,\overline{R_{ij}} = \bar{\Gamma}^{n}_{i[n,j]} + \bar{\Gamma}^{m}_{i[n}\bar{\Gamma}^{n}_{j]m}. \tag{22}$$

Palatini variation would then be supposed to give us the result

$$g_{ij} = \lambda^{-2}\,\overline{g_{ij}} \quad \text{and} \quad \bar{\Gamma}^k_{ij} = \{^{\,k}_{ij}\}_{\bar{g}}. \tag{23}$$

This would make then $\lambda_{*i} = \lambda_{,i} + \lambda\kappa_i$ identically zero [cf. (10)]. Hence we are free to include in (21) terms of the form

$g^{ij}\phi_i\lambda_{*j}$ and $g^{ij}\lambda_{*i}\lambda_{*j}$, where λ_{*i} is *defined* now as

$$\lambda_{*i} = \lambda_i + \lambda\kappa_i \tag{24}$$

with

$$\kappa_i = \tfrac{1}{4}\{{}^{\,n}_{ni}\}_g - \tfrac{1}{4}\bar{\Gamma}^{\,n}_{ni}. \tag{25}$$

The most general conformally invariant first-order Lagrangian density, which reduces to (21) if (23) is valid, reads then as (with c and d arbitrary functions of ϕ)

$$\mathscr{L}_{\text{tot}} = -\sqrt{-g}\phi^2\lambda^2[g^{ij}\bar{R}_{ij} + 4\omega(\phi)g^{ij}\phi_i\phi_j/\phi^2 + 6cg^{ij}\phi_i\lambda_{*j}/\phi\lambda + 3dg^{ij}\lambda_{*i}\lambda_{*j}/\lambda^2] + 16\pi\mathscr{L}_M. \tag{26}$$

Varying g_{ij} yields the field equations (with $G=\lambda^{-2}\phi^{-2}$)

$$\bar{G}_{ij} = -8\pi GT_{ij} - 4\omega\left(\frac{\phi_i\phi_j}{\phi^2} - \frac{1}{2}g_{ij}\frac{\phi_n\phi^n}{\phi^2}\right) - 6c\left(\frac{\phi_{(i}\lambda_{*j)}}{\phi\lambda}, - \frac{1}{2}g_{ij}\frac{\phi^n\lambda_{*n}}{\phi\lambda}\right) - 3d\left(\frac{\lambda_{*i}\lambda_{*j}}{\lambda^2} - \frac{1}{2}g_{ij}\frac{\lambda^{*n}\lambda_{*n}}{\lambda^2}\right). \tag{27}$$

Variation of $\bar{\Gamma}^{\,k}_{ij}$, on the other hand, using

$$\delta(\lambda_{*i}), = \lambda\delta\kappa_i = -\tfrac{1}{4}\lambda\delta\bar{\Gamma}^{\,n}_{ni}, \tag{28}$$

gives

$$0 = \int\sqrt{-g}d^4x\{\lambda^2\phi^2g^{ij}[(\delta\bar{\Gamma}^{\,n}_{in})_{|j} - (\delta\bar{\Gamma}^{\,n}_{ij})_{|n}] - \tfrac{3}{2}\phi\lambda^2c\phi^m\delta\bar{\Gamma}^{\,m}_{nm} - \tfrac{3}{2}\phi^2\lambda d\lambda^{*m}\delta\bar{\Gamma}^{\,n}_{nm}\}$$
$$= \int d^4x[-(\lambda^2\phi^2g^{ij}\sqrt{-g})_{|j}\delta\bar{\Gamma}^{\,n}_{in} + (\lambda^2\phi^2g^{ij}\sqrt{-g})_{|n}\delta\bar{\Gamma}^{\,n}_{ij} - \tfrac{3}{2}\sqrt{-g}\lambda^2\phi^2(c\phi^n/\phi + d\lambda^{*n}/\lambda)\delta\bar{\Gamma}^{\,m}_{mn}]\ .$$

Writing $\mathscr{G}^{ij} = \sqrt{-g}g^{ij}$ and $\beta = \lambda\phi$, this yields

$$(\beta^2\mathscr{G}^{ij})_{|k} - (\beta^2\mathscr{G}^{n(i})_{|n}\delta^{j)}_k - \tfrac{3}{2}\beta^2(c\phi_n/\phi + d\lambda_{*n}/\lambda)\cdot\mathscr{G}^{n(i}\delta^{j)}_k = 0. \tag{29}$$

Contraction over j and k gives

$$(\beta^2\mathscr{G}^{in})_{|n} = -\tfrac{5}{2}(c\phi_n/\phi + d\lambda_{*n}/\lambda)\mathscr{G}^{in}, \tag{30}$$

which, when substituted in (29), yields (with $u_{ij} = \beta^2 g_{ij}$)

$$(1/\sqrt{-u})(\sqrt{-u}\,u^{ij})_{|k} + (c\phi_n/\phi + d\lambda_{*n}/\lambda)u^{n(i}\delta^{j)}_k = 0. \tag{31}$$

Now

$$(1/\sqrt{-u})(\sqrt{-u})_{|k} = (\log\sqrt{-u})_k - \bar{\Gamma}^{\,n}_{nk} = (\log\sqrt{-g})_k + 4(\log\beta)_k - \bar{\Gamma}^{\,n}_{nk} = 4[\kappa_k + (\log\beta)_k] \tag{32}$$

such that (31) becomes

$$u_{ij|k} = 4u_{ij}[\kappa_k + (\log\beta)_k] + c(\log\phi)_{(i}u_{j)k} + d(\log\lambda)_{(i}u_{j)k} + d\kappa_{(i}u_{j)k}. \tag{33}$$

Hence, using the identity

$$\{{}^{\,n}_{in}\}_u - \bar{\Gamma}^{\,n}_{in} = \tfrac{1}{2}u^{mn}u_{mn|i}, \tag{34}$$

one has

$$8[\kappa_i + (\log\beta)_i] = 2\{{}^{\,n}_{in}\}_u - 2\bar{\Gamma}^{\,n}_{in} = u^{mn}u_{mn|i} = 16[\kappa_i + (\log\beta)_i] + c(\log\phi)_i + d(\log\lambda)_i + d\kappa_i,$$

from which one solves

$$\kappa_i = -(\log\lambda)_i - [(8+c)/(8+d)](\log\phi)_i. \tag{35}$$

This shows that κ_i, and hence also $\bar{\Gamma}^{\,n}_{ni}$, is a gradient. Are $\bar{\Gamma}^{\,k}_{ij}$ now the coefficients of a Weyl connection (and hence of an *integrable* Weyl connection)?

Take therefore the substitution of (35) in (33): One obtains

$$u_{ij|k} = 4\frac{d-c}{8+d}[u_{ij}(\log\phi)_k - u_{ik}(\log\phi)_j - u_{jk}(\log\phi)_i],$$

which must take the form[10]

$$u_{ij|k} = u_{ij}\Psi_k \tag{36}$$

with Ψ some function of ϕ. This is clearly only possible (for nonconstant ϕ) if $c=d$.

But then

$$u_{ij|k} = 0$$

or

$$\bar{\Gamma}^{\,k}_{ij} = \{{}^{\,k}_{ij}\}_u,$$

and thus

$$u_{ij} = \overline{g_{ij}}.$$

But $u_{ij} = \beta^2\overline{g_{ij}}$, so one would conclude

$$g_{ij} = \beta^{-2}\overline{g_{ij}} \tag{37}$$

instead of (23)!

Note that in the particular case $c=d=-8$ the conclusion (35) is invalid. But then (33) takes the form

$$u_{ij|k} = u_{ij}\Psi_k - u_{ik}\Psi_j - u_{jk}\Psi_i$$

with

$$\Psi_i = 4[\kappa_i + (\log\beta)_i].$$

This is only of the form (36) if $\kappa_i = -(\log\beta)_i$, from which one obtains again

$$u_{ij|k} = 0 \quad\text{or}\quad g_{ij} = \beta^{-2}\overline{g_{ij}}.$$

Now, if (37) is valid, one can rewrite (26) as

$$\mathscr{L}_{\text{tot}} = -\sqrt{-\overline{g}}(\overline{g^{ij}}\ \overline{R_{ij}} + 4\omega\,\overline{g^{ij}}\phi_i\phi_j/\phi^2 + \cdots) + 16\pi\mathscr{L}_M, \tag{38}$$

which is clearly wrong if $\bar{\Gamma}^{\,k}_{ij}$ is taken to be the connection in particle units. Variation of (38) would only give the correct field-equations if the $\overline{g_{ij}}$ was the metric in Planck units!

From these considerations we conclude that the Palatini method is not able to generate the field equations of the general scalar–tensor theory in an arbitrarily chosen unit system (e. g., the particle units), when the method is applied to an action integral similar to the one used in the Hilbert method.

More precisely, the use of the Palatini method forces one necessarily to take the connection coefficients $\Gamma_{ij}^{k(P)}$ *in Planck units* as the relevant field variables. Repeating then the above reasoning with λ_{*i} replaced by β_{*i}, one can indeed obtain $\beta_{*i} = 0$ and $g_{ij} = g_{ij}{}^{(P)}$ as desired, but the resulting field equations will turn out to be wrong. For any choice of c and d, the last terms in (27) will drop out, and what is left are the field equations (15) in Planck units, with $4\omega(\phi) + 6$ replaced by $4\omega(\phi)$. The only way to obtain the same equations for both Hilbert method and Palatini method, is to take *a priori* the gauge $\beta =$ const and to use the action integral (18); the identity of the two results then becoming a trivial matter.[3]

ACKNOWLEDGMENTS

I thank Dr. D. K. Callebaut (Universitaire Instelling Antwerpen) for constructive suggestions on these topics. This work was supported by Nationaal Fonds voor Wetenschappelijk Onderzoek-België.

[1]C. Misner, K. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).
[2]J. Ray, Nuovo Cimento B **25**, 706 (1975).
[3]M. Tsamparlis, J. Math. Phys. **19**, 555 (1977).
[4]U. Lindström, Nuovo Cimento B **32**, 298 (1976).
[5]C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[6]E. R. Harrison, Phys. Rev. D **6**, 2077 (1972).
[7]U. Lindström, Nuovo Cimento B **35**, 130 (1976).
[8]J. Ehlers, F. A. E. Pirani, and A. Schild, *General Relativity*, edited by L. O'Raifeartaigh (Oxford U. P., Oxford, 1972), p. 63.
[9]N. M. J. Woodhouse, J. Math. Phys. **14**, 495 (1973).
[10]P. A. M. Dirac, Proc. R. Soc. Lond. A **333**, 403 (1973).
[11]P. G. Bergmann, Int. J. Theor. Phys. **1**, 25 (1968).
[12]R. V. Wagoner, Phys. Rev. D **1**, 3209 (1970).
[13]K. Nordtvedt, Jr., Astrophys. J. **161**, 1059 (1970).
[14]N. Van den Bergh, Gen. Rel. Grav. (to be published).

Coherent frequency modulation in Gaussian stochastic processes

Noboru Takimoto
Department of Engineering Science, Tohoku University, Sendai, Japan

(Received 23 December 1980; accepted for publication 1 May 1981)

A study is given of the stochastic properties of a system of classical harmonic oscillators, driven by a Gaussian fluctuation and with a periodically modulated frequency. It is shown that, when the condition for the parameter amplification is almost satisfied, a phase locking occurs between the modulation and oscillation, leading to a coherent motion of the system as a whole. The difference between this type of coherent motion and the one excited by a periodic external force is pointed out. The distribution function for the process is defined and is shown to satisfy a generalized Fokker–Planck equation. A discussion is also given of the conditional probability, and some of its characteristic properties are specified in connection with the phase-locking effects.

PACS numbers: 05.40. + j, 02.50.Ey, 02.50.Fz

I. INTRODUCTION

Many previous works have contributed to studying how a coherence of motion is lost by random frequency modulations,[1-10] but, to the author's knowledge, no work appears to have studied systematically how a coherence of motion is gained by coherent frequency modulations.

The fact that a coherent external force, exerted on a system of identical harmonic oscillators, undergoing independent motions with random initial phases, leads to a coherent motion of the system as a whole is, needless to say, due to the fact that the phase of the forced oscillation is uniquely related to that of the external force. The purpose of the present paper is to show that a coherent external frequency modulation also provides a method of exciting a coherent motion of such a system. It is well known that a proper frequency modulation leads to amplification of oscillation (the parametric amplification),[11] but it seems to be less well known that, at the same time, a definite phase relationship appears between oscillation and modulation. This phase-locking effect, as it were, plays an essential role in exciting a coherent motion by frequency modulations.

In Sec. II we derive basic equations of the problem, which are applied in Sec. III, to the analysis of the parametric amplification and phase locking. Section IV is devoted to the discussion of a frequency-modulated harmonic oscillator, driven by a Gaussian fluctuation, and the distribution function describing the statistical properties of the oscillator will be derived. Then in Sec. V we show that this function satisfies a generalized Fokker–Planck equation. Finally in Sec. VI it is shown that the effect of phase locking reveals itself most clearly in the form of the conditional probability. No nonlinear effects are considered throughout.

II. BASIC EQUATIONS

A harmonic oscillator with a modulated frequency is described typically by

$$\ddot{x} + 2\lambda\dot{x} + \omega_0^2[1 + h(t)]x = 0, \tag{2.1}$$

where ω_0 is the frequency of the free oscillation, λ a small damping factor, and $h(t)$ a small dimensionless function whose form depends on the conditions of the modulation. In terms of the two-component vector $X(t) = (x_1, x_2)$ with $x_1 = \omega_0 x$ and $x_2 = \dot{x}$, we have

$$\dot{X} = \begin{pmatrix} 0 & \omega_0 \\ -\omega_0 & -2\lambda \end{pmatrix} X - h(t)\omega_0 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} X. \tag{2.2}$$

It is convenient to change the variable from $X(t)$ to $Y(t)$ through

$$X = \begin{pmatrix} \cosh k & -\sinh k \\ -\sinh k & \cosh k \end{pmatrix} Y \equiv e^{-K} Y, \tag{2.3}$$

where we have put

$$K = \begin{pmatrix} 0 & k \\ k & 0 \end{pmatrix}, \quad \tanh k = \frac{\omega_0 - (\omega_0^2 - \lambda^2)^{1/2}}{\lambda}.$$

The authors in Ref. 10 used a somewhat different transformation, but the one in the present paper is more convenient. With (2.3) substituted into (2.2) and writing $\omega = (\omega_0^2 - \lambda^2)^{1/2}$, we have

$$\dot{Y} = \begin{pmatrix} -\lambda & \omega \\ -\omega & -\lambda \end{pmatrix} Y - h(t)\omega_0 e^{K} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} e^{-K} Y. \tag{2.4}$$

For later use, we further introduce the transformation

$$Y = e^{\Gamma t} V, \quad \Gamma = \begin{pmatrix} 0 & \gamma \\ -\gamma & 0 \end{pmatrix}$$

to derive

$$\dot{V}(t) = L_0 V(t) + L_1(t) V(t) \tag{2.5}$$

with

$$L_0 = \begin{pmatrix} -\lambda & -(\gamma - \omega) \\ \gamma - \omega & -\lambda \end{pmatrix}$$

and

$$L_1(t) = -h(t)\omega_0 e^{-\Gamma t} e^{K} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} e^{-K} e^{\Gamma t}. \tag{2.6}$$

The parameter γ may be determined in a suitable way relevant to the form of $h(t)$. The full expression for $L_1(t)$ can be easily calculated to be

$$L_1(t) = -\frac{h(t)\omega_0}{2} \begin{pmatrix} \sinh 2k - \cosh 2k \sin 2\gamma t & -\cosh 2k + \cos 2\gamma t + \sinh 2k \sin 2\gamma t \\ \cosh 2k + \cos 2\gamma t - \sinh 2k \sin 2\gamma t & -\sinh 2k + \cosh 2k \sin 2\gamma t \end{pmatrix}.$$

For very weak damping $\omega_0 \gg \lambda$ we can put $k \simeq 0$, and hence

$$L_1(t) \simeq \frac{h(t)\omega_0}{2}\begin{pmatrix} \sin 2\gamma t & 1-\cos 2\gamma t \\ -1-\cos 2\gamma t & -\sin 2\gamma t \end{pmatrix}. \quad (2.7)$$

In the absence of modulation, $h(t)=0$, we can put $\gamma=\omega$ without losing generality to derive $\dot{V} = -\lambda V$ with the obvious solution $V(t) = \exp(-\lambda t)V(0)$. In the original representation, this leads to the damped oscillation

$$X = e^{-\lambda t} e^{-K} e^{\Gamma t} e^{K} X(0) = e^{-\lambda t}\left\{\cos\omega t\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin\omega t\begin{pmatrix} -\sinh 2k & \cosh 2k \\ -\cosh 2k & \sinh 2k \end{pmatrix}\right\} X(0).$$

III. PERIODIC MODULATION

The form $h(t) = h\cos 2\gamma t$, $|h| \ll 1$, corresponds to a coherent and periodic modulation, in which case (2.7) is divided into two parts:

$$L_1(t) = \frac{-h\omega_0}{4}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \frac{h\omega_0}{4} \times \begin{pmatrix} \sin 4\gamma t & 2\cos 2\gamma t - \cos 4\gamma t \\ -2\cos 2\gamma t - \cos 4\gamma t & -\sin 4\gamma t \end{pmatrix}.$$

The first term is time-independent and gives rise to a secular effect on the mechanical state of oscillation. Hence it is natural to redefine L_0 and $L_1(t)$ in (2.5) in the following way (the renormalization of L_0):

$$L_0 = \begin{pmatrix} -\lambda & -\epsilon - (\gamma-\omega) \\ -\epsilon + (\gamma - \omega) & -\lambda \end{pmatrix}, \quad \epsilon = \frac{h\omega_0}{4}, \quad (3.1)$$

$$L_1(t) = \epsilon\begin{pmatrix} \sin 4\gamma t & 2\cos 2\gamma t - \cos 4\gamma t \\ -2\cos 2\gamma t - \cos 4\gamma t & -\sin 4\gamma t \end{pmatrix}. \quad (3.2)$$

With this redefinition, (2.5) can be solved by successive approximations:

$$V(t) = e^{L_0 t}\left\{1 + \int_0^t dt' L_1^*(t') + \int_0^t dt' \int_0^{t'} dt'' L_1^*(t') L_1^*(t'') + \cdots\right\} V(0) \quad (3.3)$$

with

$$L_1^*(t) = e^{-L_0 t} L_1(t) e^{L_0 t}. \quad (3.4)$$

The secular term $\exp(L_0 t)$ plays the most important role. The two eigenvalues of L_0, denoted by $s_\pm$, can be easily evaluated to be

$$s_\pm = -\lambda \pm [\epsilon^2 - (\gamma - \epsilon)^2]^{1/2}.$$

The oscillation evidently amplifies when $s_\pm$ is real and positive:

$$\epsilon^2 > \lambda^2 + (\gamma - \omega)^2. \quad (3.5)$$

This is the condition for the parametric amplification of the lowest order.[11]

The term linear in $L_1^*(t)$ in (3.3) yields small corrections to the fundamental oscillation as well as small higher harmonics. This can be shown by considering the simplest case $\gamma = \omega$ and $\lambda = 0$. The operator $\exp(L_0 t)$ then simplifies to

$$e^{L_0 t} = \begin{pmatrix} \cosh\epsilon t & -\sinh\epsilon t \\ -\sinh\epsilon t & \cosh\epsilon t \end{pmatrix},$$

yielding

$$L_1^*(t) = -\epsilon\cos 4\gamma t\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \epsilon(\cosh 2\epsilon t\sin 4\gamma t - \sinh 2\epsilon t\cos 2\gamma t) \times \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} - \epsilon(\sinh 2\epsilon t \sin 4\gamma t - \cosh 2\epsilon t\cos 2\gamma t)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

When substituted into (3.3), the first term of $L_1^*(t)$ gives

$$e^{L_0 t}\int_0^t L_1^*(t')\,dt' = \frac{\epsilon\sin 4\gamma t}{4\gamma}\begin{pmatrix} \sinh\epsilon t & -\cosh\epsilon t \\ -\cosh\epsilon t & \sinh\epsilon t \end{pmatrix}.$$

In the original representation, this is written as

$$X(t) = \frac{\epsilon}{2\gamma}\left[(\sin 3\gamma t + \sin 5\gamma t)\begin{pmatrix} \sinh\epsilon t & -\cosh\epsilon t \\ -\cosh\epsilon t & \sinh\epsilon t \end{pmatrix} + (\cos 3\gamma t - \cos 5\gamma t)\begin{pmatrix} -\cosh\epsilon t & \sinh\epsilon t \\ \sinh\epsilon t & -\cosh\epsilon t \end{pmatrix}\right]\cdot X(0).$$

The term retained thus yields higher harmonics of 3γ and 5γ of order h. In the same way, the terms proportional to $\sin 2\gamma t$ and $\cos 2\gamma t$ in $L_1^*(t)$ are shown to yield corrections to the fundamental oscillation as well as higher harmonics, each of order h. On the other hand, the terms nonlinear in $L_1^*(t)$ involve, in addition to still higher harmonics of order h^n, $n > 2$, higher-order secular effects (details are shown in Appendix A), which, however, do not affect the qualitative features of the following arguments.

The fundamental oscillation in the original representation is written as

$$X(t) = e^{-K} e^{\Gamma t} e^{L_0 t} e^{K}\cdot X(0) \equiv \beta(t)\cdot X(0). \quad (3.6)$$

This expression enables us to study the phase relationship between the modulation and oscillation. Take $\lambda \ll \omega_0$ and $\gamma = \omega$, for simplicity. Then $\exp(K)$ can be approximated by a unit matrix and $\exp(L_0 t)$ reduces to

$$e^{L_0 t} = e^{-\lambda t}\begin{pmatrix} \cosh\epsilon t & -\sinh\epsilon t \\ -\sinh\epsilon t & \cosh\epsilon t \end{pmatrix}.$$

If ϵ is positive (modification for $\epsilon < 0$ is trivial), we have for $\epsilon t \gg 1$

$$e^{L_0 t} \simeq \frac{1}{\sqrt{2}} e^{(\epsilon - \lambda)t}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

and (3.6) simplifies to

$$X(t) \simeq \frac{1}{\sqrt{2}} e^{(\epsilon - \lambda)t}[x_1(0) - x_2(0)]\begin{pmatrix} \cos(\omega t + \pi/4) \\ -\sin(\omega t + \pi/4) \end{pmatrix}. \quad (3.7)$$

Clearly, in this limit, the phase of oscillation is uniquely related to that of the modulation. Note that the relationship is independent of the sign of $|\epsilon| - \lambda$; i.e., it is the same whether the oscillation amplifies or damps.

The physical significance of this result can be visualized most easily, if we consider the motion of a simple pendulum of mass m and length l, whose point of support oscillates vertically according to the law $a \cos 2\gamma t$, $|a| \ll l$. If the angle between the string and vertical is taken to be x, then the equation of motion for this pendulum for small $|x|$ and for $\omega_0 \simeq \gamma$ is given by

$$\ddot{x} + 2\lambda \dot{x} + \omega_0^2 (1 + h \cos 2\gamma t) x = 0,$$

where $\omega_0^2 = g/l$, $h = 4a/l$, and we have introduced the damping proportional to $\dot{x}$. For the particular case $\gamma = \omega$ and $\omega_0 \gg \lambda$, the solution of this equation for $|h|\omega_0 t \gg 1$ approaches (3.7). The correspondence between the position of the point of support and that of the mass is shown in Fig. 1 in the case $\epsilon > 0$ and $x_1(0) > x_2(0)$. Three positions of the mass, A, B, and O, correspond to the neutral point N. To position A corresponds the phase $\omega t = 7\pi/4 \bmod 2\pi$, and to B the phase $3\pi/4$, whereas to the origin correspond the two phases $\pi/4$ and $5\pi/4$.

This phase-locking property of the modulation leads to the following remarkable consequence. Consider a system of a large number of identical harmonic oscillators, each undergoing independent oscillation. If the initial phases of the oscillators are random, the motion of the system as a whole is, of course, incoherent. Now, let the frequency of the oscillators be modulated coherently. It is clear that a proper modulation would lead to a coherent motion of the system as a whole. However, the resultant coherent motion is not such that all oscillators in the system move in unison with the same amplitude and phase. According to (3.7), the amplitude of an individual oscillator, even after a sufficient time, still depends on the initial condition of motion, and the phase also depends on the initial conditions, but the dependence is only upon the sign of $x_1(0) - x_2(0)$, i.e., the modulation determines the phase of oscillation within $\pm \pi$. These properties of the parametric amplification are in remarkable contrast with those of the coherent motion excited by a periodic external force, in which case, after a sufficient time, all oscillators move in unison independently of the initial conditions. It will be shown in Sec. VI that these features of the periodic frequency modulation have salient effects on the statistical properties of a harmonic oscillator, driven by a fluctuating force.

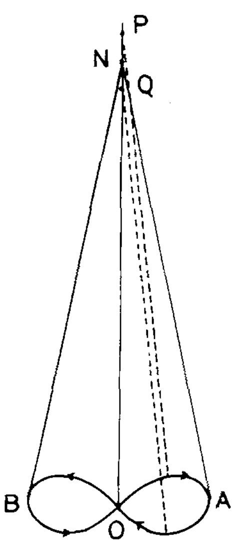


FIG. 1. A simple pendulum with the point of support oscillating vertically according to the law $a \cos 2\omega t$, $a > 0$. The sign has been fixed, so that Q may be its position at $\omega t = 0, \bmod \pi$. After a sufficient time, a definite correspondence appears between the phase of the mass and that of the point of support, as shown in the figure.

IV. MODULATION OF A HARMONIC OSCILLATOR DRIVEN BY A GAUSSIAN FLUCTUATION

Let a frequency-modulated harmonic oscillator be driven by a fluctuating force $f(t)$ with the stochastic properties

$$\langle f(t) \rangle = 0, \quad \langle f(t) f(t') \rangle = F(t - t'), \tag{4.1}$$

where $\langle\ \rangle$ denotes stochastic averaging. It is assumed that the stochastic process is stationary and Gaussian.

When the stochastic force is added, (2.2) is extended to

$$\dot{X}(t) = \begin{pmatrix} 0 & \omega_0 \\ -\omega_0 & -2\lambda \end{pmatrix} \cdot X(t) - h\omega_0 \cos 2\gamma t \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \cdot X(t) + f(t) \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The transformation $X(t) = e^{-K} e^{\Gamma t} \cdot V(t)$ yields

$$\dot{V}(t) = L_0 \cdot V(t) + L_1(t) \cdot V(t) + f(t) e^{-\Gamma t} e^{K} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{4.2}$$

where L_0 and $L_1(t)$ are defined by (3.1) and (3.2), respectively. If the nonsecular effects are ignored, the solution including the lowest order secular effects is

$$V(t) = e^{L_0 t} \left[V(0) + \int_0^t dt' f(t') e^{-L_0 t'} e^{-\Gamma t'} e^{K} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]. \tag{4.3}$$

In the original representation, this is written as

$$X(t) = \beta(t) \cdot [X(0) + A(t)], \tag{4.4}$$

where $\beta(t)$ is defined by (3.6), and

$$A(t) = \int_0^t dt' f(t') \beta^{-1}(t') \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{4.5}$$

If the initial value $X(0)$ is also a random variable with the probability distribution $P(X(0),0)$, then the average value of a quantity $Q(X(t))$ is given by

$$\langle\langle Q(x(t)) \rangle\rangle = \int dX(0) \langle P(X(0),0) Q(X(t)) \rangle. \tag{4.6}$$

Note that two averaging processes are involved in this expression. Change of the variable from $X(0)$ to $X(t) = \beta(t) \cdot [X(0) + A(t)]$ leads to

$$\langle\langle Q(X(t)) \rangle\rangle = \|\beta\|^{-1} \int dX(t) \times \langle P(\beta^{-1} \cdot X(t) - A(t), 0) Q(X(t)) \rangle.$$

Now define

$$P(X,t) = \|\beta\|^{-1} \langle P(\beta^{-1} \cdot X(t) - A(t), 0) \rangle. \tag{4.7}$$

Then we have

$$\langle\langle Q(X(t)) \rangle\rangle = \int dX\, P(X,t) Q(X). \tag{4.8}$$

Equation (4.8) permits the interpretation that $P(X,t)$ represents the distribution at t. The initial form (4.6) is based on the picture in which the dynamical variable $X(t)$ changes in time but the distribution function remains unchanged, whereas the final form (4.8) is based on the picture in which the converse is the case. Such a change of picture is introduced whenever a Fokker–Planck type of equation for the distribution function is derived from the Langevin equation.

V. A GENERALIZED FOKKER–PLANCK EQUATION

In this section, the differential equation for $P(X,t)$ will be derived for the δ-correlated case: $F(t-t') = 2F_0\delta(t-t')$. Extension to general cases is treated in Appendix B. In terms of the Fourier transform

$$P(X,t) = (2\pi)^{-2}\int d^2k\, e^{ik\cdot X}\psi(k), \tag{5.1}$$

(4.7) takes the form

$$P(X,t) = \|\beta\|^{-1}(2\pi)^{-2}\int d^2k\, \langle \exp[ik\cdot(\beta^{-1}\cdot X - A)]\rangle \psi(k). \tag{5.2}$$

Being a linear combination of the Gaussian variable, A is also Gaussian, so that

$$\langle \exp[ik\cdot A(t)]\rangle = \exp\{-\tfrac{1}{2}k\cdot\langle A(t)\tilde{A}(t)\rangle\cdot k\},$$

where $\tilde{}$ denotes transposing a vector or matrix. Now define a matrix $D(t)$ by

$$\begin{aligned} D(t) &= \langle A(t)\tilde{A}(t)\rangle \\ &= 2\int_0^t dt'\,\beta^{-1}(t')\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\tilde{\beta}^{-1}(t'). \end{aligned} \tag{5.3}$$

Evidently $D(t)$ is symmetric and positive-definite.

With these expressions substituted into (5.2), we obtain

$$\begin{aligned} P(X,t) &= \|\beta\|^{-1}(2\pi)^{-2}\int d^2k\, \exp(ik\cdot\beta^{-1}\cdot X) \\ &\quad\times \exp(-\tfrac{1}{2}\tilde{k}\cdot D\cdot k)\psi(k) \\ &= \|\beta\|^{-1}(2\pi)^{-2}\exp(\tfrac{1}{2}\tilde{\nabla}\cdot\beta D\tilde{\beta}\cdot\nabla)\int d^2k \\ &\quad\times \exp(i\tilde{k}\cdot\beta^{-1}\cdot X)\psi(k) \\ &= \|\beta\|^{-1}\exp(\tfrac{1}{2}\tilde{\nabla}\cdot\beta D\tilde{\beta}\cdot\nabla)P(\beta^{-1}\cdot X,0). \end{aligned}$$

Differentiation with respect to t leads to a Fokker–Planck type of equation

$$\begin{aligned} \frac{\partial P(X,t)}{\partial t} &= -\Bigg[\left(\frac{d}{dt}\|\beta\|\right)\Big/\|\beta\| \\ &\quad + \frac{1}{2}\frac{d}{dt}((\beta D\tilde{\beta})_{ij})\frac{\partial^2}{\partial x_i\partial x_j} \\ &\quad + \exp(\tfrac{1}{2}\tilde{\nabla}\cdot\beta D\tilde{\beta}\cdot\nabla)(\dot{\beta}^{-1})_{ij}x_j\beta_{ki}\frac{\partial}{\partial x_k} \\ &\quad \times \exp(-\tfrac{1}{2}\tilde{\nabla}\cdot\beta D\tilde{\beta}\cdot\nabla)\Bigg]P(X,t) \\ &= (\tfrac{1}{2}\tilde{\nabla}\cdot\beta\dot{D}\tilde{\beta}\cdot\nabla - \tilde{\nabla}\cdot\dot{\beta}\beta^{-1}\cdot X)P(X,t), \end{aligned} \tag{5.4}$$

together with the identity $[(d/dt)\|\beta\|]/\|\beta\| = \dot{\beta}_{ij}(\beta^{-1})_{ji}$. It is easily seen that

$$\beta(t)\dot{D}(t)\tilde{\beta}(t) = 2F_0\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$$

and

$$\dot{\beta}\beta^{-1} = e^{-K}(\Gamma + e^{\Gamma t}L_0e^{-\Gamma t})e^{K}. \tag{5.5}$$

Equation (5.4) thus describes a nonstationary Markovian, Gaussian process.[12]

In the absence of modulation, $\epsilon = 0$, we can put $\gamma = \omega$ to obtain

$$\dot{\beta}\beta^{-1} = \begin{pmatrix}0 & \omega_0\\ -\omega_0 & 2\lambda\end{pmatrix}. \tag{5.6}$$

Equation (5.4) then reduces to the familiar form

$$\begin{aligned} \frac{\partial P(X,t)}{\partial t} &= \left[F_0\frac{\partial^2}{\partial x_2^2} - \omega_0\left(x_2\frac{\partial}{\partial x_1} - x_1\frac{\partial}{\partial x_2}\right)\right. \\ &\quad \left. + 2\lambda\frac{\partial}{\partial x_2}x_2\right]P(X,t), \end{aligned} \tag{5.7}$$

which describes the Brownian motion of a simple harmonic oscillator.[13]

VI. THE CONDITIONAL PROBABILITY

The solution of (5.4) for the special initial condition $P(X,0) = \delta(X - X_0)$ is called the conditional probability, denoted by $P(X_0,0;X,t)$. Substituting $\psi(k) = \exp(-i\tilde{k}\cdot X_0)$ into (5.2) and then evaluating the integral, we obtain the conditional probability

$$\begin{aligned} P(X_0,0;X,t) &= \|\beta\|^{-1}(2\pi)^{-2}\int d^2k \\ &\quad\times \exp[-\tfrac{1}{2}\tilde{k}\cdot D\cdot k + i\tilde{k}\cdot(\beta^{-1}\cdot X - X_0)] \\ &= (2\pi)^{-1}\|D_\beta(t)\|^{-1/2} \\ &\quad\times \exp[-\tfrac{1}{2}(X - \beta\cdot X_0)^{\sim}\cdot D_\beta^{-1}(t)\cdot(X - \beta\cdot X_0)] \end{aligned} \tag{6.1}$$

with the symmetric and positive definite matrix

$$D_\beta(t) = \beta(t)D(t)\tilde{\beta}(t). \tag{6.2}$$

The distribution (6.1) has a center at $\beta(t)\cdot X_0$, which is the solution of the equation of motion in the absence of the stochastic force. It starts from the initial value X_0 and moves along a spiral, which either converges at the origin or diverges at infinity, according as $\lambda^2 > \epsilon^2$ or $\lambda^2 < \epsilon^2$ (the divergence is, of course, suppressed if nonlinear effects are included). The properties of the distribution are determined completely by the matrix $D_\beta(t)$ or its inverse. Its full expression is

$$D_\beta(t) = 2F_0\beta(t)\int_0^t dt'\,\beta^{-1}(t')\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\tilde{\beta}^{-1}(t')\tilde{\beta}(t). \tag{6.3}$$

Evaluation of $D_\beta(t)$ is carried out in Appendix C under various physical conditions, and the results will be cited in what follows.

For $\epsilon = 0$, we have

$$\begin{aligned} D_\beta(t) &= \frac{F_0}{2\lambda}\left\{\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \exp\left[\begin{pmatrix}0 & \omega_0\\ -\omega_0 & -2\lambda\end{pmatrix}t\right]\right. \\ &\quad \left.\times\exp\left[\begin{pmatrix}0 & -\omega_0\\ \omega_0 & -2\lambda\end{pmatrix}t\right]\right\}. \end{aligned} \tag{6.4}$$

For $\lambda t \gg 1$, this reduces to

$$D_\beta(t) = \frac{F_0}{2\lambda}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{6.5}$$

and (6.1) gives the stationary distribution

$$P(X_0,0;X,t) = (\lambda/\pi F_0)\exp[-(\lambda/F_0)(x_1^2+x_2^2)]. \quad (6.6)$$

The origin of the damping and fluctuating force have not been specified so far. If they are caused by a heat reservoir in thermal equilibrium into which the oscillator is immersed, then (6.6) provides the well-known fluctuation-dissipation relation $F_0/\lambda = 2k_B T$, where k_B and T are the Boltzmann constant and the temperature of the reservoir, respectively.

For $\epsilon \neq 0$, three cases can be considered. For $\lambda^2 > \epsilon^2$, the center of the distribution converges at the origin, and its asymptotic form is given by

$$\begin{aligned} P(X_0,0;X,t) = {} & [(\lambda^2-\epsilon^2)^{1/2}/\pi F_0] \\ & \times \exp\{-(1/F_0)[(\lambda+\epsilon\sin 2\omega t)x_1^2 \\ & + (\lambda-\epsilon\sin 2\omega t)x_2^2 + 2\epsilon\cos 2\omega t\, x_1x_2]\}. \end{aligned} \quad (6.7)$$

It is convenient to define the variables ξ and η along the principal axes by

$$\begin{pmatrix}\xi\\ \eta\end{pmatrix} = \begin{pmatrix}\cos(\omega t+\pi/4) & -\sin(\omega t+\pi/4)\\ \sin(\omega t+\pi/4) & \cos(\omega t+\pi/4)\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}. \quad (6.8)$$

Note that those states represented by (3.7) are on the ξ axis. In terms of the new variables, (6.7) takes the form

$$\begin{aligned} P(X_0,0;X,t) = {} & [(\lambda^2-\epsilon^2)^{1/2}/\pi F_0]\exp\{-(1/F_0)[(\lambda-\epsilon)\xi^2 \\ & + (\lambda+\epsilon)\eta^2]\}. \end{aligned} \quad (6.9)$$

The distribution is now not isotropic; the equation $P(X_0,0;X,t) = \text{const}$ defines in the phase space an ellipse, hereafter referred to as the distribution ellipse, whose principal axes rotate with the angular velocity ω (Fig. 2). This remarkable feature reminds us of the discussion in Sec. III on the phase-locking behavior of the modulation. There we have seen that if the modulation continues indefinitely, there appears a definite phase relationship between the modulation and oscillation. For $\omega t = 3\pi/4$, for example, the oscillation tends to take the phase A or B (see Fig. 1), accordingly as $x_1(0) - x_2(0)$ is negative or positive. Of course, such a rigorous phase locking cannot be expected to occur for the Brownian motion, but it is plausible that for $\omega t = 3\pi/4$, mod2π, the mechanical states A and B have higher probability than any other states. This is the physical implication of the distribution ellipses drawn in Fig. 2.

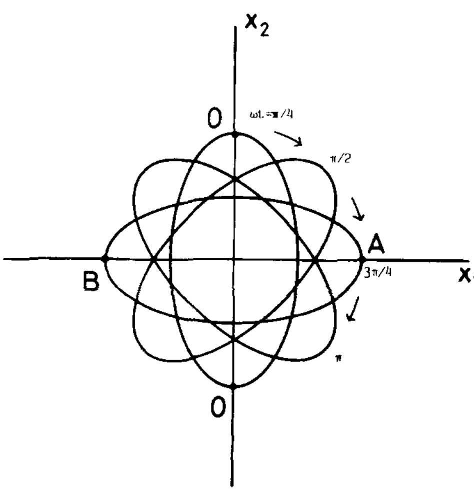


FIG. 2. The distribution ellipses in the case $\epsilon^2 < \lambda^2$, showing the rotation of the principal axes. As $|\epsilon|$ approaches λ, the major semiaxis becomes much larger than the minor semiaxis, indicating a continuous onset of the coherent motion associated with the phase locking. It is easily seen that the coherent motion is not such that all oscillators move in unison with the same amplitude and phase.

It should be pointed out that a similar rotation of the distribution ellipses has been reported already in 1945 in the case of a simple harmonic oscillator.[13] In this case the rotation occurs transiently; namely, an initial two-dimensional δ function in the phase space first becomes a narrow ellipse elongated in the x_2 direction, which then rotates with angular velocity ω and at the same time broadens out until it approaches the Maxwell–Boltzmann distribution centered at the origin.

For $\epsilon^2 = \lambda^2$, the center of the distribution, $\beta(t)\cdot X_0$, approaches a circle defined by (3.7). Writing $X' = X - \beta(t)\cdot X_0$, and then defining (ξ',η') related to X' by the same transformation as (6.9), we find for $\lambda t \gg 1$

$$\begin{aligned} & P(X_0,0;X,t) \\ & = \frac{1}{\pi F_0}\left(\frac{\lambda}{t}\right)^{1/2}\exp\left[-\frac{2\lambda}{F_0}\left(\frac{\xi'^2}{4\lambda t}+\eta'^2\right)\right], \quad \epsilon > 0. \end{aligned} \quad (6.10)$$

The phase-locking effect reveals itself in a more enhanced form in the fact that the major semiaxis is along the ξ' direction and increases as $\sqrt{t}$.

Finally for $\epsilon^2 > \lambda^2$, the center moves along an expanding spiral, and we find

$$\begin{aligned} & P(X_0,0;X,t) \\ & = \frac{(\epsilon^2-\lambda^2)^{1/2}}{\pi F_0}e^{-(\epsilon-\lambda)t} \\ & \quad \times \exp\left[-\frac{1}{F_0}\left((\epsilon-\lambda)e^{-2(\epsilon-\lambda)t}\xi'^2 + (\epsilon+\lambda)\eta'^2\right)\right], \end{aligned} \quad (6.11)$$

provided $\epsilon > 0$. The phase-locking effect is still more enhanced, as it should be.

The conditional probability averaged over the Maxwell–Boltzmann distribution is also of some interest; we assume that the initial state of the system has the distribution

$$P(X_0,0) = (\|E\|^{1/2}/2\pi)\exp(-\tfrac{1}{2}\tilde{X}_0\cdot E\cdot X_0), \quad E = (m/2k_B T)I,$$

where m is the mass of the oscillator and I a unit matrix. Then it is easily found that

$$\begin{aligned} P(X,t) &= \int P(X_0,0)P(X_0,0;X,t)\,dX_0 \\ &= (\|A\|^{1/2}/2\pi)\exp(-\tfrac{1}{2}\tilde{X}\cdot A\cdot X), \\ A &= \tilde{\beta}^{-1}E(I+DE)^{-1}\beta^{-1}. \end{aligned}$$

D is shown to be proportional to $e^{2\lambda t}$ (see Appendix C), and, hence, for large t, $A(t)$ approaches

$$A(t) \simeq \tilde{\beta}^{-1}(t)D^{-1}(t)\beta^{-1}(t) = D_\beta^{-1}(t).$$

Therefore, $P(X,t)$ for large t is the same as (6.1), (6.10), or (6.11), provided (ξ',η') is replaced by (ξ,η).

APPENDIX A: HIGHER ORDER RENORMALIZATION

For a periodic modulation, (2.5) is rewritten as

$$\dot{V}(t) = L_0 \cdot V(t) + \epsilon(e^{2i\gamma t}L_2 + e^{4i\gamma t}L_4 + \text{c.c.})\cdot V(t), \quad \text{(A1)}$$

where L_0 is the same as (3.1), and

$$L_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad L_4 = \frac{1}{2}\begin{pmatrix} -i & -1 \\ -1 & i \end{pmatrix}.$$

With the use of the Laplace transform

$$\widehat{V}(s) = \int_0^\infty e^{-st}V(t)\,dt,$$

(A1) can be written as

$$(s - L_0)\cdot\widehat{V}(s) - \epsilon\{L_2\cdot[\widehat{V}(s-2i\gamma) + \widehat{V}(s+2i\gamma)] + L_4\cdot\widehat{V}(s-4i\gamma) + L_4^*\cdot\widehat{V}(s+4i\gamma)\} = V(0).$$

Now rewrite this equation as

$$(s - L_0 - \epsilon^2\Delta(s))\cdot\widehat{V}(s) - \epsilon\{L_2\cdot[\widehat{V}(s-2i\gamma) + \widehat{V}(s+2i\gamma)] + L_4\cdot\widehat{V}(s-4i\gamma) + L_4^*\cdot\widehat{V}(s+4i\gamma)\} + \epsilon^2\Delta(s)\cdot\widehat{V}(s) = V(0), \quad \text{(A2)}$$

with $\epsilon^2\Delta(s)$ representing higher order secular effects, and define

$$K(s) = s - L_0 - \epsilon^2\Delta(s).$$

Then from (A2) follows

$$\widehat{V}(s-2i\gamma) = K^{-1}(s-2i\gamma)(V(0) + \epsilon\{L_2\cdot[\widehat{V}(s-4i\gamma) + \widehat{V}(s)] + L_4\cdot\widehat{V}(s-6i\gamma) + L_4^*\cdot\widehat{V}(s+2i\gamma)\} - \epsilon^2\Delta(s-2i\gamma)\cdot\widehat{V}(s-2i\gamma)),$$

and similar equations for $\widehat{V}(s+2i\gamma)$ and $\widehat{V}(s\pm 4i\gamma)$. Substitution of these into (A2) yields

$$\begin{aligned}(K(s) &+ \epsilon^2\{\Delta(s) - L_2[K^{-1}(s-2i\gamma) + K^{-1}(s+2i\gamma)]L_2 \\ &- L_4K^{-1}(s-4i\gamma)L_4^* - L_4^*K^{-1}(s+4i\gamma)L_4\})\cdot\widehat{V}(s) \\ &= (I + \epsilon\{L_2(K^{-1}(s-2i\gamma) + K^{-1}(s+2i\gamma)) \\ &\quad + L_4K^{-1}(s-4i\gamma) + L_4^*K^{-1}(s+4i\gamma)\})\cdot V(0) \\ &\quad + \epsilon^2L_2(K^{-1}(s-2i\gamma)\{L_2\cdot\widehat{V}(s-4i\gamma) \\ &\quad + L_4\cdot\widehat{V}(s-6i\gamma) + L_4^*\cdot\widehat{V}(s+2i\gamma) + \cdots\}) \\ &\quad + \epsilon^2[L_2\Delta(s-2i\gamma)\cdot\widehat{V}(s-2i\gamma) + \cdots]. \end{aligned} \quad \text{(A3)}$$

Further iterations give rise to secular effects of higher order. Evidently it is natural to put

$$\Delta(s) = L_2\{K^{-1}(s-2i\gamma) + K^{-1}(s+2i\gamma)\}L_2 + L_4K^{-1}(s-4i\gamma)L_4^* + L_4^*K^{-1}(s+4i\gamma)L_4, \quad \text{(A4)}$$

which is an implicit equation for $\Delta(s)$.

The properties of the fundamental oscillation are determined by the solution of

$$\det\|K(s)\| = 0 \quad \text{(A5)}$$

near the origin, where $K^{-1}(s\pm 2i\gamma)$ and $K^{-1}(s\pm 4i\gamma)$ can be approximated by $(\pm 2i\gamma)^{-1}$ and $(\pm 4i\gamma)^{-1}$, respectively. Therefore

$$\Delta(s) \simeq \frac{i}{4\gamma}(L_4L_4^* - L_4^*L_4) + \frac{1}{4\gamma}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then (A5) leads to

$$(s + \lambda)^2 = \epsilon^2 - (\gamma - \omega - \epsilon^2/4\gamma)^2.$$

In the case $\lambda = 0$, the condition for amplification is

$$|h\omega_0|/4 > |\gamma - \omega_0 - h^2\omega_0/64|, \quad \text{(A6)}$$

which is a known result (see Ref. 11).

APPENDIX B: NON-MARKOVIAN, GAUSSIAN PROCESSES

To start with, a brief review of non-Markovian, Gaussian processes will be given in a form convenient to later analyses. The simplest equation that includes a damping with memory and a stochastic force may be written as

$$\dot{x}(t) + \int_0^t \lambda(t - t')x(t')\,dt' = f(t), \quad t > 0, \quad \text{(B1)}$$

with $\lambda(t - t')$ representing a memory effect, and $f(t)$ a Gaussian variable, satisfying $\langle f(t)\rangle = 0$, $\langle f(t)f(t')\rangle = F(t - t')$. The solution is

$$x(t) = \chi(t)x(0) + \int_0^t \chi(t')f(t - t')\,dt', \quad \chi(0) = 1,$$

where $\chi(t)$ is the inverse Laplace transform of $[s + \hat{\lambda}(s)]^{-1}$ with

$$\hat{\lambda}(s) = \int_0^\infty e^{-st}\lambda(t)\,dt.$$

Let $P(x(0),0)$ describe the distribution of the initial value $x(0)$. Then the average value of $x(t)$ is given by

$$\begin{aligned}\langle\langle x(t)\rangle\rangle &= \int dx(0)\,\langle P(x(0),0)x(t)\rangle \\ &= [1/\chi(t)]\int dx\,\langle P((x - A(t))/\chi(t),0)x\rangle \\ &\equiv \int dx\,P(x,t)x,\end{aligned}$$

where

$$A(t) = \int_0^t \chi(t')f(t - t')\,dt'.$$

In terms of the Fourier transform

$$P(x,0) = (1/2\pi)\int e^{ikx}\psi(k)\,dk,$$

$P(x,t)$ takes the form

$$\begin{aligned}P(x,t) &= [1/2\pi\chi(t)]\int\langle\exp\{ik[x - A(t)]/\chi(t)\}\rangle\psi(k)\,dk \\ &= [1/2\pi\chi(t)]\int\exp[ikx/\chi(t) - \tfrac{1}{2}k^2D(t)/\chi^2(t)]\psi(k)\,dk\end{aligned} \quad \text{(B2)}$$

with

$$\begin{aligned}D(t) &= \langle A(t)A(t)\rangle \\ &= \int_0^t dt'\int_0^t dt''\,\chi(t')\chi(t'')F(t' - t'').\end{aligned} \quad \text{(B3)}$$

In the limit $t\to\infty$, $\chi(t)$ for a passive system is expected to approach zero, and $D(t)$ a finite value, denoted by D_∞ in the following. Application of the method of steepest descents to (B2) yields the stationary distribution

$$\begin{aligned}P(x,\infty) &= 1/(2\pi D_\infty)^{1/2}\psi(0)e^{-x^2/2D}, \\ &= 1/(2\pi D_\infty)^{1/2}e^{-x^2/2D},\end{aligned} \quad \text{(B4)}$$

whose Fourier transform is

$$\psi_\infty(k) = \int e^{-ikx} P(x,\infty)\,dx = e^{-D_\infty k^2/2}.$$

The stationary distribution should not depend on t. Thus, if $\psi_\infty(k)$ is substituted into (B2) and the integral is evaluated, then the result must be identical with $P(x,\infty)$. From this self-consistency requirement follows

$$D(t) + \chi^2(t) D_\infty = D_\infty. \tag{B5}$$

Equation (B5) holds obviously at $t = 0$. Differentiation gives

$$\int_0^t \chi(t')F(t-t')\,dt' + \dot\chi(t)D_\infty = 0.$$

The Laplace transform of this relation leads to

$$[s + \hat\lambda(s)]^{-1}\hat F(s) + \{s[s+\hat\lambda(s)]^{-1} - 1\}D_\infty = 0,$$

whence follows the fluctuation–dissipation relation

$$\hat F(s) = \hat\lambda(s)D_\infty,$$

or

$$\langle f(t)f(t')\rangle = \lambda(t-t')D_\infty. \tag{B6}$$

With the use of (B5), (B2) simplifies to

$$\begin{aligned}
P(X,t) &= [1/2\pi\chi(t)] \\
&\quad\times \int \exp\{ikx/\chi(t) - \tfrac{1}{2}k^2[1/\chi^2(t) - 1]D_\infty\} \\
&\quad\times \psi(k)\,dk \\
&= 1/\{2\pi[1-\chi^2(t)]D_\infty\}^{1/2} \int d\xi P(\xi,0) \\
&\quad\times \exp\{-\tfrac{1}{2}(x - \chi(t)\xi)^2/(1-\chi^2(t))D_\infty\}.
\end{aligned} \tag{B7}$$

It follows easily that[12]

$$\frac{\partial}{\partial t}P(X,t) = -\frac{\dot\chi}{\chi}\left\{\frac{\partial}{\partial x}[x\,P(x,t)] + D_\infty\frac{\partial^2}{\partial x^2}P(x,t)\right\}. \tag{B8}$$

If $\chi(t)$ is monotonic, one can define a generalized time T through

$$T = -\log\chi(t)\int_0^\infty \chi(t)\,dt$$

to reduce (B8) to the form

$$\frac{\partial}{\partial T}P(x,T) = \alpha\left\{\frac{\partial}{\partial x}[x\,P(x,T)] + D_\infty\frac{\partial^2}{\partial x^2}P(x,T)\right\}, \tag{B9}$$

$$\alpha = \left[\int_0^\infty \chi(t)\,dt\right]^{-1}.$$

Formally, (B9) is identical with the Fokker–Planck equation in a stationary Markovian, Gaussian process. Note, however, that t or T in (B8) or (B9) is a specialized time coordinate in that, as is obvious from (B1), incoherent effects are supposed to set in at $t = T = 0$.

All the results can be extended with a slight modification to the problem in hand; the equation to be studied is

$$\begin{aligned}
\dot X(t) &= \begin{pmatrix} 0 & \omega_0 \\ -\omega_0 & 0\end{pmatrix}\cdot X(t) - 2\int_0^t \lambda(t-t')\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\cdot X(t')\,dt' \\
&\quad + h\cos 2\gamma t X(t) + f(t)\begin{pmatrix}0\\1\end{pmatrix}.
\end{aligned}$$

It is assumed that $\lambda(t-t')$ is very small and γ is nearly equal to ω_0. The transformation $X(t) = e^{\Gamma t}\cdot V(t)$ yields

$$\begin{aligned}
\dot V(t) &= \begin{pmatrix} 0 & -\epsilon - (\gamma-\omega_0) \\ -\epsilon + (\gamma - \omega_0) & 0\end{pmatrix}\cdot V(t) \\
&\quad + L_1(t)\cdot V(t) - \int_0^t \lambda(t-t')e^{-\Gamma(t-t')} \\
&\quad \times\left[I + \begin{pmatrix} -\cos 2\gamma t' & \sin 2\gamma t' \\ \sin 2\gamma t' & \cos 2\gamma t'\end{pmatrix}\right]\cdot V(t') \\
&\quad + f(t)e^{-\Gamma t}\cdot\begin{pmatrix}0\\1\end{pmatrix}.
\end{aligned}$$

It was shown in Sec. III and in Appendix A that the term involving $L_1(t)$ gives rise to small corrections to the secular effects as well as small higher harmonics. The Laplace transform of the above equation with this term omitted yields

$$\begin{aligned}
s\hat V(s) - V(O) &= L_0(s)\cdot\hat V(s) + \hat\lambda_\Gamma(s)[L_3\cdot\hat V(s - 2i\gamma) + L_3^*\cdot\hat V(s+2i\gamma)] \\
&\quad + \hat f_\Gamma(s)\begin{pmatrix}0\\1\end{pmatrix}.
\end{aligned} \tag{B10}$$

with

$$L_0(s) = \begin{pmatrix} 0 & -\epsilon - (\gamma - \omega_0) \\ -\epsilon + (\gamma-\omega_0) & 0\end{pmatrix} - \hat\lambda_\Gamma(s),$$

$$\hat\lambda_\Gamma(s) = \int_0^\infty e^{-st}\lambda(t)e^{-\Gamma t}\,dt,$$

$$L_3 = \tfrac{1}{2}\begin{pmatrix}1 & i \\ i & -1\end{pmatrix},\quad \hat f_\Gamma(s) = \int_0^\infty e^{-st}f(t)e^{-\Gamma t}\,dt.$$

Equation (B10) is very similar to (A2), and analyses from now on go parallel with those in Appendix A. Taking account of the secular effects to order $\hat\lambda^2_\Gamma(s)$, we find

$$\begin{aligned}
J(s)\cdot\hat V(s) &= \{I + \hat\lambda_\Gamma(s)[L_3 J^{-1}(s - 2i\gamma) \\
&\quad + L_3^* J^{-1}(s + 2i\gamma)]\}\cdot V(0) \\
&\quad + [\hat f_\Gamma(s) + \hat\lambda_\Gamma(s)L_3 J^{-1}(s - 2i\gamma)\hat f_\Gamma(s - 2i\gamma)]\cdot\begin{pmatrix}0\\1\end{pmatrix}
\end{aligned} \tag{B11}$$

with

$$\begin{aligned}
J(s) &= s - L_0(s) + \hat\lambda_\Gamma(s)L_3 J^{-1}(s - 2i\gamma)\hat\lambda_\Gamma(s - 2i\gamma)L_3^* \\
&\quad + \hat\lambda_\Gamma(s)L_3^* J^{-1}(s + 2i\gamma)\hat\lambda_\Gamma(s+2i\gamma)L_3.
\end{aligned} \tag{B12}$$

The fundamental oscillation in the lowest order approximation is given by

$$V(t) = \nu(t)\cdot V(0) + \int_0^t dt' f(t-t')\nu(t')e^{-\Gamma(t-t')}\cdot\begin{pmatrix}0\\1\end{pmatrix},$$

where $\nu(t)$ is the inverse Laplace transform of $J^{-1}(s)$. In the original representation, we have

$$X(t) = e^{\Gamma t}\nu(t)\cdot X(0) + e^{\Gamma t}\int_0^t dt' f(t-t')\nu(t')e^{-\Gamma(t-t')}\cdot\begin{pmatrix}0\\1\end{pmatrix}.$$

The condition for amplification is that the real part of the solution of $J(s) = 0$ be positive.

Define $\chi(t) = e^{\Gamma t}\nu(t)$ and write

$$X(t) = \chi(t)\cdot X(0) + A(t),$$

$$A(t) = \int_0^t dt' f(t-t')e^{\Gamma(t-t')}\chi(t')e^{-\Gamma(t-t')}\cdot\begin{pmatrix}0\\1\end{pmatrix}.$$

The distribution function $P(X,t)$ is shown to take the form

$$P(X,t) = [2\pi\|\chi(t)\|]^{-1}\int \langle\exp\{ik\cdot\chi^{-1}(t)\cdot[X - A(t)]\}\rangle \times\psi(k)\,d^2k$$
$$= [2\pi\|\chi(t)\|]^{-1}\int \exp\{ik\cdot\chi^{-1}(t)\cdot X - \tfrac{1}{2}\tilde{k}\cdot\chi^{-1}(t)D_\Gamma(t)\tilde{\chi}^{-1}(t)\cdot k\}\psi(k)\,d^2k,$$

in which

$$D_\Gamma(t) = \int_0^t dt'\int_0^t dt''\,F(t'-t'')e^{\Gamma(t-t')}\chi(t') \times e^{-\Gamma(t-t')}\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}e^{\Gamma(t-t'')}\tilde{\chi}(t'')e^{-\Gamma(t-t'')}. \quad \text{(B13)}$$

It can be shown after somewhat tedious calculations that

$$\frac{\partial P(X,t)}{\partial t} = \left[\frac{1}{2}\tilde{\nabla}\cdot\chi\frac{d}{dt}(\chi^{-1}D_\Gamma\tilde{\chi}^{-1})\tilde{\chi}\cdot\nabla - \tilde{\nabla}\cdot\chi\dot{\chi}^{-1}\cdot X\right] \times P(X,t). \quad \text{(B14)}$$

The conditional probability can also be calculated easily to be

$$P(X_0,0;X,t) = \{2\pi[\|D_\Gamma(t)\|]^{1/2}\}^{-1} \times\exp\{-\tfrac{1}{2}[X-\chi(t)\cdot X_0]^\sim\cdot D_\Gamma^{-1}(t)\cdot[X-\chi(t)\cdot X_0]\}. \quad \text{(B15)}$$

The physical meaning of the expansion (B12) is of some interest. To see this, we put $\lambda(t) = \lambda\delta(t)$, which yields $\hat{\lambda}_\Gamma(s) = \lambda$. Then (B12) reduces to

$$J(s)\simeq\begin{pmatrix} s+\lambda & -\epsilon-(\gamma-\omega_0)\\ -\epsilon+(\gamma-\omega_0) & s+\lambda\end{pmatrix} + \frac{i\lambda^2}{2\gamma}(L_3L_3^* - L_3^*L_3)$$
$$\simeq\begin{pmatrix} s+\lambda & -\epsilon-(\gamma-\omega)\\ -\epsilon+(\gamma-\omega) & s+\lambda\end{pmatrix},$$

where we have put

$$\omega = \omega_0 - \lambda^2/2\gamma \simeq \omega_0 - \lambda^2/2\omega_0, \quad \text{(B16)}$$

and $J^{-1}(s\pm 2i\gamma)$ has been replaced by $(\pm 2i\gamma)^{-1}$. The last approximation is permissible so long as the fundamental oscillation is concerned. It should be mentioned that the right-hand side of (B16) is the first two terms in the expansion of $(\omega_0^2-\lambda^2)^{1/2}$ in terms of λ^2. Thus the expansion (B12) to the order considered represents the frequency renormalization. In the same approximation, $\chi(t)$ and $D_\Gamma(t)$ are shown to agree with $\beta(t)$ and $D_\beta(t)$ defined in Sec. VI, respectively.

APPENDIX C: EVALUATION OF $D_\beta(t)$

When $\beta(t)$ is substituted from (4.5) into (6.4), $D_\beta(t)$ takes the form

$$D_\beta(t) = 2F_0e^{-K}e^{\Gamma t}e^{L_0t}\int_0^t dt'\,e^{-L_0t'}e^{-\Gamma t'}e^K\begin{pmatrix}0&0\\0&1\end{pmatrix} \times e^Ke^{\Gamma t'}e^{-\tilde{L}_0t}e^{-\Gamma t}e^{-K}. \quad \text{(C1)}$$

For $\epsilon = 0$, Γ commutes with L_0, and we can make use of the properties of e^K, as exhibited in Sec. II, to transform back to

$$\beta(t) = e^{-K}e^{\Gamma t}e^{L_0t}e^K = \exp\left[\begin{pmatrix}0 & \omega_0\\ -\omega_0 & -2\lambda\end{pmatrix}t\right]. \quad \text{(C2)}$$

Then (C1) gives

$$D_\beta(t) = 2F_0\int_0^t dt'\exp\left[\begin{pmatrix}0 & \omega_0\\ -\omega_0 & -2\lambda\end{pmatrix}t'\right]\begin{pmatrix}0&0\\0&1\end{pmatrix} \times\exp\left[\begin{pmatrix}0 & -\omega_0\\ \omega_0 & -2\lambda\end{pmatrix}t'\right].$$

With the use of the identity

$$\begin{pmatrix}0 & \omega_0\\ -\omega_0 & -2\lambda\end{pmatrix} + \begin{pmatrix}0 & -\omega_0\\ \omega_0 & -2\lambda\end{pmatrix} = -4\lambda\begin{pmatrix}0&0\\0&1\end{pmatrix},$$

we can write $D_\beta(t)$ as

$$D_\beta(t) = \frac{-F_0}{2\lambda}\int_0^t dt'\frac{d}{dt'}\beta(t')\tilde{\beta}(t'),$$

which leads to (6.4) after integration.

Written explicitly, (C2) takes the form

$$\exp\left[\begin{pmatrix}0 & \omega_0\\ -\omega_0 & -2\lambda\end{pmatrix}t\right] = e^{-\lambda t}e^{-K}\exp\left[\begin{pmatrix}0 & \omega\\ -\omega & 0\end{pmatrix}t\right]e^K.$$

Substitution of this into (6.4) yields

$$D_\beta(t) = \frac{F_0}{2\lambda}\left\{\begin{pmatrix}1&0\\0&1\end{pmatrix} - e^{-2\lambda t}e^{-K}\exp\left[\begin{pmatrix}0&\omega\\-\omega&0\end{pmatrix}t\right]e^{2K}\times\exp\left[\begin{pmatrix}0&-\omega\\ \omega&0\end{pmatrix}t\right]e^{-K}\right\}$$
$$= \frac{F_0}{2\lambda}\left[\begin{pmatrix}1&0\\0&1\end{pmatrix} - e^{-2\lambda t}\begin{pmatrix}\cosh k & -\sinh k\\ -\sinh k & \cosh k\end{pmatrix}\begin{pmatrix}\cos\omega t & \sin\omega t\\ -\sin\omega t & \cos\omega t\end{pmatrix} \times\begin{pmatrix}\cosh 2k & \sinh 2k\\ \sinh 2k & \cosh 2k\end{pmatrix}\begin{pmatrix}\cos\omega t & -\sin\omega t\\ \sin\omega t & \cos\omega t\end{pmatrix}\begin{pmatrix}\cosh k & -\sinh k\\ -\sinh k & \cosh k\end{pmatrix}\right].$$

For $\lambda t \gg 1$, this reduces to (6.5).

Evaluation of $D_\beta(t)$ for $\epsilon\neq 0$ is extremely tedious; we shall do this only for $\omega_0\gg\lambda$ and $\gamma=\omega$, in which case e^K may be approximated by a unit matrix and L_0 becomes symmetric:

$$L_0 = -\begin{pmatrix}\lambda & \epsilon\\ \epsilon & \lambda\end{pmatrix} = \tilde{L}_0.$$

Therefore

$$D_\beta(t) = F_0e^{\Gamma t}e^{L_0t}\int_0^t dt'\,e^{-L_0t'}\left[\begin{pmatrix}1&0\\0&1\end{pmatrix} - \begin{pmatrix}\cos 2\omega t' & \sin 2\omega t'\\ \sin 2\omega t' & -\cos 2\omega t'\end{pmatrix}\right]e^{-L_0t'}e^{L_0t}e^{-\Gamma t}. \quad \text{(C3)}$$

It is easily seen that the first term in the curly bracket yields, after integration, quantities of order $|\epsilon \pm \lambda|^{-1}$, whereas contributions from the second term are of order ω^{-1}. Hence, for small enough ϵ and λ, the second term is negligible, and we obtain

$$D_\beta(t) \simeq F_0 e^{\Gamma t} \int_0^t dt'\, e^{2L_0 t'} e^{-\Gamma t} = F_0 e^{\Gamma t} (2L_0)^{-1} (e^{2L_0 t} - 1) e^{-\Gamma t},$$

where we have put $\lambda^2 - \epsilon^2 \neq 0$. From now on the calculation is straightforward; the result is

$$D_\beta(t) = \frac{F_0}{2(\lambda^2 - \epsilon^2)} \begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix} \begin{pmatrix} \lambda & -\epsilon \\ -\epsilon & \lambda \end{pmatrix} \times \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - e^{-2\lambda t} \begin{pmatrix} \cosh 2\epsilon t & -\sinh 2\epsilon t \\ -\sinh 2\epsilon t & \cosh 2\epsilon t \end{pmatrix} \right] \times \begin{pmatrix} \cos\omega t & -\sin\omega t \\ \sin\omega t & \cos\omega t \end{pmatrix}, \tag{C4}$$

and

$$\|D_\beta(t)\| = [F_0/4(\lambda^2 - \epsilon^2)](1 - e^{-4\lambda t} - 2e^{-2\lambda t}\cosh 2\epsilon t). \tag{C5}$$

The inverse matrix is

$$D_\beta^{-1}(t) = \frac{2}{F_0}(1 + e^{-4\lambda t} - 2e^{-2\lambda t}\cosh 2\epsilon t)^{-1} \begin{pmatrix} \cos\omega t & \sin\omega t \\ -\sin\omega t & \cos\omega t \end{pmatrix} \times \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - e^{-2\lambda t} \begin{pmatrix} \cosh 2\epsilon t & \sinh 2\epsilon t \\ \sinh 2\epsilon t & \cosh 2\epsilon t \end{pmatrix} \right] \times \begin{pmatrix} \lambda & \epsilon \\ \epsilon & \lambda \end{pmatrix} \begin{pmatrix} \cos\omega t & -\sin\omega t \\ \sin\omega t & \cos\omega t \end{pmatrix}. \tag{C6}$$

For $\lambda^2 > \epsilon^2$, the second term in the curly bracket vanishes in the limit $|\epsilon| t \gg 1$, and it follows that

$$D_\beta^{-1}(t) \simeq \frac{2}{F_0} \left[\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \epsilon \begin{pmatrix} \sin 2\omega t & \cos 2\omega t \\ \cos 2\omega t & -\sin 2\omega t \end{pmatrix} \right] \tag{C7}$$

and

$$\|D_\beta(t)\| \simeq F_0/4(\lambda^2 - \epsilon^2).$$

For $\lambda^2 < \epsilon^2$, we have in the limit $|\epsilon| t \gg 1$

$$D_\beta^{-1}(t) \simeq \frac{1}{F_0} \Big\{ [(|\epsilon| + \lambda) + (|\epsilon| - \lambda) e^{-2(|\epsilon| - \lambda)t}] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mathrm{sgn}(\epsilon)[(|\epsilon| + \lambda) - (|\epsilon| - \lambda) e^{-2(|\epsilon| - \lambda)t}] \times \begin{pmatrix} \sin 2\omega t & \cos 2\omega t \\ \cos 2\omega t & -\sin 2\omega t \end{pmatrix} \Big\} \tag{C8}$$

and

$$\|D_\beta(t)\| \simeq F_0 e^{2(|\epsilon| - \lambda)t}/4(\epsilon^2 - \lambda^2). \tag{C9}$$

Finally, when $\lambda^2 - \epsilon^2 = 0$, direct integration of (C3) leads to

$$D_\beta^{-1}(t) = \frac{\lambda}{2F_0} \left[\left(\frac{4}{1 - e^{-4\lambda t}} + \frac{1}{\lambda t} \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \mathrm{sgn}(\epsilon) \left(\frac{4}{1 - e^{-4\lambda t}} - \frac{1}{\lambda t} \right) \begin{pmatrix} \sin 2\omega t & \cos 2\omega t \\ \cos 2\omega t & -\sin 2\omega t \end{pmatrix} \right] \tag{C10}$$

and

$$\|D_\beta(t)\| = F_0^2 t (1 - e^{-4\lambda t})/4\lambda. \tag{C11}$$

[1] W. Anderson, J. Phys. Soc. Japan **9**, 316 (1954).
[2] R. Kubo, J. Phys. Soc. Japan **9**, 935 (1954).
[3] R. Kubo, J. Math. Phys. **4**, 174 (1963).
[4] R. L. Stratonovich, *Topics in the Theory of Random Noise II* (Gordon and Breach, New York, 1967), Chap. 10.
[5] R. Kubo, J. Phys. Soc. Japan **26** Suppl. **1** (1969).
[6] R. C. Bourret, Physica **54**, 623 (1971).
[7] R. C. Bourret, U. Frisch, and A. Pouquet, Physica **65**, 303 (1973).
[8] M. Tokuyama and H. Mori, Prog. Theor. Phys. **55**, 411 (1976).
[9] N. G. van Kampen, Physics Reports **24**, 171 (1976).
[10] B. J. West, K. Lindenberg, and V. Seshadri, Physica A **102**, 470 (1980).
[11] L. D. Landau and E. M. Lifshitz, *Mechanics* (Pergamon, Oxford, 1960), §27.
[12] R. F. Fox, Phys. Rep. **48**, 179 (1978).
[13] M. C. Wang and G. E. Uhlenbeck, Rev. Mod. Phys. **17**, 323 (1945).

Generalized decoupling theorem in quantum field theory

Edward B. Manoukian
Department of National Defence, Royal Military College of Canada, Kingston, Ontario K7L 2W3, Canada

(Received 2 December 1980; accepted for publication 24 April 1981)

Previous analyses of the decoupling theorem, in Euclidean space, considered only *one* mass scale in the theory becoming large and had the stringent constraint of allowing *no* zero mass particles in the theory. We generalize the decoupling theorem in both respects. We prove the vanishing property of renormalized Feynman amplitudes, with subtractions, when any subset of the masses in the theory become large and, in general, at *different* rates, thus providing different large mass scales in the theory. This theorem is then extended and we give sufficiency conditions for the validity of the decoupling theorem when any subset of the remaining nonasymptotic masses are scaled to *zero* and, in general, at different rates. All the subtractions of renormalization are carried out at the origin of momentum space. The proof applies for theories with derivative couplings and with higher spin fields as well.

PACS numbers: 11.10.Jj

I. INTRODUCTION

Previous derivations[1,2] of the decoupling theorem, in Euclidean space, considered that only one mass scale in the theory becomes large and had the constraint that no zero mass particles be allowed in the theory. In this paper we generalize the decoupling theorem in both respects. We prove the vanishing property of renormalized Feynman amplitudes $\mathscr{A}$, *with* subtractions, when any subset of the masses becomes large and, in general, at *different* rates. We then apply a previous analysis of the zero mass behavior of renormalized Feynman amplitudes[3] to give sufficiency conditions for the validity of the decoupling theorem, i.e., of the vanishing property of $\mathscr{A}$, when any subset of the remaining nonasymptotic masses are scaled to zero and, in general, at different rates. We then give some examples. For many interesting applications of the decoupling theorem see, for example, Refs. 4–8. An expected universality of the fundamental interactions at very high energies (cf. Refs. 4 and 9) makes it interesting to investigate the decoupling theorem when different "heavy" masses (providing an energy scale for the strengths of the interactions) are involved in a theory. All subtractions[10] of renormalization are carried out at the origin of momentum space with the degree of divergence of a subdiagram coinciding with its dimensionality. The proof applies for theories with derivative couplings and with higher spin fields as well.

II. BASIC THEOREMS

A renormalized Feynman amplitude associated with a proper and connected graph G is, in Euclidean space, of the form

$$\mathscr{A}(P,\mu) = \int_{\mathbb{R}^{4n}} dk\, R\,(P,k,\mu), \tag{1}$$

where

$$P = (p_1^0,\dots,p_m^3), \quad k = (k_1^0,\dots,k_n^3), \tag{2}$$

$$\mu = (\mu^1,\dots,\mu^\rho), \quad \mu^i > 0,\ i = 1,\dots,\rho,$$

$$R = \frac{\mathscr{P}(P,k,\mu)}{\prod_l [Q_l^2 + \mu_l^2]}. \tag{3}$$

P, k, and μ denote, respectively, the set of the components of the independent external momenta of G, the set of the integration variables, and the set of the masses associated with G. $\mathscr{P}$ is a polynomial in the elements in the sets in (2) and, in general, is also a polynomial in the $(\mu^i)^{-1}$ as well. The Q_l are of the form:

$$Q_l = \sum_{i=1}^{n} a_l^i k_i + \sum_{j=1}^{m} b_l^j p_j. \tag{4}$$

Consider the graph G. A line l joining a vertex v_i to a vertex v_j in G will be represented by a propagator (in Euclidean space) of the form

$$D_{ijl}^{+}(Q_{ijl},\mu_{ijl}) = \frac{P_{ijl}(Q_{ijl},\mu_{ijl})}{[Q_{ijl}^2 + \mu_{ijl}^2]}, \tag{5}$$

where Q_{ijl} is the momentum carried by the line l, μ_{ijl} is the mass associated with the line and coincides with one of the masses in (2). P_{ijl} is a polynomial in Q_{ijl},μ_{ijl} and, in general of μ_{ijl}^{-1} as well. In general, for a propagator D_{ijl}^{+}, we assume that

$$\deg_{\mu_{ijl}} D_{ijl}^{+} \leqslant -1, \tag{6}$$

[For example, for spins 0, 1, 2: $D_{ijl}^{+} = O(\mu_{ijl}^{-2})$ and for spins 1/2,3/2: $D_{ijl}^{+} = O(\mu_{ijl}^{-1})$, and (7) is also true], and that

$$\deg_{Q_{ijl},\mu_{ijl}} P_{ijl}(Q_{ijl},\mu_{ijl}) \leqslant \deg_{Q_{ijl}} P_{ijl}(Q_{ijl}\mu_{ijl}),$$

$$\deg_{Q_{ijl},\mu_{ijl}} D_{ijl}^{+} \leqslant \deg_{Q_{ijl}} D_{ijl}^{+}. \tag{7}$$

The momentum Q_{ijl} carried by the line l in G will be written[10,11] as

$$Q_{ijl} = k_{ijl} + q_{ijl}, \tag{8}$$

where k_{ijl} is a linear combination of the integration variables only, and q_{ijl} is a linear combination of the external momenta of G [see (4)]. By definition, a subdiagram is called proper if the subdiagram has no external lines and if any *one* of its lines is removed then the number of its connected parts does not increase. Accordingly, in particular, by a proper subdiagram we do not necessarily mean a connected subdiagram. For any proper subdiagram $g \subset G$, a line l joining a vertex v_i to a vertex v_j, all pertaining to g, will carry a momentum Q_{ijl} written in the form

$$Q_{ijl} = k^g_{ijl} + q^g_{ijl} = k_{ijl} + q_{ijl}, \tag{9}$$

where k^g_{ijl} is a linear combination of the integration variables, and q^g_{ijl} is a linear combination of the external momenta of g. We introduce the sets $k^g = \{k^g_{ijl}\}$, $q^g = \{q^g_{ijl}\}$ with i, j, and l pertaining to the subdiagram g. Similarly we write $k \equiv k^G = \{k_{ijl}\}, q \equiv q^G = \{q_{ijl}\}$ for the whole graph G. Let g' be any proper subdiagram containing the proper subdiagram g. Then with i, j, and l pertaining to the subdiagram g we have[11]

$$k^g_{ijl} = k^g_{ijl}(k^{g'}), \quad q^g_{ijl} = q^g_{ijl}(k^{g'}, q^{g'}), \tag{10}$$

where the dependence of the q^g_{ijl} on $k^{g'}$ is only through those $k^{g'}_{ijl}$ in g'/g. g'/g denotes the subdiagram g' with g in it shrunk to a point.

We consider a $(4n + 4m + \rho)$-dimensional Euclidean space $\mathbb{R}^{4n+4m+\rho}$. Let I be a $4n$-dimensional subspace of $\mathbb{R}^{4n+4m+\rho}$ and let E be a complement of I in $\mathbb{R}^{4n+4m+\rho}$, and we may then write the direct sum: $\mathbb{R}^{4n+4m+\rho} = I \oplus E$. I will be associated with the $4n$ integration variables. We further introduce a $4m$-dimensional subspace E_1 of E, and introduce the orthogonal complement E_2 of E_1 in E, and write $E = E_1 \oplus E_2$. The subspace E_2 is of ρ dimensions and will be associated with the masses in (2). We may write $\mathbb{R}^{4n+4m+\rho} = I \oplus E_1 \oplus E_2$. We also introduce the projection operations $\Lambda(E)$, $\Lambda(I \oplus E_1)$, $\Lambda(I \oplus E_2)$, $\Lambda(I)$ along the subspaces E, $I \oplus E_1$, $I \oplus E_2$, I, respectively.

Let $\mathbf{P}$ be a vector in $\mathbb{R}^{4n+4m+\rho}$ such that the elements in the sets P, k, μ in (2) may be written as some linear combinations of the (standard) components of P. Suppose $\mathbf{P}$ is of the form[12]

$$\mathbf{P} = \mathbf{L}_1\eta_1\eta_2 \cdots \eta_k + \dots + \mathbf{L}_r\eta_r \cdots \eta_k + \dots + \mathbf{L}_k\eta_k + \mathbf{C}, \tag{11}$$

where $1 \leqslant k \leqslant 4n + \rho$, and $\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_k$ are k independent vectors in $\mathbb{R}^{4n+4m+\rho}$, and $\eta_1, \eta_2, \dots, \eta_k$ are strictly positive parameters. $\mathbf{C}$ is a vector confined to a finite region in $\mathbb{R}^{4n+4m+\rho}$ *such that* $\mu^i \neq 0$ for all $i = 1, \dots, \rho$. For any r in $1 \leqslant r \leqslant k$, the vectors $\mathbf{L}_1, \mathbf{L}_2, \dots, \mathbf{L}_r$ span a subspace S_r. Throughout we suppose that $\Lambda(I \oplus E_2)S_r$ is the zero subspace: $\Lambda(I \oplus E_2)S_r = \{0\}$, for all $1 \leqslant r \leqslant k \leqslant 4n + \rho$. Since the integrations variables, in particular, may be written as some linear combinations of the components of $\mathbf{P}$ in (11) it follows that for any proper subdiagram g the k^g_{ijl} may be written as some linear combinations of the components of the vector $\mathbf{P}$ as well. The latter, in particular, means that for r fixed in $1 \leqslant r \leqslant k \leqslant 4n + \rho$, then a k^g_{ijl} of g may (or may not) depend on the parameter η_r in (11).

Let r be *fixed* in $1 \leqslant r \leqslant k$. Suppose that $\Lambda(E)S_r \neq \{0\}$. Then the renormalized Feynman integr*and*[10] may be written as[13]

$$R = \sum_N \tau^{(N)}, \tag{12}$$

where

$$\tau^{(N)} = \prod_{g \in N} (\delta^N_g - T_g) I_G \tag{13}$$

and I_G is the *un*renormalized integrand associated with G. N is a set of proper subdiagrams such that (i) $G \in N$, (ii) if g, g' are in N then $g \not\subseteq g'$ or $g' \not\subseteq g$, (iii) and finally let N be arranged in an increasing order: $\{\dots, g, g', \dots\}$ such that $g \not\subseteq g'$; then with g and g' as consecutive subdiagrams in N, all the $k^{g'}_{ijl}$ in g'/g are either all dependent on the parameter η_r or are all independent of the parameter η_r. T_g denotes the Taylor operation, with respect to the external momenta of g, about the origin up to the order $d(g)$—the dimensionality of the subdiagram g. If $d(g) < 0$, then we set $T_g \equiv 0$. The δ^N_g are defined as follows. Let $\underline{g}, g, \bar{g}$ can be any consecutive elements in N such that $\underline{g} \not\subseteq g \not\subseteq \bar{g}$. If all the k^g_{ijl} in $g/\underline{g}$ are dependent on η_r and all the $k^{\bar{g}}_{ijl}$ in $\bar{g}/g$ are independent of η_r, then $\delta^N_g = 1$, otherwise $\delta^N_g = 0$. On the other hand, $\delta^N_G = 1$. The sum in (12) is over *all* distinct N sets.

Let N be a set in (12). We write

$$N = H_1(N) \cup H_2(N), \tag{14}$$

such that a proper subdiagram g is in $H_1(N)$ if all the k^g_{ijl} in $g/\underline{g}$ are dependent on η_r, and g is in $H_2(N)$ if all the k^g_{ijl} in $g/\underline{g}$ are independent of η_r. We also write

$$H_1(N) = F_1(N) \cup F_2(N), \tag{15}$$

where $g \in F_1(N)$ if all the $k^{\bar{g}}_{ijl}$ in $\bar{g}/g$ are dependent on η_r, and $g \in F_2(N)$ if all the $k^{\bar{g}}_{ijl}$ in $\bar{g}/g$ are independent of η_r. In these notations, for $g \not\subseteq G, \delta^N_g = 0$ if $g \in H_1(N) \cup F_1(N)$, and $\delta^N_g = 1$ if $g \in F_2(N)$. Finally we set

$$\rho(g) = 4 \sum_{\substack{g' \in H_1(N) \\ g' \subset g}} L(g'/\underline{g}'), \tag{16}$$

where $L(g)$ denotes the number of independent loops in g. We note, in particular, that if $g \in H_1(N)$, then the sum in (16) obviously goes over subdiagrams $g' \in H_1(N)$ with $g' \not\subseteq g$. We also note from (13) that we may introduce the recursion relation

$$\tau_g(N) = (\delta^N_g - T_g)\tau_{\underline{g}}(N), \tag{17}$$

with

$$\tau^{(N)} \equiv \tau_G(N). \tag{18}$$

The subtraction scheme in (12) is not to be confused with the one in Ref. 11 (c.f. Ref. 10).

As before, suppose that $\Lambda(E)S_r \neq \{0\}$ with r *fixed* in $1 \leqslant r \leqslant k$. We then prove the following lemma for any subdiagram $g \not\subseteq G$ in N. The corresponding situation for the graph G will be treated separately.

Lemma: Let $\tau_g(N)$ be as defined in (17). For all $g' \subset g$, with $g' \in H_1(N)$, we scale all the $k^{g'}_{ijl}$ in $g'/\underline{g}'$, which by definition depend on η_r, by a parameter λ, and we scale all those masses μ_{ijl} in g, depending on η_r, by the parameter λ as well. If $g \in F_2(N)$ then

$$\deg_\lambda \tau_g(N) \leqslant \min[d(g), -1] - \rho(g), \tag{19}$$

and if $g \in H_2(N)$, then

$$\deg_\lambda \tau_g(N) \begin{cases} \leqslant -1 - \rho(g), \\ \quad \text{or} \\ = 0, \end{cases} \tag{20}$$

where the $= 0$ condition in (20) holds if there is no subdiagram $g' \subset g$ in N with $g' \in F_2(N)$, and there are no masses μ_{ijl} in the lines in g which depend on the parameter η_r.

The proof is by induction. Suppose that the lemma is true for $\underline{g} \not\subset g$ in N, we then prove the lemma for g as well. In addition to these hypotheses, suppose, as part of the induction hypotheses, that if $\underline{g} \in F_1(N) \cup H_2(N)$; then

$$\deg_{\eta_r} \tau_{\underline{g}}(N) \leqslant d(\underline{g}) - \rho(\underline{g}) \tag{21}$$

and that $\tau_{\underline{g}}(N)$ has a structure such as

$$\tau_{\underline{g}}(N) = P_{\underline{g}}(k^{\underline{g}}, q^{\underline{g}}, \mu) f^1_{\underline{g}}(k^{\underline{g}}, \mu) f^2_{\underline{g}}(k^{\underline{g}}, \mu), \tag{22}$$

where $P_{\underline{g}}$ is a polynomial in the elements in $k^{\underline{g}}$, $q^{\underline{g}}, \mu$ and, in general, in the $(\mu^i)^{-1}$ as well. The $f^n_{\underline{g}}$ are any functions of the form

$$f^n_g(k^g, q^g, \mu) = \prod_{\substack{g' \in H_n(N) \\ g' \subset g}} \prod_{ijl}^{g'/g'} [(\Theta^{g'}_{ijl})^2 + \mu^2_{ijl}]^{-\sigma^n_{ijl}}, \tag{23}$$

for $n = 1, 2$, $\Theta^{g}_{ijl} \equiv Q_{ijl}$, and for $g' \not\subset g, g' \in H_n(N)$, $\Theta^{g'}_{ijl} \equiv k^{g'}_{ijl}$. The σ^n_{ijl} are strictly positive integers. We also denote by $f^n(k^g, q^g, \mu)$ any function of the form

$$f^n(k^g, q^g, \mu) = \prod_{ijl}^{g/g} [Q^2_{ijl} + \mu^2_{ijl}]^{-\sigma^n_{ijl}}. \tag{24}$$

If the q^g_{ijl} in the $g/\underline{g}$, in (23) and (24), are set equal to zero, then the corresponding functions are denoted by $f^n_g(k^g, \mu)$ and $f^n(k^g, \mu)$, respectively. Finally, if $\underline{g} \in F_2(N)$, then we suppose that the $\tau_{\underline{g}}(N)$ has a structure such as

$$\tau_{\underline{g}}(N) = P_{\underline{g}}(k^{\underline{g}}, q^{\underline{g}}, \mu) f^1_{\underline{g}}(k^{\underline{g}}, q^{\underline{g}}, \mu) f^2_{\underline{g}}(k^{\underline{g}}, \mu). \tag{25}$$

We prove the above lemma together with the results through (21)–(25) for the subdiagram g itself.

Proof: (i) Suppose that $g \in H_2(N)$. Then $\underline{g} \in H_2(N) \cup F_2(N)$. We write $k^{\underline{g}} = k^{\underline{g}}(k^g)$ and $q^{\underline{g}} = q^{\underline{g}}(k^g, q^g)$. Suppose that $\underline{g} \in H_2(N)$. From the induction hypotheses we may then write

$$\begin{aligned}\tau_g(N) = -\sum_{\substack{A,B \\ a}} (q^g)^{A+B+a} P^A_g(k^g, 0, \mu) &\times f^2_B(k^g, \mu) f^1_{\underline{g}}(k^{\underline{g}}(k^g), \mu) f^2_{\underline{g}}(k^{\underline{g}}(k^g), \mu) \\ &\times P^a_{\underline{g}}(k^{\underline{g}}(k^g), q^{\underline{g}}(k^g, 0), \mu),\end{aligned} \tag{26}$$

where

$$|A| + |B| \leqslant d(g), \tag{27}$$

$$\deg_\lambda P^a_{\underline{g}} + \deg_\lambda f^1_{\underline{g}} + \deg_\lambda f^2_{\underline{g}} \begin{cases} \leqslant -1 - \rho(\underline{g}), \\ \quad \text{or} \\ \quad = 0. \end{cases} \tag{28}$$

The $= 0$ condition in (28) holds if $\underline{g}$ contains no masses depending on η_r, and if there is no subdiagram $g' \subset \underline{g}$, with $g' \in F_2(N)$ [i.e., in particular, if $\rho(\underline{g}) = 0$]. Since the k^g_{ijl} in $g/\underline{g}$ are independent of η_r, it follows from (26) and (28) that

$$\deg_\lambda \tau_g(N) \begin{cases} \leqslant -1 - \rho(g), \\ \quad \text{or} \\ \quad = 0, \end{cases} \tag{29}$$

with $\rho(g) = \rho(\underline{g})$, and the $= 0$ condition in (29) holds if no masses in g depend on η_r, and if there is no subdiagram $g' \subset g$, with $g' \in F_2(N)$. Suppose that $\underline{g} \in F_2(N)$, then $\tau_g(N)$ has the form

$$\begin{aligned}\tau_g(N) = -\sum_{\substack{A,B \\ a,b}} (q^g)^{A+B+a+b} P^A_g(k^g, 0, \mu) f^2_B(k^g, \mu) &\times f^2_{\underline{g}}(k^{\underline{g}}(k^g), \mu) f^1_{\underline{g}b}(k^{\underline{g}}(k^g), q^{\underline{g}}(k^g, 0), \mu) \\ &\times P^a_{\underline{g}}(k^{\underline{g}}(k^g), q^{\underline{g}}(k^g, 0), \mu).\end{aligned} \tag{30}$$

We note from (10) that all the denominators in $f^1_{\underline{g}b}(k^{\underline{g}}(k^g), q^{\underline{g}}(k^g, 0), \mu)$ are dependent on η_r since the $q^{\underline{g}}(k^g, 0)$ are independent of η_r. According to the induction hypotheses we then have

$$\deg_\lambda P^a_{\underline{g}} + \deg_\lambda f^1_{\underline{g}b} + \deg_\lambda f^2_{\underline{g}} \leqslant \min[d(\underline{g}), -1] - \rho(\underline{g}). \tag{31}$$

From (30) and (31) we the see that (29) is again true. Finally from (26) and (30) we also note that $\tau_g(N)$ has a structure as in (22).

(ii) Suppose that $g \in F_1(N)$. Then $\underline{g} \in H_2(N) \cup F_1(N)$. According to the induction hypotheses we then have

$$\begin{aligned}\tau_g(N) = -\sum_{\substack{A,B \\ a}} (q^g)^{A+B+a} P^A_g(k^g, 0, \mu) &\times f^1_B(k^g, \mu) f^1_{\underline{g}}(k^{\underline{g}}(k^g), \mu) \\ &\times f^2_{\underline{g}}(k^{\underline{g}}(k^g), \mu) P^a_{\underline{g}}(k^{\underline{g}}(k^g), q^{\underline{g}}(k^g, 0), \mu),\end{aligned} \tag{32}$$

where

$$|a| + \deg_\lambda P^a_{\underline{g}} + \deg_\lambda f^1_{\underline{g}} + \deg_\lambda f^2_{\underline{g}} \leqslant d(\underline{g}) - \rho(\underline{g}). \tag{33}$$

On the other hand, we have

$$|A| + |B| + \deg_\lambda P^A_g + \deg_\lambda f^1_B \leqslant d(g/\underline{g}) - 4L(g/\underline{g}), \tag{34}$$

which together with (32) and (33) imply that (21) is true. From (32) we also note that $\tau_g(N)$ has a structure as in (22).

(iii) Finally, suppose that $g \in F_2(N)$. Then $\underline{g} \in F_1(N) \cup H_2(N)$. Then we may write

$$\begin{aligned}\tau_g(N) = \sum_a (q^g)^a P^a_{\underline{g}}(k^{\underline{g}}(k^g), q^{\underline{g}}(k^g, 0), \mu) &\times f^1_{\underline{g}}(k^{\underline{g}}(k^g), \mu,) \\ &\times f^2_{\underline{g}}(k^{\underline{g}}(k^g), \mu)(1 - T^{d(g) - |a|}) I_{g/\underline{g}},\end{aligned} \tag{35}$$

where $T^{d(g) - |a|}$ denotes the Taylor operation with respect to the external momenta of g up to the order $d(g) - |a|$, for $d(g) \geqslant a$. For $d(g) \geqslant |a|$, the basic property of the remainder of a Taylor operation implies that

$$\deg_\lambda (1 - T^{d(g) - |a|}) I_{g/\underline{g}} \leqslant \deg_\lambda I_{g/\underline{g}} - d(g) + |a| - 1. \tag{36}$$

We also have from (7) that

$$\deg_\lambda I_{g/\underline{g}} \leqslant d(g/\underline{g}) - 4L(g/\underline{g}). \tag{37}$$

From (33), (35)–(37) we then obtain for $d(g) \geqslant 0$ that

$$\deg_\lambda \tau_g(N) \leqslant -1 - \rho(g), \tag{38}$$

where $\rho(g) = 4L(g/\underline{g}) + \rho(\underline{g})$. On the other hand, if $d(g) < 0$, then (33), (35), and (37), with $T_g \equiv 0$, imply that

$$\deg_\lambda \tau_g(N) \leqslant d(g) - \rho(g). \tag{39}$$

From (38) and (39) we may then write

$$\deg_{\lambda} \tau_g(N) \leqslant \min[d(g), -1] - \rho(g). \tag{40}$$

Finally, from (35) and the definition of the operation $(1 - T^{d(g) - |a|})$, we conclude that $\tau_g(N)$ has a structure as in (25). This completes the proof of the lemma together with the results through (21)–(25) for the subdiagram g itself.

We now apply the above lemma to the graph G itself. If $G \in F_2(N)$, then we may immediately conclude from (19) that

$$\deg_{\eta_r} \tau_G(N) \leqslant \min[d(G), -1] - \rho(G), \tag{41}$$

where we have used the fact that since $\Lambda\,(I \oplus E_2)S_r = \{0\}$, the external momenta of G are independent of η_r, and we may freely replace $\deg_{\lambda}$ by $\deg_{\eta_r}$ in (41). If $G \in H_2(N)$ then for $d(G) < 0$ we have from (20) and (6) that

$$\deg_{\eta_r} \tau_G(N) \leqslant -1 - \rho(G), \tag{42}$$

where we have used the fact that the k^{G}_{ijl} in $G/\underline{G}$ are independent of η_r in the case $G \in H_2(N)$. If $d(G) \geqslant 0$, then from (20) we note that

$$\deg_{\eta_r}(-T_G) I_{G/\underline{G}} \tau_{\underline{G}}(N) \leqslant -1 - \rho(G). \tag{43}$$

Accordingly from (42) and (43) we have for $g \in H_2(N)$ that

$$\deg_{\eta_r} \tau_G(N) \leqslant -1 - \rho(G). \tag{44}$$

Since not necessarily all of the four components of the k^{g}_{ijl} in $g/\underline{g}$ with $g \in H_1(N)$ are dependent on η_r, we have quite generally

$$r = \dim \Lambda\,(E)S_r \leqslant \rho(G). \tag{45}$$

By summing over all the N sets in (12) we then obtain from (41), (44), and (45) that for $\Lambda\,(E)S_r \neq \{0\}$,

$$\deg_{\eta_r} R \leqslant -1 - \dim \Lambda\,(E)S_r. \tag{46}$$

On the other hand, if $\Lambda\,(E)S_r = \{0\}$, we have directly from (6) that

$$\deg_{\eta_r} R \leqslant -1. \tag{47}$$

We are almost ready to state the generalized decoupling theorem. To this end we decompose E_2 into ρ one-dimensional orthogonal subspaces generated by orthogonal vectors $\mathbf{L}'_1,\ldots,\mathbf{L}'_\rho$. In particular, we choose the $\mathbf{L}'_i$ to have only one nonvanishing component μ^i, respectively, for $i = 1,\ldots,\rho$. We then write for the renormalized Feynman amplitude

$$\begin{aligned}&\mathscr{A}(P,\eta_1\ldots\eta_k\mu^1,\ldots,\eta_k\mu^k,\mu^{k+1},\ldots,\mu^\rho)\\ &\quad\equiv \mathscr{A}(\mathbf{L}'_1\eta_1\ldots\eta_k + \ldots + \mathbf{L}'_k\eta_k + \mathbf{C}),\end{aligned} \tag{48}$$

where $1 \leqslant k \leqslant \rho$, and $\mathbf{C}$ is a vector confined to a finite region in E *with* $\mu^{k+1},\ldots,\mu^\rho \neq 0$.

Since the renormalized Feynman integrand R satisfies[13] the power counting theorem criterion[12] we then have the following theorem.

Theorem 1: The power asymptotic coefficients $\alpha_I(S_i), i = 1,\ldots,k$, for $\mathscr{A}$ in (48) are all bounded above by -1 thus establishing the *vanishing property of $\mathscr{A}$ when the parameters $\eta_1,\ldots,\eta_k \to \infty$ independently.*

Since the logarithmic asymptotic coefficients[14,15] cannot change the vanishing property of $\mathscr{A}$ we will not carry out the explicit construction of these coefficients as the underlying analysis becomes quite cumbersome.

We now generalize Theorem 1 when any subset of the remaining nonasymptotic masses are led to go to zero. To this end, for any line carrying a mass μ_{ijl} that we wish to scale to zero, we write for the corresponding propagator

$$D^{+}_{ijl}(Q_{ijl},\mu_{ijl}) + \frac{\widehat{P}_{ijl}(Q_{ijl},\mu_{ijl})}{[Q^2_{ijl} + \mu^2_{ijl}]}, \tag{49}$$

in the expression for I_G, where $\widehat{P}(Q_{ijl},\mu_{ijl})$ is a polynomial in Q_{ijl} and μ_{ijl} but not of $(\mu_{ijl})^{-1}$, such that

$$D^{+}_{ijl}(Q_{ijl},0) = \frac{\widehat{P}_{ijl}(q_{ijl},0)}{Q^2_{ijl}}, \tag{50}$$

denotes the zero mass propagator. {In general we may also allow positive powers of $[Q^2_{ijl} + \mu^2_{ijl}]^{-1}$ in (49) as long as the correct dimensionality of D^{+}_{ijl} is taken when carrying out the subtractions of renormalization.}

We scale any subset of the masses in $\{\mu^{k+1},\ldots,\mu^\rho\}$, say the masses in the set $\{\mu^{k+s+1},\ldots,\mu^\rho\}$, $s \leqslant \rho - k - 1$, as follows:

$$\left.\begin{array}{l}\mu^{k+s+1} \to \lambda_1\mu^{k+s+1}\\ \mu^{k+s+2}\lambda_1\lambda_2\mu^{k+s+2}\\ \vdots\\ \mu^\rho \to \lambda_1\lambda_2\ldots\lambda_{\rho-s-k}\mu^\rho\end{array}\right\}. \tag{51}$$

We choose the external momenta of the graph G in question to be nonexceptional. This, in particular, means that *all* the external *momenta* carried by the external lines *to* the graph G impinging on its external vertices are *nonzero*. Then according to Ref. 3 the limits $\lambda_1,\lambda_2,\ldots,\lambda_{\rho-s-k} \to 0$ of $\mathscr{A}$ exist if the following is true.

Sufficiency conditions: Let i be any integer in $1 \leqslant i \leqslant \rho - s - k$. Let T_i be the set of *all* subdiagrams $G_i \subset G$ such that the G_i contain all the external vertices of G but not necessarily all of its lines. Also, if $G_i \in T_i$ then all the lines in G/G_i (if not empty) do not carry any external momenta and contain only masses from the set $\{\mu^{k+s+i},\ldots,\mu^\rho\}$; and any external line of G_i depend on elements from the set P and/or the set $\{\mu^1,\ldots,\mu^{k+s+i-1}\}$. Then if the following two conditions are true for every i in $1 \leqslant i \leqslant \rho - s - k$, the limit $\lambda_1,\lambda_2,\ldots,\lambda_{\rho-s-k} \to 0$ of $\mathscr{A}$ exists: (1) For any $G_i \in T_i, d(G_i) \leqslant d(G)$. (2) If $d(G_i) = d(G)$ then G_i has no proper subdiagram $g_i \subset G_i$ such that all of its masses are from the set $\{\mu^{k+s+i},\ldots,\mu^\rho\}$, and the dimensionality of each of the connected components of g_i are non-negative. In particular we note that these conditions imply that the graph G itself is not to contain a proper subdiagram g such that all of its masses are from the set $\{\mu^{k+s+1},\ldots,\mu^\rho\}$ and the dimensionality of each of the components of g are non-negative. The first condition (1) implies that the powers of the parameters $(1/\lambda_i)$ for $\lambda_i \to 0$ are non-positive, and condition (2) implies that the *logarithmic* coefficients associated with those parameters $(1/\lambda_i)$, having zero power asymptotic coefficients, are identically equal to *zero*. These two properties then imply the existence of $\lim \lambda_1,\lambda_2,\ldots,\lambda_{\rho-s-k} \to 0$ of $\mathscr{A}$.

We consider the renormalized Feynman amplitude:

$$\mathscr{A}(\mathbf{L}_1'\eta_1\eta_2\cdots\eta_k + \cdots + \mathbf{L}_k'\eta_k + \mathbf{L}_{k+s+1}'\lambda_1 + \cdots + \mathbf{L}_\rho'\lambda_1\cdots\lambda_{\rho-s-k} + \mathbf{C}), \tag{52}$$

where $\mathbf{C}$ is now a vector confined to a finite region in E with $\mu^{k+1},\ldots,\mu^{k+s} \neq 0$, and we may then state the following theorem.

Theorem 2: If the sufficiency conditions for all $i = 1,\ldots,\rho - s - k$ above are satisfied, then the power asymptotic coefficient $\alpha_r(S_j)$ with $j = 1,\ldots,k$, of $\mathscr{A}$ associated with the parameters $\eta_1,\ldots,\eta_j,\ldots,\eta_k$ are all bounded above by -1 thus establishing the *vanishing property of* $\mathscr{A}$ *when* $\eta_1,\eta_2,\ldots,\eta_k \to \infty;\lambda_1,\lambda_2,\ldots,\lambda_{\rho-s-k} \to 0$, *independently*.

III. EXAMPLES

Consider the renormalized Feynman amplitudes associated with the graphs in Figs. 1 and 2 in quantum electrodynamics. We write the photon propagator (in the Feynman gauge) as $D_{\mu\nu}(Q) = g_{\mu\nu}/(Q^2 + \mu^2)$. All the renormalized amplitudes associated with these latter graphs vanish when the mass m of the fermion is led to go to infinity. Also any subdiagram g of any of the graphs in Figs. 1 and 2, say a graph G, is such that $d(g) \leqslant d(G)$. We also note that the subdiagram G does not contain any proper subdiagram which has all its masses from the set $\{\mu\}$. Accordingly, the limit $\mu \to 0$ of all the amplitudes $\mathscr{A}$ of the above mentioned graphs exist, and in the limit $m \to \infty, \mu \to 0$ they vanish.

It is quite instructive to explicitly derive the vanishing property of the renormalized amplitude associated with the graph in Fig. 1 with $\mu = 0$. The explicit expression for it is given by (for $q^2 > 0$)

$$\mathscr{A}(q,m,0) = \gamma q a\left(\frac{q^2}{m^2}\right) + m b\left(\frac{q^2}{m^2}\right), \tag{53}$$

$$a\left(\frac{q^2}{m^2}\right) = -\frac{\alpha}{2\pi}\int_0^1 x\,dx\,\ln\left(1 + \frac{q^2}{m^2}x\right), \tag{54}$$

$$b\left(\frac{q^2}{m^2}\right) = -\frac{\alpha}{\pi}\int_0^1 dx\,\ln\left(1 + \frac{q^2}{m^2}x\right). \tag{55}$$

We use the identity $(x \geqslant 0)$,

$$\ln\left(1 + \frac{q^2}{m^2}x\right) = \frac{q^2}{m^2}x - \frac{q^4}{m^4}x^2\int_0^1 \frac{y\,dy}{[1 + (q^2/m^2)xy]}, \tag{56}$$

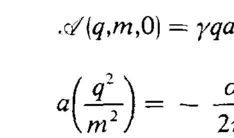


FIG. 1. Lowest order fermion self-energy graph in quantum electrodynamics.

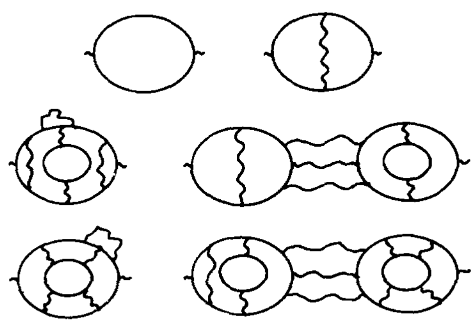


FIG. 2. Some low and high order photon self-energy graphs.

and hence the bound

$$\left|\ln\left(1 + \frac{q^2}{m^2}x\right)\right| \leqslant \frac{q^2}{m^2}x + \frac{q^4x^2}{2m^4}, \tag{57}$$

to bound the expressions in (54) and (55) as

$$\left|a\left(\frac{q^2}{m^2}\right)\right| \leqslant \frac{\alpha}{6\pi}\frac{q^2}{m^2} + \frac{\alpha}{16\pi}\frac{q^4}{m^4}, \tag{58}$$

$$\left|b\left(\frac{q^2}{m^2}\right)\right| \leqslant \frac{\alpha}{2\pi}\frac{q^2}{m^2} + \frac{\alpha}{6\pi}\frac{q^4}{m^4}. \tag{59}$$

Accordingly upon scaling m by η, with $\eta > 1$, we obtain

$$\left|a\left(\frac{q^2}{\eta^2 m^2}\right)\right| \leqslant \frac{1}{\eta}C^a\left(\frac{q^2}{m^2}\right), \tag{60}$$

$$\left|\eta m b\left(\frac{q^2}{\eta^2 m^2}\right)\right| \leqslant \frac{1}{\eta}mC^b\left(\frac{q^2}{m^2}\right), \tag{61}$$

where

$$C^a\left(\frac{q^2}{m^2}\right) = \frac{\alpha}{6\pi}\frac{q^2}{m^2} + \frac{\alpha}{16\pi}\frac{q^4}{m^4}, \tag{62}$$

$$C^b\left(\frac{q^2}{m^2}\right) = \frac{\alpha}{2\pi}\frac{q^2}{m^2} + \frac{\alpha}{6\pi}\frac{q^4}{m^4}, \tag{63}$$

in conformity with the above theorems. Other examples may be similarly treated. The analysis in this paper is rigorously carried out in Euclidean space. In a forthcoming paper some of the results obtained here will be generalized to Minkowski space as well, where many interesting applications have been, and are being, worked out.

ACKNOWLEDGMENT

This work was supported by the Chief Research and Development Award under CRAD No. 3610–637.

[1] E. B. Manoukian, "*On the Decoupling Theorem in Field Theory*" J. Math. Phys. **22**, 572 (1981).
[2] J. Ambjørn, Commun. Math. Phys. **67**, 109 (1979).
[3] E. B. Manoukian, J. Math. Phys. **21**, 1218 (1980).
[4] Y. Kazama and Y. – P. Yao, Phys. Rev. Lett. **43**, 1562 (1979).
[5] D. Toussaint, Phys. Rev. D **18**, 1626 (1978).
[6] E. G. J. Collins, F. Wilczek, and A. Zee, Phys. Rev. D **18**, 242 (1978).
[7] E. C. Poggio, H. R. Quinn, and J. B. Zuber, Phys. Rev. D **15**, 1630 (1977).
[8] T. Appelquist and J. Carazzonne, Phys. Rev. D **11**, 2856 (1975).
[9] D. Huff and O. Prewett, Eds. *The Nature of the Physical Universe* (Wiley, New York, 1979).
[10] E. B. Manoukian, Phys. Rev. D **14**, 966, 2202 (E) (1976).
[11] W. Zimmermann, Commun. Math. Phys. **15**, 208 (1969).
[12] S. Weinberg, Phys. Rev. **118**, 838 (1960).
[13] E. B. Manoukian, Phys. Rev. D **15**, 535 (1977) and preprint.
[14] J. P. Fink, J. Math. Phys. **9**, 1389 (1968).
[15] E. B. Manoukian, J. Math. Phys. **19**, 917 (1978).

Electric charge is vectorlike

S. Barr and A. Zee
Department of Physics, University of Washington, Seattle, Washington 98195

(Received 27 January 1981; accepted for publication 18 May 1981)

Electric charge is vectorlike in the real world. We formulate this fact in terms of an infinite hierarchy of trace conditions. Contemporary physics offers an understanding of only the first two conditions. We discuss the requirement of a vectorlike charge in the context of grand unification. We give an argument against elementary Higgs fields. A number of group theoretic assertions bearing on the issue of vectorlike electric charge are proven. We argue that the observation of color nontriplet quarks would signal a rich fermionic spectrum. Several unsolved mathematical problems are stated.

PACS numbers: 11.10.Lm, 11.30.Na

A fundamental, and striking, feature of our world is the vectorial character of the electric charge. Let us explain what we mean. At some fundamental level, Nature is described by a field theory with some number of fermion fields. We adopt the convention, commonly used in the grand unification literature, of treating all fermion fields as two-component left-handed Weyl fields.[1] Thus, for example, the electron family consists of 15 fermion fields: e^-, e^+, ν, u, $\bar{u}$, d, and $\bar{d}$. Then in the real world the electric charge Q has the noteworthy property that, for every fermion field with charge $+q$, there is a fermion field with charge $-q$ if $q \neq 0$. We say that the electric charge is vectorlike.

In the real world, charge is certainly vectorlike. Otherwise, there will be massless charged [2] fermion fields (provided that the electromagnetic gauge invariance is not spontaneously broken). The theory will be beset by such severe infrared difficulties that an S matrix presumably cannot be defined in the conventional way. Also, experiments strongly indicate that the electromagnetic current is purely vector, without an axial vector piece.

However, even if the S matrix may not exist, the field theory could still be perfectly sensible in the sense that all off-shell Green's functions are well-defined. Indeed, the only requirement for quantum electrodynamics to be renormalizable and hence defined is that it be anomaly-free [3]: $\mathrm{tr}Q^3 = 0$. For example, quantum electrodynamics with a charge-2 fermion field ψ_L and eight charge-(-1) fermion fields ψ_{iL} ($i = 1,...,8$) should be a sensible field theory. It is renormalizable, and all (off-shell) Green's functions may be calculated. What does the physical spectrum of such a theory look like? Unfortunately, field theorists are at present far from being able to answer such question. Perhaps the local U(1) invariance is dynamically broken. One could conceivably consider vacuum condensation of the order parameters $\psi_L C\psi_L$, $\psi_L C\psi_{iL}$, $\psi_{iL} C\psi_{jL}$. But strong forces are not available to drive the formation of condensates.

Why does Nature arrange things so that electric charge is vectorlike? We cannot provide a truly satisfactory answer to this fundamental question. Perhaps some day one could show that field theories with a nonvectorlike electric charge are not sensible. It is not unreasonable to conjecture that such theories will dynamically break electric charge. The scope of this paper is more modest. Our purpose here is to explore this question, to examine some examples which might shed some light, and to prove some assertions under certain specific circumstances.

The mystery of why electric charge is vectorlike is underscored in the context of grand unified theories of weak, electromagnetic, strong, and possibly technistrong interactrions. In these theories, we have the condition $\mathrm{tr}Q = 0$ in addition to $\mathrm{tr}Q^3 = 0$. (Thus, the simple example above would not be acceptable in this context.) These theories describe the world in terms of a gauge theory based on some simple gauge group G, which is subsequently broken down to some subgroup [in the real world this is presumably $\mathrm{SU}(3)_{\mathrm{color}} \times \mathrm{U}(1)$]. This subgroup is to contain one and only one U(1) factor, which is then identified with electromagnetism. [Why the gauge group G should be broken down to contain one and only one U(1) factor[4,2] is another mysterious, and at present unanswerable, question.] The question raised in this paper is then the following: Why in the world should this particular U(1) generator, when evaluated over the fermions, be vectorlike?

The situation is especially mysterious if elementary Higgs fields are responsible for the symmetry breaking. In general, there is no particular connection between the fermion representations and the Higgs representations used. With dynamical symmetry breaking caused by fermion bilinear condensation,[5] there might be a deep connection. This might be considered as another argument in favor of the view that elementary Higgs fields are to be regarded as unnatural. It is tempting to conjecture that the dynamics of symmetry breaking is such as to leave no nonvectorlike U(1) unbroken.[6]

In the Georgi–Glashow[7] SU(5) theory, electric charge is vectorlike with the usual $\bar{5}$ and 10 assignment of fermions. The group SU(5) is first broken into $\mathrm{SU}(3)\times\mathrm{SU}(2)\times\mathrm{U}(1)$ by an adjoint of Higgs and then into $\mathrm{SU}(3)\times\mathrm{U}(1)$ by a fundamental of Higgs. This symmetry-breaking chain uniquely fixes the charge generator Q to be $Q = (1/3, 1/3, 1/3, -1, 0)$. We use $(a_1, a_2,...,a_N)$ to denote a diagonal $N \times N$ matrix with diagonal elements equal to $(a_1, a_2, ..., a_N)$.) To illustrate our discussion, let us suppose that the fermion content is *not* the usual one but corresponds to the anomaly-free representation $9[\bar{1}] + \{2\}$. (We use the following notation for the irreducible representations of SU(n): $[k]$ denotes the totally antisymmetric tensor with k upper indices, $[\bar{k}] = [n-k]$

 0022-2488/81/102263-06$01.00

denotes the totally antisymmetric tensor with k lower indices, $\{k\}$ denotes the totally symmetric tensor with k upper indices, $[^k_m]$ denotes the traceless tensor with k upper and m lower indices antisymmetric among the upper and lower indices, and so forth.)

With the standard breaking of $SU(5) \to SU(3) \times SU(2) \times U(1)$, one can easily work out the fermion spectrum to be

$(\mathbf{6}, \mathbf{1}; -\frac{2}{3})$,

$(\mathbf{3}, \mathbf{2},; \frac{2}{3}, -\frac{1}{3})$,

$(\mathbf{1}, \mathbf{3}; 2, 1, 0)$,

nine $(\bar{\mathbf{3}}, \mathbf{1}; \frac{1}{3})$,

nine $(\mathbf{1}, \mathbf{2}; 0, -1)$.

The first two numbers in the parenthesis denote the SU(3) and SU(2) representations, respectively. The numbers after the semicolon indicate the electric charges. Electric charge is certainly not vectorlike. For instance, the lepton sector consists of one field with charge $+2$, one field with charge $+1$, nine fields with charge -1, and ten neutral fields. (Incidentally, in this particular model $\mathrm{tr}Q^3$ vanishes in the quark and lepton sectors separately.) Note also that this model has quarks belonging to a color sextet.[8] The question of whether there is a deep connection between having only color-triplet quarks and the electric charge being vectorlike naturally suggests itself. In an SU(5) theory, the requirement that quarks are color-triplets implies that electric charge is vectorlike. However, the converse is not true. We will exhibit a counterexample to the converse statement.

The example above shows that, given some anomaly-free respresentation of SU(5), it is extremely unlikely that the Georgi–Glashow charge when evaluated over the fermions will be vectorlike. It appears truly remarkable that the fermion assignment $[4] + [2]$ (i.e.,$\bar{5} + 10$) leads to a vectorlike charge.

Not being able to answer the dynamical question of why electric charge is vectorlike, we retreat from physics to mathematics. In this paper we would like to pose the following mathematical question: Given some simple group G [say $SU(N)$] and some anomaly-free fermion representation R, does there exist a charge generator $Q = (a_1, a_2, ..., a_N)$ such that Q is vectorlike over the fermions? Mathematically, this corresponds to the requirement that the representation R be real under the U(1) of electromagnetism.

Note that the statement that Q is vectorlike is equivalent to saying that

$$\mathrm{tr}Q^{2k+1} = 0 \qquad \text{for } k = 0, 1, 2, ..., \infty. \tag{1}$$

[This trace is over all fermions. Q being vectorlike means that for every fermion of charge $+q$ there is also one of charge $-q$. Their contributions to the trace in Eq. (1) thus cancel against each other.] What we are saying is that our present understanding of physics tells us that there are good reasons (simple grand unification and renormalizability) to demand $\mathrm{tr}Q = 0$ and $\mathrm{tr}Q^3 = 0$. (Note these reasons only became known not that long ago.) Will we ever discover deep reasons for demanding the entire set of conditions in (1)?

In a gauge theory broken down to a U(1) of electromagnetism, there are additional constraints on Q. For instance, in an $SU(N)$ theory with $Q = (a_1, a_2,...,a_N)$, the numbers a_k must be such that there exist sets of integers n_k so that $\sum_{k=1}^{N} n_k a_k = 0$. This holds whether the symmetry breaking is due to elementary or to composite Higgs fields.

In the real world, Q is actually vectorlike over the quark and lepton sectors separately. Also, color is vectorlike. In contrast, baryon number B and lepton number L, which are not associated with a local invariance, may not be vectorlike. [In the standard SU(5) model, B is vectorlike but L is not.] We will not investigate here the implications of demanding vectorlike color.

Henceforth in this paper we shall use the following terminology: For a given group G, an anomaly-free set of fermion representations, R, and a charge generator Q, which is vectorlike over R will be referred to as a "solution".

We now list a number of (trivial) observations.

Observation 1: If the representation R is real under the group G, then Q is vectorlike.

Observation 2: For any representation R, Q will be vectorlike over R if Q is represented in the fundamental representation by a matrix of the form

$$\mathrm{diag}(a, -a, b, -b, ..., c, -c, 0, 0, ..., 0).$$

A charge generator of this form will be referred to as "automatically" vectorlike.

We refer to these two classes of solutions ("real" solutions and "automatically vectorlike" solutions") as "trivially" vectorlike solutions. Note the Georgi–Glashow charge is not trivially vectorlike. We do not wish to imply, however, that Nature may not choose a trivially vectorlike electric charge. In what follows we will exhibit infinite classes of *nontrivial* solutions. However, we have not been able to answer the question of whether these solutions comprise *all* solutions, though we have found no others. Thus it remains an open problem to find the general solution to the stated mathematical problem.

Clearly, since the only simple groups that admit of complex representations are $SU(N)$, $SO(4N+2)$, and $E(6)$, only these groups can have nontrivial solutions. We will discuss each of these cases in turn. First we make two remarks:

Remark 1: If $\{R_{(N)}, Q_N = (a_1, ... a_N)\}$ is a solution for $SU(N)$, then $\{R_{(N+1)}, Q_N = (a_1, ..., a_N, 0)\}$ is a solution of $SU(N+1)$ if $R_{(N+1)}$, decomposes to $R_{(N)}$ under an $SU(N)$ subgroup of $SU(N+1)$.

Remark 2: If $\{R_1, Q\}$ and $\{R_2, Q\}$ are solutions for a group G, then $\{(R_1 \otimes R_2), Q\}$ is a solution as well. (Note that if R_1 and R_2 are anomaly-free, then so is $R_1 \otimes R_2$.)

We now exploit these two remarks to construct infinite classes of solutions.

I. SU(N)

Case A: R contains only totally antisymmetric representations.

In this case, by Remark 1, if we find a representation $R_{(N+1)}$ of $SU(N+1)$ such that $R_{(N+1)}$ reduces to a *real* representation under an $SU(N)$ subgroup of $SU(N+1)$, then $Q = (a_1, a_2, ... a_N, 0)$ is a solution. If N is odd, then $R_{(N+1)}$ is also real (as can easily be verified if we note that

$[k]_{N+1} \to [k]_N + [k-1]_N$) and so we obtain just another trivial solution. But if $N = 2p$, then we have the following nontrivial solution:

$$\mathrm{SU}(2p+1), \qquad Q = (a_1, \ldots, a_{2p}, 0), \qquad R = \{[2], [4], \ldots, [2p]\} \tag{2}$$

with a_k arbitrary except that $\Sigma a_k = 0$.

In the light of this discussion, the Georgi–Glashow solution

$$\{\mathrm{SU}(5), Q = (1/3, 1/3, 1/3, -1, 0), R_{\mathrm{GG}} = [2] + [4]\}$$

is revealed to be a solution precisely because when one reduces SU(5) to SU(4), $R_{\mathrm{GG}} \to [0] + [1] + [2] + [3] = \mathbf{1} + \mathbf{4} + \mathbf{6} + \bar{\mathbf{4}}$, a real representation of SU(4).

The solution in Eq.(2) can be embedded in an orthogonal group leading to the following nontrivial solution for SO($4p+2$):

$$\mathrm{SO}(4p+2), \qquad Q = \sum_{i=1}^{2p} a_i \epsilon_i, \qquad R = (2^{2p})\text{-dimensional spinor}. \tag{3}$$

[Of course, SO(10) is a special case of (3) with $p = 2$.] The notation for Q is as follows.[9] We denote the components of the spinor representation of SO($4p+2$) by

$$|\epsilon_1 \epsilon_2 \cdots \epsilon_{2p+1}\rangle, \tag{4}$$

where $\epsilon_i = \pm 1$ and $\Pi_{k=1}^{2p+1} \epsilon_k = +1$ (or -1). This notation makes obvious why (3) represents a solution. Since $a_{2p+1} = 0$, we can reverse the sign of Q by letting $\epsilon_k \to -\epsilon_k$, $k = 1, \ldots, 2p$, and $\epsilon_{2p+1} \to \epsilon_{2p+1}$.

Evidently, we can apply remark (1) repeatedly to the solution in Eq. (2) to generate additional nontrivial solutions. For example, starting with the Georgi–Glashow solution for SU(5) we can construct the nontrivial solution for SU(6): $R = 2[\bar{1}] + [2] = \bar{\mathbf{6}} + \bar{\mathbf{6}} + \mathbf{15}$ and $Q = (a, b, c, -(a+b+c), 0, 0)$. [Incidentally, this 27 dimensional representation may be embedded in E(6)]. Repeating this procedure an arbitrary number of times, we find the following nontrivial solutions:

$$\mathrm{SU}(2p), \qquad Q = (a_1, \ldots, a_{2k}, 0, 0, 0, \ldots, 0),$$
$$R = \{\eta_1[1], \eta_2[2], \ldots, \eta_{p-1}[p-1]\},$$
$$\eta_l = \sum_{q=1}^{p-k} (-1)^{q+l} \beta_q \binom{p-l+q-1}{p-l-q} \cdots,$$

where $\beta_1, \ldots, \beta_{p-k}$ are arbitrary integers; (4a)

$$\mathrm{SU}(2p+1), \qquad Q = (a_1, \ldots, a_{2k}, 0, \ldots, 0),$$
$$R = \{\eta_1[1], \ldots, \eta_p[p]\},$$
$$\eta_l = \sum_{q=0}^{p-k} (-1)^{q+l} \beta_q \binom{p-l+q}{p-l-q} \cdots, \tag{4b}$$

where $\beta_0, \ldots, \beta_{p-k}$ are arbitrary integers. (Here we interpret $-[r]$ as $[r]^*$.)

We have not succeeded in finding any other nontrivial solutions for SU(N) with only antisymmetric fermion representations (see the Appendix). If any exist, then it is easy to show that there must be such nontrivial solutions for which Q has no vanishing diagonal entries (see the Appendix). [All of the solutions in (4) necessarily have a Q with some vanishing diagonal entries because of how they were constructed.]

It seems probable that (4a) and (4b) represent the most general nontrivial solutions for SU(N) involving only antisymmetric fermion representations. We have, however, not been able to prove this assertion.

Case B: R contains representations of mixed symmetry under permutation.

We may use Remark 2 to generate an infinite number of solutions. Given that for SU(N) and charge Q the sets of representations $R_1, R_2, \cdots$ are solutions, then $R \equiv \Pi_i R_i$ is also a solution.

Let us look at an example. Consider SU(5) again with the charge operator $Q = (1, 0, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3})$. We may construct a nontrivial solution involving mixed representations by multiplying the Georgi–Glashow solution $R_{\mathrm{GG}} = [2] + [4]$ by the real representation $R_{\mathrm{real}} = [2] + [3]$:

$$R = R_{\mathrm{GG}} \times R_{\mathrm{real}}$$
$$= \begin{bmatrix} 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \text{(real representations)},$$

i.e.,

$$R = \overline{50} + 40 + 10 + \text{(real representations)}.$$

Suppose we take this to be an SU(5) model with the standard pattern of breakings

$$\mathrm{SU}(5) \to \mathrm{SU}(3)_c \times \mathrm{SU}(2) \times \mathrm{U}(1)$$
$$\to \mathrm{SU}(3)_c \times \mathrm{U}(1)_{\mathrm{em}};$$

then we display the resulting content in Table I.

This model thus represents a counterexample to the conjecture that an SU(5) theory which admits a vectorlike electric charge and which has purely V-A weak interaction necessarily contains no quarks which are nontriplets of color. ("Shiny quarks" in the terminology of the second paper in Ref.8.) Our model *appears* to be the smallest (in the number of fields) SU(5) model which contains shiny quarks and a vectorlike electric charge. Thus, the observation of a shiny quark will imply either the existence of at least 100 new fermion fields or that SU(5) is incorrect as a classification scheme. Within the grand unification framework, the existence of shiny quarks is correlated with a rich fermionic spectrum.[10]

The same construction goes through for other groups, of course. For instance, in SO(10), solutions may be generated by multiplying $16 \times 16 = 10 + 120 + 126$ and $16 \times 10 = 16 + 144$. Thus, the following is always a solution:

any group G, Q,

$$R = \prod_i R_i, \qquad \text{where } R_i = \text{solution for } G, Q. \tag{5}$$

Are there any other nontrivial solutions for SU(N) which are not of the form discussed here? We have not found any. Whether any exist is, however, an open question.

Given a specific group and a specific representation R, it is not difficult, though it is tedious, to search for a charge operator Q which is vectorlike over R. The obvious procedure consists of determining the maximum and minimum charges in R, matching them, and then repeating the proce-

TABLE I. (d,I_3) denotes a d-dimensional representation of SU(3) of color, with value I_3 of the third component of weak isospin.

$Q =$	2	$\frac{5}{3}$	$\frac{4}{3}$	1	$\frac{2}{3}$	$\frac{1}{3}$	0	$-\frac{1}{3}$	$-\frac{2}{3}$	-1	$-\frac{4}{3}$	$-\frac{5}{3}$	-2
$10 \rightarrow$				$(1,0)$	$(3,\frac{1}{2})$			$(3,-\frac{1}{2})$	$(\bar{3},0)$				
$40 \rightarrow$				$(8,0)$	$(\bar{6},\frac{1}{2})$	$(\bar{3},1)$		$(\bar{6},-\frac{1}{2})$	$(\bar{3},0)$	$(1,\frac{1}{2})$		$(\bar{3},-1)$	$(1,-\frac{1}{2})$
					$(3,\frac{1}{2})$			$(3,-\frac{1}{2})$	$(\bar{3},0)$				
$50 \rightarrow$	$(1,0)$	$(3,\frac{1}{2})$	$(6,1)$		$(3,-\frac{1}{2})$	$(6,0)$	$(8,\frac{1}{2})$		$(6,-1)$	$(8,-\frac{1}{2})$	$(\bar{6},0)$		
						$(\bar{3},0)$							
total no. of fields $=$	1	3	6	9	15	12	8	12	15	9	6	3	1 = 100

color sixes: $\begin{pmatrix} \frac{4}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix}_L$

leptons: $\begin{pmatrix} -1 \\ -2 \end{pmatrix}_L$, $(+1)_L$, $(+2)_L$

quarks: $\begin{pmatrix} \frac{5}{3} \\ \frac{2}{3} \end{pmatrix}_L$, $\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}_L$, $\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}_L$ color sixes: $\begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}_L$, $(-\frac{4}{3})_L$

anti-quarks: $\begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ -\frac{5}{3} \end{pmatrix}_L$, $(-\frac{2}{3})_L$ $(-\frac{2}{3})_L$, $(+\frac{1}{3})_L$ color octets: $\begin{pmatrix} 0 \\ -1 \end{pmatrix}_L$, $(+1)_L$

dure (see the Appendix). We have labored over a number of cases in this manner, and the experience we gained from these exercises tends to suggest that, given an arbitrary anomaly-free representation, the existence of a *nontrivially* vectorlike electric charge is rather unlikely.

II. SO(4N + 2)

We need only concern ourselves with spinorial representations of SO($4N + 2$), as all of the tensor representations are real and thus are trivially vectorlike.

The case of SO($4N + 2$) is reducible to the case of U($2N + 1$) in the following sense. Q defined on the vector representation is a ($4N + 2$) by ($4N + 2$) antisymmetric real matrix. It can therefore be brought by a basis transformation to the block form [H]

$$\begin{pmatrix} 0 & a & & & & 0 \\ -a & 0 & & & & \\ & & \ddots\,0 & b & & \\ & & -b & 0 & & \\ & 0 & & & \ddots\,0 & c \\ & & & & -c & 0 \end{pmatrix}.$$

The U(1) generated by this matrix is easily seen to belong to a U($2N+1$) subgroup of SO($4N + 2$). Thus a knowledge of nontrivial solutions for U($2N + 1$) will give us nontrivial solutions for SO($4N + 2$).

Here we will explicitly discuss only the 2^{2N}-dimensional spinor representation of SO($4N + 2$). Under a U($2N + 1$) subgroup this spinor decomposes into $\{[0] + [2] + \cdots + [2N]\}$. It can be shown (see the Appendix) that this is a solution for *all* and *only* those Q which are of the form $Q = (a_1, \ldots, a_N, 0)$ [where we have written Q as an U($2N + 1$) generator]. In fact, for $\Sigma a_k = 0$, this is just the SO($4N + 2$) embedding of the SU($2N + 1$) solution given in Eq. (2).

This solution is covered by our Remark 1. The form of Q is such that we may break SO($4N + 2$) down to SO($4N$), under which all representations are real.

Indeed, the preceding remark implies that higher-dimensional spinorial representations of SO($4N + 2$) also are solutions for $Q = (a_1, \ldots, a_N, 0)$ [written again as a U($2N + 1$) generator]. All of these spinorial representations may be constructed as products of 2^{2N}-dimensional spinors and vectors.

III. E(6)

E(6) has a maximal subgroup that may be denoted $SU(3)_A \times SU(3)_B \times SU(3)_C$. The generators λ_{Ia}, $I =$ A, B, C, $a = 3$ or 8 are six mutually commuting generators of E(6) (which is a rank 6 group). In some basis we may express the charge generator as a linear combination of the six

$$\lambda_{Ia}: Q = \sum_{\substack{I = A,B,C \\ a = 3,8}} \alpha_{Ia} \lambda_{Ia}.$$

Under the $SU(3)_A \times SU(3)_B \times SU(3)_C$ subgroup the fundamental representation which has dimension 27 transforms as

$$27 = (3,\bar{3},1) + (1,3,\bar{3}) + (\bar{3},1,3).$$

If, therefore, $\alpha_{Aa} = \alpha_{Ba}$ (or $\alpha_{Ba} = \alpha_{Ca}$ or $\alpha_{Ca} = \alpha_{Aa}$), then clearly Q is vectorlike over the 27-dimensional respresentation. For then, under the Weyl reflection symmetry which takes $SU(3)_A \leftrightarrow SU(3)_B$, $Q \rightarrow Q$, and $27 \rightarrow 27^*$. This gives us a four-parameter family of solutions. Under the $SU(6) \times SU(2)$ subgroup of E(6), this four-parameter family can be written

$$\begin{pmatrix} a & & & & & \\ & b & & & & \\ & & c & & & \\ & & & -(a+b+c) & & \\ & & & & d & \\ & & & & & -d \end{pmatrix} + \begin{pmatrix} d & \\ & -d \end{pmatrix}.$$

This charge generator is vectorlike not only over the 27 but over all E(6) representations, which transform as $(r_1,r_2,r_3) + (r_3,r_1,r_2) + (r_2,r_3,r_1)$ under the $SU(3)_A \times SU(3)_B \times SU(3)_C$ subgroup.

We have found no solutions that are not of this form.

IV. CONCLUDING REMARKS

Let us look at an SU(5) model in which the fermions are in the anomaly-free set of representations consisting of some number of "families" of $[2]_L + [4]_L$. Let us assume that the vacuum expectation value of some Higgs in the adjoint representation breaks SU(5) down to a smaller group at some mass scale M. SU(5) may be broken to one of the following groups depending on the vacuum expectation value of the adjoint Higgs: $SU(4)\times U(1)$, $SU(3)\times SU(2)\times U(1)$, $SU(3)\times U(1)^2$, $SU(2)^2\times U(1)^2$, $SU(2)\times U(1)^3$, or $U(1)^4$. In *any* of these cases in the absence of further breaking there are unbroken U(1) gauge interactions remaining. Furthermore—if there is no further breaking—*at least one* of these unbroken U(1)'s is *nonvectorlike*. For example, if SU(5) breaks as in the Georgi–Glashow model to $SU(3)_C \times SU(2)\times U(1)_y$ *and does not break further*, then we are left with an exact U(1) of hypercharge which is clearly not vectorlike over the fermions:

$$Y(5)^* = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, \quad Y(10) = \begin{pmatrix} & -\frac{1}{6} & \\ & -\frac{1}{6} & \frac{2}{3} \\ -1 & -\frac{1}{6} & \frac{2}{3} \\ & -\frac{1}{6} & \frac{2}{3} \\ & -\frac{1}{6} & \\ & -\frac{1}{6} & \end{pmatrix}. \qquad (6)$$

Suppose now that a Higgs, ϕ^{α}, in the fundamental representation of SU(5) acquires a nonvanishing vacuum expectation value. Then whatever U(1) generators remain unbroken must have at least one vanishing diagonal entry [i.e., if $\langle \phi^k \rangle \neq 0$, then $Q = (a_1,...,a_5)$ will only remain unbroken if $a_k = 0$]. But *all* such generators are vectorlike over a family of fermions [Eq. (2)], so that a further breaking done by the vacuum expectation value of a fundamental Higgs is sufficient to insure that *only vectorlike* $U(1)$ *charges remain unbroken*! *Without* further breaking, on the other hand, it must be that some nonvectorlike U(1) charges remain unbroken.

There might be a significant clue here to the structural properies of grand unified theories. To the extent that all the nontrivial solutions we found are such that at least one field in the fundamental representation has vanishing charge, it appears that breaking by a Higgs in the fundamental representation is necessary (or, equivalently, condensation in the fundamental channel, if one prefers). Grand unified theories presently on the market all have this feature.

We conclude that a deep theoretical understanding of why charge is vectorlike may be important to a further development of fundamental theories. The requirement of a vectorlike charge in the context of grand unification appears to have far reaching implications.

Further work, such as an exhaustive listing of all nontrivial solutions, would be welcome. We have also, in this paper, avoided the difficult dynamical question of why a theory would choose a symmetry-breaking channel so as to guarantee a vectorlike charge.

ACKNOWLEDGMENTS

One of us (A. Z.) thanks F. Gürsey and B. Stech for informative discussions on E(6). This work was supported in part by the U. S. Department of Energy under Contract DE-AC06-76ER01388.

APPENDIX

As mentioned in the text, it is an open question whether the solution given in Eq. (4) is the most general nontrivial solution for SU(N) involving only antisymmetric fermion representations. If *other* nontrivial solutions exist for this case then there must exist such nontrivial solutions for which $Q = (a_1,...,a_N)$ *has no vanishing diagonal elements*, i.e., $a_i \neq 0$ for all $i = 1,...,N$. [This is obvious. For suppose all nontrivial solutions for SU(N) with antisymmetric fermion representations have Q with some diagonal elements vanishing. Suppose $(N - M)$ of the a vanish. We can look at such a solution under the SU(M) subgroup of SU(N) in which Q has no vanishing diagonal elements. This gives us a solution for SU(M) which by assumption must be a trivial one. But this implies that under SU(M) the fermion representations are real. This in turn means that the original SU(N) solution was of the form given in Eq. (4).]

In the search for such exceptional solutions (we have found none), the following theorem simplifies things.

Theorem: Let SU(N) have a solution consisting of a set R of antisymmetric fermion representations (containing no real representations) and a charge $Q = (a_1,a_2,...,a_N)$ with $a_i \neq 0$ (all i). We may, without loss of generality, assume $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_k > 0 > a_{k+1} \geqslant ... \geqslant a_N$. Then there $\exists\ i,j$ such that $i < k < j$, and such that either $[i], \overline{j} \in R$ or $\overline{[i]}$, $[j] \in R$ and such that $[l]$, $\overline{[l]} \notin R$ for any $i < l < j$ (in particular $[k]$ and $\overline{[k]} \notin R$).

This theorem is proven by demanding that $Q_{\max} = -Q_{\min}$, where $Q_{\max}$ is the largest fermion charge and $Q_{\min}$ is the smallest fermion charge.

Corollary: For SU($2p + 1$), $R = \{[2],[4],...,[2p]\}$ the only nontrivial solutions have Q of the form $Q = (a_1,...,a_{2p},0)$.

Proof: Suppose $Q = (a_1,...,a_{2p+1})$, $a_i \neq 0$. Then aply the theorem. For *every* $[l]$ either $[l]$ or $\overline{[l]} \in R$ (if l = even, $[l] \in R$; if l is odd, $\overline{[l]} \in R$. But we know that $[k]$ and $\overline{[k]}$ do not belong to R. Thus we have a contradiction. Thus one of the a_i must vanish.

[1] H. Weyl, Z. Phys. **56**, 330 (1929).

[2] People have argued that massless charged particles are not allowed: K. M. Case and S. Gasiorowicz, Phys. Rev. **125**, 1055 (1962); T. Pradhan, Nuovo Cimento **38**, 883 (1965).

[3] S. Adler, Phys. Rev. **177**, 2426 (1969); J. Bell and R. Jackiw, Nuovo Cimento A **60**, 47 (1969).

[4] One might speculate on the existence of more than one "photon" in the

world. S. M. Barr and A. Zee, Phys. Lett. B **92**, 297 (1980).
[5]S. Dimopolous, S. Raby, and L. Susskind, "Tumbling Gauge Theories," Stanford Preprint ITP 653, Stanford University, 1979.
[6]The explicit examples given in Ref. 5 do not contain any broken local U(1)'s.
[7]H. Georgi and S. L. Glashow, Phys. Rev. Lett. **32**, 438 (1974).
[8]G. Karl, Phys. Rev. D **14**, 2374 (1976); F. Wilczek and A. Zee, Phys. Rev. D **16**, 860 (1977); C. B. Dover, T. K. Gaisser, and G. Steigman, Phys. Rev. Lett. **42**, 1117 (1979); H. Georgi and S. L. Glashow, HUTP-79/A027, 1979.
[9]F. Wilczek and A. Zee, unpublished.
[10]This point has also been emphasized in the fourth paper cited in Ref. 8.
[11]For example, J. N. Herstein, *Topics in Algebra* (Xerox College Publishing, Waltham, Mass., 1964), p.306.

Some applications of Postnikov systems in gauge theories

L. Bonora, A. de Pantz, and P. Pasti
Instituto di Fisica dell'Universitá di Padova, Instituto Nazionale di Fisica Nucleare—Sezione di Padova, 35100 Padova, Italy

(Received 30 September 1980; accepted for publication 1 May 1981)

We use Postnikov systems for spaces relevant to gauge theories in order to derive results on characteristic classes and to prove Singer-like theorems about the impossibility of a continuous gauge choice on compactified space–time manifolds.

PACS numbers: 11.10.Np

In recent years classical manifolds such as spheres, tori, and projective spaces have been growing more and more important in gauge field theory due to the rich structure one can obtain on these geometrical configurations. This is achieved by compactifying the Euclidean space–time (or a subspace of it) with suitable conditions at infinity.

Whereas the geometrical approach to gauge theories built over spheres has been widely studied, the same attention has not been paid, for instance, to the geometry of the torus, though a gauge theory on such a manifold seems to have interesting features as to the confinement problem.[1]

In this paper we report some results relevant to the geometry of a gauge theory defined on tori and other compact manifolds, obtained by using the Postnikov system algorithm. We shall first study in detail theories with an $\mathrm{SU}(N)$ or an $\mathrm{SU}(N)/Z_N$ gauge group, and use this case as a pattern for other usual semisimple groups.

Let us recall briefly what a Postnikov system is.[2–4] If a topological space X is simply connected, one can find a family of spaces X_n, $n\geqslant 2$, such that

$$\begin{aligned} K(\pi_4,4)&\xrightarrow{i_4} X_4 \xrightarrow{k_4} K(\pi_5,6),\\ &\qquad\ \downarrow p_4\\ K(\pi_3,3)&\xrightarrow{i_3} X_3 \xrightarrow{k_3} K(\pi_4,5),\\ &\qquad\ \downarrow p_3\\ & \qquad X_2 = K(\pi_2,2),\end{aligned}\tag{1}$$

where the p_n are fibrations with fiber $K(\pi_{n+1}, n+1)$, k_n inducing maps, i_n inclusion maps. $K(\pi_n,m)$ are Eilenberg–MacLane spaces, and π_n are the homotopy groups of X. The X_n spaces approximate X, in the sense that if we denote by K an n-dimensional complex, then the homotopy classes $[K,X]=[K,X_n]$.[5]

We first apply this construction to classify the inequivalent principal G-bundles, $G=\mathrm{SU}(N)/Z_N$ over a manifold M. The classification is given by the homotopy classes $[M, BG]$, where BG is the base space of the universal G-bundle. As $\pi_n(BG)\approx\pi_{n-1}(G)$, and G is connected, then BG is simply connected and there exists a Postnikov system for it. Let $M^{(4)}$ be a four-dimensional compact oriented manifold. Then we are interested in $[M^{(4)},BG_4]$. Following a procedure analogous to Ref. 4, we can derive from (1) the following exact sequence:

$$0\rightarrow Z\xrightarrow{i_4^*}[M^{(4)},BG_4]\xrightarrow{p_3^*}H^2(M^{(4)},Z_N)\rightarrow 0,\tag{2}$$

where i_4^*,p_3^* are induced by i_4 and p_3, respectively.

One must be careful in dealing with this exact sequence because it is an exact sequence of sets as $[M^{(4)}, BG_4]$ has no natural group structure and, by itself, it means only that the image of i_4^* is mapped by p_3^* into the trivial element of $H^2(M^{(4)}, Z_N)$. However, going through the Postnikov construction (see the Appendix), one realizes that to every element of $H^2(M^{(4)},Z_N)$ there corresponds a subset of $[M^{(4)},BG_4]$ which may be given the same group structure as Z. So we write formally $[M^{(4)}, BG_4]/Z$ to mean a set in one-to-one correspondence with $H^2(M^{(4)}, Z_N)$.

We see that inequivalent bundles over a manifold $M^{(4)}$ are classified according to the group of integers and the group $H^2(M^{(4)}, Z_N)$. If $M^{(4)}$ is the four-dimensional torus T^4, we get by standard methods[5] $H^2(T^4,Z_N)\approx 6Z_N$ (that is, the direct sum of six groups Z_N), which means that for any Chern class $\in H^4(T^4,Z)\approx Z$ we have N^6 inequivalent fiber bundles, or, in physical language, for any instanton sector there are N^6 gauge inequivalent periodic boundary conditions.[1] We give some more examples in four dimensions. $H^2(S^4,Z_N)=0$, $H^2(RP^4,Z_N)\approx 0(Z_2)$ if N is odd (even), $H^2(S^2\times S^2,Z_N)\approx 2Z_N$, $H^2(S^3\times S^1,Z_N)=0$.

If M is a lower-dimensional manifold, the Postnikov system gives simply $[M,BG]\approx H^2(M,Z_N)$. Then if M is the three-dimensional torus T^3, we have $H^2(T^3,Z_N)\approx 3Z_N$. Moreover, $H^2(S^3,Z_N)=0$, $H^2(RP^3,Z_N)\approx H^2(RP^4, Z_N)$, $H^2(S^2\times S^1, Z_N)\approx Z_N$.

In two dimensions there exists a classification of 2-manifolds or surfaces[6] M_g, with g a nonnegative integer, and N_h, h a positive integer. One finds easily $H^2(M_g,Z_N)\approx Z_N$ for all g and $H^2(N_h,Z_N)\approx 0(Z_2)$ if N is odd (even) for all h. We recall that $M_0=S^2$, $M_1=T^2$, $N_1=RP^2$.

The Postnikov construction turns out to be very useful also in another context, that is, in showing the impossibility of a continuous gauge choice in a gauge theory defined on compactified spaces. Let us consider a gauge field theory defined on a principal fiber bundle $P(M,G)$. Let $\mathscr{C}$ be the set of connections on P, and $\mathscr{G}$ the group of gauge transformations, that is to say, the group of automorphisms of P,[7] defined by $p\rightarrow p\cdot\gamma(p)$ for all $p\in P$, where $\gamma: P\rightarrow G$ is any smooth function satisfying $\gamma(pg)=g^{-1}\gamma(p)g$, $\forall g\in G$. If $\omega\in\mathscr{C}$, then $\gamma\in\mathscr{G}$ transforms ω into

$$\omega'=\mathrm{ad}\gamma^{-1}\omega+\gamma^{-1}d\gamma.\tag{3}$$

The action of $\mathscr{G}$ splits $\mathscr{C}$ into orbits, In the path integral approach to quantum field theory the problem arises of choosing a representative for each orbit in a continuous way. This problem has already been appropriately settled in fiber

 0022-2488/81/102269-03$01.00

bundle language and solved in the case $M = S^3$ or S^4 by Singer[8]: When G is a semisimple Lie group, there is no continuous gauge choice. When M is a less simple manifold, Singer's proof must be partially modified. This is what we want to illustrate here. At first we choose $G = \mathrm{SU}(N)$. In order to have a well-defined mathematical structure, we consider the subset $\mathscr{C}$ of irreducible connections and the (sub) group $\tilde{\mathscr{G}} = \mathscr{G}/Z_N$ of gauge transformations, where Z_N is the center of $\mathrm{SU}(N)$.

Indeed, as one sees from Eq. (3), the γ's that leave unchanged a fixed connection ω must obey the equation $[\gamma,\Omega] = 0$, where Ω is the curvature of ω. On the basis of the Ambrose–Singer theorem,[9] γ must commute with the Lie algebra of the holonomy group. When $\omega \in \tilde{\mathscr{C}}$, it follows that γ is constant and belongs to the center of the group G. The converse is trivial and therefore is one wishes a free group action on $\mathscr{C}$ one must refer to the group $\tilde{\mathscr{G}}$. As a consequence, the space $N = \tilde{\mathscr{C}}/\tilde{\mathscr{G}}$ turns out to be[8] the base space of a principal fibre bundle whose group is $\tilde{\mathscr{G}}$. The possibility of choosing a representative for each orbit in a continuous way would imply the existence of a global cross section in this principal fibre bundle, that is, its triviality. Therefore, $\pi_i(\tilde{\mathscr{C}}) = \pi_i(N) \oplus \pi_i(\tilde{\mathscr{G}})$. Now, as Singer showed, all the homotopy groups $\pi_i(\tilde{\mathscr{C}})$ are trivial; so a necessary condition for the existence of a continuous gauge choice is

$$\pi_i(\tilde{\mathscr{G}}) = 0 \quad \text{for all } i. \tag{4}$$

From now on, our procedure will consist in taking Eqs. (4) as hypothesis and looking for contradictions with them. Let us define $\mathscr{G}_0$ as the subset of $\mathscr{G}$ formed by the gauge transformations taking the value of $e \in \mathrm{SU}(N)$ in a fixed point $x_0 \in M$.[10] We take from Ref. 7 the following fundamental theorem: If $\dim M \leqslant 4$, $\mathscr{G}_0$ is weakly homotopically equivalent to the set $(\mathrm{SU}(N))^M$ of all maps $f: M \to \mathrm{SU}(N)$ such that $f(x_0) = e$.[5]

Let us consider the following fibrations, $0 \to Z_N \to \mathscr{G} \to \tilde{\mathscr{G}} \to 0$, $0 \to \mathscr{G}_0 \to \mathscr{G} \to \mathrm{SU}(N) \to 0$, and the relevant exact sequences

$$\begin{aligned} &\to 0 \to \pi_1(\mathscr{G}) \to \pi_1(\tilde{\mathscr{G}}) \to Z_N \to \pi_0(\mathscr{G}) \to \pi_0(\tilde{\mathscr{G}}),\\ &\to 0 \to \pi_1(\mathscr{G}_0) \to \pi_1(\mathscr{G}) \to 0 \to \pi_0(\mathscr{G}_0) \to \pi_0(\mathscr{G}) \to 0. \end{aligned} \tag{5}$$

Inserting Eq. (4), we get

$$\pi_0(\mathscr{G}_0) = Z_N, \quad \pi_1(\mathscr{G}_0) = 0. \tag{6}$$

Now

$$\begin{aligned} &\pi_0(\mathscr{G}_0) = \pi_0 \mathrm{SU}(N)^M = [M,\mathrm{SU}(N)],\\ &\pi_1(\mathscr{G}_0) = [S^1 \wedge M, \mathrm{SU}(N)], \end{aligned} \tag{7}$$

where the symbol $\wedge$ means smash product. In order to compute the equivalence classes in Eqs. (7), we may write down a Postnikov system for $X = G = \mathrm{SU}(N)$ as this group is 2-connected:

$$\begin{aligned} &K(\pi_5,5) \to G_5\\ &\qquad\qquad\quad\ \downarrow\\ &K(\pi_4,4) \to G_4 \to K(\pi_5,6)\\ &\qquad\qquad\quad\ \downarrow\\ &K(\pi_3,3) = G_3 \to K(\pi_4,5). \end{aligned} \tag{8}$$

We recall that $\pi_i = \pi_i\,(\mathrm{SU}(N))$. From the fibrations of this system we can extract the following interesting exact sequences of groups[11]:

$$\to H^2(M,\pi_3) \to H^4(M,\pi_4) \to [M,G_4] \to H^3(M,\pi_3) \to H^5(M,\pi_4), \tag{9a}$$

$$\to H^1(M,\pi_3) \to H^3(M,\pi_4) \to [S^1 \wedge M, G_4] \to H^2(M,\pi_3) \to H^4(M,\pi_4), \tag{9b}$$

$$\to H^4(M,\pi_5) \to [S^1 \wedge M, G_5] \to [S^1 \wedge M, G_4] \to H^5(M,\pi_5). \tag{9c}$$

We recall that since $\dim M \leqslant 4$, $H^5(M,\pi_4) = H^5(M,\pi_5) = 0$,[4] $[M,\mathrm{SU}(N)] \approx [M,G_4]$, and $[S^1 \wedge M, \mathrm{SU}(N)] \approx [S^1 \wedge M, G_5]$.

Now let us first consider the case $N \geqslant 3$. Then $H^4(M,\pi_4) = 0$, and from (9a) we get the exact sequence

$$0 \to [M,G_4] \to H^3(M,Z) \to 0, \tag{10}$$

so that, in this case,

$$\pi_0(\mathscr{G}_0) \approx H^3(M,Z) \approx [M,G_4]. \tag{11}$$

If	$M = S^4$,	then	$\pi_0(\mathscr{G}_0) \approx 0.$
If	$M = S^3$,	then	$\pi_0(\mathscr{G}_0) \approx Z.$
If	$M = T^4$,	then	$\pi_0(\mathscr{G}_0) \approx 4Z.$
If	$M = T^3$,	then	$\pi_0(\mathscr{G}_0) \approx Z.$
If	$M = RP^4$,	then	$\pi_0(\mathscr{G}_0) \approx 0.$
If	$M = RP^3$,	then	$\pi_0(\mathscr{G}_0) \approx Z.$
If	$M = S^2 \times S^2$,	then	$\pi_0(\mathscr{G}_0) \approx 0.$
If	$M = S^3 \times S^1$,	then	$\pi_0(\mathscr{G}_0) \approx Z.$
If	$M = S^2 \times S^1$,	then	$\pi_0(\mathscr{G}_0) \approx Z.$

For a two-dimensional manifold $\pi_0(\mathscr{G}_0) \approx 0$. In any case we find a contradiction with Eq. (6). For completeness, we report that, taking into account Eqs. (9b) and (9c), one finds new contradictions when $M = T^4, T^3, S^2 \times S^2, S^2 \times S^1, RP^4, RP^3$.

Now we concentrate on the particular case $N = 2$. Then $\pi_4 \approx Z_2$.

We may take advantage of the exact sequence

$$\to H^4(M,Z_2) \to [M,G_4] \to H^3(M,Z) \to 0. \tag{12}$$

One immediately sees that for all two-three-four-dimensional manifolds already mentioned $[M,G_4] \not\approx Z_2$. The only exceptions may be S^4, RP^4, $S^2 \times S^2$ for which Eq. (12) is not conclusive. For $S^2 \times S^2$ and RP^4 [12] we use the sequences (9b) and (9c), from which we get

$$0 \to [S^1 \wedge M, G_4] \to Z \oplus Z \to Z_2 \quad \text{for } M = S^2 \times S^2, \tag{13a}$$

$$0 \to Z_2 \to [S^1 \wedge M, G_4] \to Z_2 \to Z_2 \quad \text{for } M = RP^4, \tag{13b}$$

$$\to [S^1 \wedge M, G_5] \to [S^1 \wedge M, G_4] \to 0 \quad \text{for both.} \tag{13c}$$

From (13a) and (13b) we see that in both cases $[S^1 \wedge M, G_4]$ is not trivial, so that from (13c) $[S^1 \wedge M, G_5] \neq 0$, then, by Eq. (7) $\pi_1(\mathscr{G}_0) \neq 0$. So we are again in contradiction.

So sum up, when $G = \mathrm{SU}(N)$, we have verified that for all manifolds considered a continuous gauge choice is impossible. When the gauge group is $G' = \mathrm{SU}(N)/Z_N$ we need only a minor specification. Indeed it is enough to observe that $\mathscr{G}'$ coincides with(it is homeomorphic to) $\tilde{\mathscr{G}} = \mathscr{G}/Z_N$, where $\mathscr{G}$ is the group of $\mathrm{SU}(N)$-valued gauge transformations. Then the previous discussion needs no change, and at the end we arrive at the same conclusion.

Now let us consider a gauge theory with a compact semisimple, connected, but not necessarily simply connected, group G. In this case the previously quoted Singer's theorem does not hold, because classifying inequivalent G-bundles requires more characteristic classes besides the second Chern class. But it still holds for $\bar{G}$, the universal covering group of G. So if $\bar{\mathscr{G}}$ is the group of $\bar{G}$-valued gauge transformations, we have the following fibration:

$$0 \to \bar{\mathscr{G}}_0 \to \bar{\mathscr{G}} \to \bar{G} \to 0, \tag{14}$$

where $\bar{\mathscr{G}}_0$ is, as above, the subgroup of $\bar{\mathscr{G}}$ containing only the gauge transformations with a fixed point. So Singer's theorem and the fibration (14) mean

$$\pi_0(\bar{\mathscr{G}}) \approx [M,\bar{G}], \quad \pi_1(\bar{\mathscr{G}}) \approx [S^1 \wedge M,\bar{G}] \tag{15}$$

if $\dim M \leqslant 4$. Now we must link $\bar{\mathscr{G}}$ with the very gauge group of the theory, that is, $\mathscr{G}$, the group of G-valued gauge transformations. We have the fibration

$$0 \to \pi_1(G) \to \bar{G} \to G \to 0.$$

As $\pi_1(G)$ belongs to the center of $\bar{G}$ (which is finite), we get a new fibration

$$0 \to \pi_1(G) \to \bar{\mathscr{G}} \to \tilde{\mathscr{G}} \to 0. \tag{16}$$

Let C be the center of G, which is also finite. Then

$$0 \to C \to \mathscr{G} \to \tilde{\mathscr{G}} \to 0, \tag{17}$$

where $\tilde{\mathscr{G}} = \mathscr{G}/C$, is a fibration. On the basis of the previous discussion, a gauge choice is possible only if Eqs. (4) are satisfied. Consequently, from (15), (16), and (17) we must have

$$\to 0 \to [S^1 \wedge M,\bar{G}] \to 0 \to \pi_1(G) \to [M,\bar{G}] \to C. \tag{18}$$

Let us consider some examples. Let $G = \mathrm{SO}(n)$, $n \geqslant 6$. Then $\pi_1(\mathrm{SO}(n)) = Z_2$. Let us write down a Postnikov system for $\bar{G}$. Then we get, for example, an exact sequence like (9a), where π_i are to be interpreted as homotopy groups of $\bar{G}$, $\pi_i(\bar{G}) = \pi_i(G)$ for $i \geqslant 2$. We get $[M,\bar{G}_4] \approx H^3(M,Z)$ for $\dim M \leqslant 4$, as $\pi_4(G) = 0$. For all manifolds considered, $H^3(M,Z)$ is either trivial or a finite direct sum of Z groups. Therefore, we get immediately a contradiction with the sequence (18). The same result can be obtained for SO(5). For $G = \mathrm{SO}(4)$, as well as for other semisimple groups commonly used in gauge theories (i.e., products of simple groups) the extension is straightforward.

For the sake of completeness we quote also the exceptional groups and the symplectic groups. The first case is fairly simple due to the fact that the fourth and fifth homotopy groups are trivial, the sixth homotopy group is different from zero ($\approx Z_3$) only for G_2,[13] and the previous proof extends easily. The sympletic groups have the same first five homotopy groups as SU(2), so that one has the same result as in that case.

APPENDIX

All fibrations involved in a Postnikov system are induced by a principal fibration having a space of paths as total space and a space of loops as fiber. Therefore, any such fibration is also a principal fibration.[2] This qualification implies much more than a simple (Serre) fibration. In fact a Serre fibration (E, p, B, F) is a principal fibration if it has the following (briefly stated)properties[2,14]:

(1) There is a multiplicationlike binary operation m: $F \times F \to F$, which is homotopy associative and has a two-sided unit and a two-sided homotopy inverse;

(2) there is a map $\mu: F \times E \to E$, which acts fibrewise and coincides with m when restricted to $F \times F$;

(3) if we call $E^* \subset E \times E$ the subset $\{(z_1,z_2): p(z_1) = p(z_2)\}$, there exists a map h: $E^* \to F$ such that $\mu \circ (h,\pi_1) \simeq \pi_2$, where $\pi_i(z_1,z_2) = z_i$.

Now let M be any space. Then for a principal fibration the following theorem holds:

Let $v, v' \in [M,E]$, then $p^*(v) = p^*(v')$ if and only if there exists $w \in [M,F]$ such that $\mu^*(w,v) = v'$. It turns out that $w = h^*(v,v')$. Here p^*, μ^*, h^* are of course induced by p, μ, and h, respectively.

Now we split $[M,E]$ into subsets, each of which corresponds to an element of $[M,B]$. This is possible because p^* is onto. Let us consider one such subset $\{v: p^*(v) = u \in [M,B]\}$, and pick out from it an arbitrary element v_0. Let us put $\phi^*(w) = \mu^*(w,v_0)$ and $\psi^*(v) = h^*(v_0,v)$. Then, due to the above-stated theorem, ϕ^* maps $[M,B]$ onto the subset $p^{*-1}(u)$ and ϕ^* and ψ^* turn out to be inverse to each other. Therefore, ϕ^* is a one-to-one map of $[M,F]$ onto $p^{*-1}(u)$, and in general every subset of $[M,E]$ corresponding to a fixed element of $[M,B]$ is in one-to-one correspondence with $[M,F]$.

The set $[M,F]$ is endowed with a group structure, due to the properties of the map m. One can obviously transfer this group structure into every subset of $[M,E]$ corresponding to a fixed element of $[M,B]$. However, there is no natural way of doing that.

[1] G.'t Hooft, Nucl. Phys. B **153**, 141 (1979); G. Mack, Preprint MPI-PAE/PTh 41/79 (1979).

[2] R. E. Mosher and M. C. Tangora, *Cohomology Operations and Applications in Homotopy Theory* (McGraw-Hill, New York, 1966).

[3] E. H. Spanier, *Algebraic Topology* (McGraw-Hill, New York, 1966).

[4] S. J. Avis and C. J. Isham, lectures at the Cargése Summer School on Recent Advances in Gravitation (1978).

[5] We take $Y^X [X,Y]$, to mean, respectively, the set of maps and the homotopy classes of maps from the pointed space X to the pointed space Y.

[6] C. R. F. Maunder, *Algebraic Topology* [Van Nostrand Reinhold, London, 1970); A. Dold, *Lectures in Algebraic Topology* [Springer, New York, 1971).

[7] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).

[8] I. M. Singer, Comm. Math. Phys. **60**, 7 (1978).

[9] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963).

[10] Any element of $\mathscr{G}$ can be viewed as a family of G-valued functions, each of which is defined in a trivializing neighborhood in M, subjected to suitable compatibility relations.

[11] $[M,G_4]$, $[S^1 \wedge M,G_5]$, etc., in the sequences (9), can be given a group structure, due to the 2-connectedness of SU(N). As a matter of fact, for example $[M,G_4]$ is an abelian group.[2]

[12] We give up considering S^4 here because our method is quite plethoric in this case, though, of course, it leads to the same results as Singer's.

[13] A. Borel, Bull. Am. Math. Soc. **50**, 397 (1955); R. Bott and H. Samelson, Am. J. Math. **80**, 964 (1959).

[14] F. P. Peterson and N. Stein, Am. J. Math. **81**, 281 (1959).

The linear integral equations associated with any antiexact solution of the Yang–Mills free field equations

Dominic G. B. Edelen
Center for the Application of Mathematics, Lehigh University, Bethlehem, Pennsylvania 18015

(Received 29 April 1980; accepted for publication 15 August 1980)

Any system of r antiexact 1-forms, A^{α}, that solves the Yang–Mills free field equations associated with a semisimple r-parameter gauge group is shown to satisfy a system of linear inhomogeneous integral equations. These equations take the form $A^{\alpha} = A^{\alpha}_i\, dx^i$, $A^{\alpha}_i = R^{\alpha}_i + Y^{\alpha j}_{i\rho}(A^{\rho}_j)$, where $R^{\alpha}_i\, dx^i$ are r antiexact 1-forms, each of which solves Maxwell's free field equations, and $Y^{\alpha j}_{i\rho}$ are linear Riemann–Graves integral operators that are also linear in the structure constants of the gauge group. The standard iteration procedure for solving such systems may be looked upon as generating a sequence of corrections to the R^{α}_i fields that account for the nonabelian character of the gauge group.

PACS numbers: 11.10.Np, 02.30.Rz

1. MATHEMATICAL PRELIMINARIES

Most of the results obtained in this paper come about in a simple and direct manner by use of the calculus of exterior differential forms together with certain extensions that are given in this section. We assume that the reader is familiar with the standard structure of the exterior calculus so that we may use the following symbols without further remarks:

$\wedge$ for the exterior product of forms,
$\lrcorner$ for the inner multiplication of a vector with a form,
d for the exterior derivative of a form.

The arena for our discussion is a four-dimensional flat manifold M_4, that is the Cartesian product of three-dimensional Eculidean space with the real line. It is sufficient for our purposes to work with a fixed global coordinate cover $\{x^i\}$ relative to which the metric tensor of M_4 is given by $((g_{ij})) = \text{diag}(1,1,1,-1)$ and hence $\det(g_{ij}) = -1$. Results for other coordinate covers may be obtained directly from the standard mapping properties of vector fields and exterior differential forms.

Vector fields on M_4 are written as

$$V = v^i\, \partial_i\,,$$

with the usual summation convention and $\partial_i \equiv \partial/\partial x^i$. The class of exterior forms of degree k on M_4 is denoted by Λ^k. We use

$$\mu = dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 = \frac{1}{4!} e_{ijkl}\, dx^i \wedge dx^j \wedge dx^k \wedge dx^l\,, \tag{1.1}$$

for the volume 4-form of M_4. Starting with μ, construct a sequence of forms of decreasing degree by

$$\mu_i = \partial_i \lrcorner \mu\,, \quad \mu_{ji} = \partial_j \lrcorner \mu_i = \partial_j \lrcorner \partial_i \lrcorner \mu = -\mu_{ij}\,, \quad \mu_{kji} = \partial_k \lrcorner \mu_{ji}\,, \tag{1.2}$$

so that $\{\mu_i\}$ is a basis for Λ^3, $\{\mu_{ji} \mid j < i\}$ is a basis for Λ^2, etc. These forms exhibit the following useful properties (Ref. 1, pp. 480–486):

$$d\mu_i = 0, \quad d\mu_{ji} = 0, \quad d\mu_{kji} = 0, \qquad dx^i \wedge \mu_j = \delta^i_j\, \mu, \quad dx^i \wedge \mu_{jk} = \delta^i_j\, \mu_k - \delta^i_k\, \mu_j\,, \tag{1.3}$$

and hence

$$P = p^i \mu_i, \quad Q = \tfrac{1}{2} q^{ij}\, \mu_{ij}, \quad q^{ij} = -q^{ji}, \tag{1.4}$$

yield

$$dP = (\partial_i\, p^i)\, \mu, \quad dQ = (\partial_i\, q^{ij})\, \mu_j\,, \qquad V \lrcorner P = v^j\, p^i\, \mu_{ji}, \quad V \lrcorner Q = \tfrac{1}{2} v^k q^{ij} \mu_{kij}\,. \tag{1.5}$$

Since M_4 is star shaped with respect to any of its points, it is simplest to take the center of M_4 to be the origin, in which case we define the vector field X by

$$X = x^i \partial_i\,. \tag{1.6}$$

If W is a form of degree k on M_4, then W can be written uniquely as

$$W = W_{i_1 \cdots i_k}(x^l)\, dx^{i_1} \wedge dx^{i_2} \wedge \cdots \wedge dx^{i_k}\,,$$

in which case we define the λ-dependent 1-parameter family of k-forms $\widetilde{W}(\lambda)$ by

$$\widetilde{W}(\lambda) = W_{i_1 \cdots i_k}(\lambda x^l)\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}\,. \tag{1.7}$$

The linear homotopy operator, H, is defined on k-forms W by means of

$$HW = \int_0^1 X \lrcorner\, \widetilde{W}(\lambda)\, \lambda^{k-1}\, d\lambda\,, \tag{1.8}$$

and satisfies the fundamental identity

$$W = dHW + H\, dW, \tag{1.9}$$

whereby the Poincaré lemma is established for M_4 (Ref. 1, pp. 414–502; Ref. 2, pp. 211–5).

The operation H gives rise to the collection of "antiexact" exterior forms which have very useful properties. The reader is referred to Ref. 1, Chap. V of the Appendix for the derivations as well as an account of what happens under mappings and under a change of center. Since

$$HHW \equiv 0, \tag{1.10}$$

(1.9) shows that any form W has an *exact part*

$$W_e = d\,(HW + d\eta) = d\xi, \tag{1.11}$$

and ξ is unique under the stipulation $\xi \in \ker H$ (i.e., $HW + d\eta \in \ker H \rightleftarrows d\eta = 0$). Any form W also has a unique *antiexact part*

 0022-2488/81/102272-04$01.00

$$W_a = H\,dW, \tag{1.12}$$

and $H\,dW \in \ker H$. The class of *antiexact forms*, $\mathscr{A}$ is defined by

$$\mathscr{A} = \{ W \in \Lambda \mid W \in \ker H \}\,. \tag{1.13}$$

This is equivalent to

$$\mathscr{A} = \{ W \in \Lambda \mid X \lrcorner W = 0, \quad W(0) = 0 \quad \text{for} \quad \deg(W) > 0 \}, \tag{1.14}$$

from which it follows that *$\mathscr{A}$ is a submodule of the module Λ of exterior forms on M_4*. Further, and most important of all, *H is the inverse of d for any $W \in \mathscr{A}$*:

$$W = H\,dW, \quad \forall W \in \mathscr{A}\,. \tag{1.15}$$

For example, suppose B is a given element of $\mathscr{A}$ and we want to obtain all $W \in \mathscr{A}$ such that

$$dW = G + B \wedge W.$$

Since $W \in \mathscr{A}$, $B \in \mathscr{A}$ imply $B \wedge W \in \mathscr{A}$ and $W = H\,dW$, application of H to both sides of the above equation yield

$$W = H\,dW = HG + H(B \wedge W) = HG,$$

and we are done.

2. STATEMENT OF THE PROBLEM

We shall consider classical gauge fields associated with an r-parameter Lie group G. These fields are represented by the r Yang–Mills 1-form potentials

$$A^{\alpha} = A_i^{\alpha}(x)\,dx^i, \quad \alpha = 1,\ldots,r\,. \tag{2.1}$$

Greek indices have the range 1 through r and the summation convention is also observed with respect to these indices. Let $u^\alpha = u^\alpha(x)$ and $C^{\alpha}_{\beta\gamma}$, respectively, denote the parameters and the constants of structure of the group G in an appropriate representation. The action of G on the 1-form potentials gives rise to the gauge transformations

$$\grave{A}^{\alpha} = A^{\alpha} - du^{\alpha} - C^{\alpha}_{\rho\beta}\,A^{\rho}u^{\beta}\,, \tag{2.2}$$

and to the connection 1-forms

$$\omega^{\alpha}_{\beta} = C_{\rho}{}^{\alpha}{}_{\beta}\,A^{\rho}\,, \tag{2.3}$$

that characterize the nonabelian aspects of G. The curvature 2-forms Ω^{α}_{β} associated with ω^{α}_{β} assume the form

$$\Omega^{\alpha}_{\beta} = C_{\rho}{}^{\alpha}{}_{\beta}\,F^{\rho}, \tag{2.4}$$

where the 2-forms

$$F^{\alpha} = dA^{\alpha} + \omega^{\alpha}_{\beta} \wedge A^{\beta}/2\,, \tag{2.5}$$

are the "field strengths" (i.e., $F^\alpha = \frac{1}{2} F^{\alpha}_{ij}\,dx^i \wedge dx^j$). They satisfy the equations

$$dF^{\alpha} + \omega^{\alpha}_{\beta} \wedge F^{\beta} = 0, \tag{2.6}$$

for any and every choice of the A^{α}'s as a consequence of the Bianchi identities of Ω^{α}_{β} and (2.4). Although the F^{α}'s given by (2.5) are not invariant under the action of G, the equations (2.6) are invariant. [Equations (2.5) and (2.6) are the nonabelian gauge generalization of the first half of Maxwell's equations

$$dF = 0, \quad F = dA,$$

with $F = \frac{1}{2} F_{ij}\,dx^i \wedge dx^j$ and $A = A_i\,dx^i$].

The generalization of the second half of Maxwell's equations to the nonabelian gauge theory obtains directly from the demand that a *gauge invariant* action functional with Lagrangian L be rendered stationary relative to the choice of the A^{α}'s:

$$\partial_i \left(\frac{\partial L}{\partial(\partial_i A^{\alpha}_j)} \right) - \frac{\partial L}{\partial A^{\alpha}_j} = 0\,. \tag{2.7}$$

In the interest of simplicity, we restrict our attention to what may be called the free Yang–Mills situation in which the Lagrangian is taken to be a gauge invariant function of the components F^{α}_{ij} of $F^{\alpha} = \frac{1}{2} F^{\alpha}_{ij}\,dx^i \wedge dx^j$. If we set

$$G_{\alpha}{}^{ij} = \partial L/\partial F^{\alpha}_{ij}, \quad G_{\alpha} = \tfrac{1}{2} G^{ij}_{\alpha}\,\mu_{ij}\,, \tag{2.8}$$

Then the use of (2.5) and (2.3) show that (2.7) is equivalent to

$$dG_{\alpha} - \omega^{\beta}_{\alpha} \wedge G_{\beta} = 0\,. \tag{2.9}$$

[Note that $G_\alpha = \frac{1}{2} G^{ij}_{\alpha}\,\mu_{ij}$ and (1.4) imply that

$$dG_{\alpha} = (\partial_i G^{ij}_{\alpha})\,\mu_j\,,$$

and that $\omega^{\alpha}_{\beta} = \omega^{\alpha}_{\beta i}\,dx^i$ and (1.3) give

$$\omega^{\beta}_{\alpha} \wedge G_{\beta} = (\omega^{\beta}_{\alpha i}\,G^{ij}_{\beta})\,\mu_j\,.$$

Thus (2.8) and (2.9) generalize the second half,

$$G^{ij} = \partial L/\partial F_{ij}, \quad \partial_i G^{ij} = 0,$$

of Maxwell's equations.]

If the group G is semisimple then the Cartan–Killing metric

$$C_{\alpha\beta} = C_{\alpha}{}^{\lambda}{}_{\nu}\,C_{\beta}{}^{\nu}{}_{\lambda} = C_{\beta\alpha} \tag{2.10}$$

is nonsingular, while the Jacobi identity and (2.3) show that

$$-\,\omega^{\rho}_{\alpha} C_{\rho\beta} = \omega^{\rho}_{\beta} C_{\alpha\rho}\,. \tag{2.11}$$

Thus, if we introduce the 2-forms G^{α} by

$$G_{\alpha} = C_{\alpha\gamma} G^{\gamma}\,, \tag{2.12}$$

it follows that

$$dG_{\alpha} - \omega^{\beta}_{\alpha} \wedge G_{\beta} = C_{\alpha\nu}(dG^{\nu} + \omega^{\nu}_{\gamma} \wedge G^{\gamma})\,.$$

In this instance (2.9) is equivalent to the field equations

$$dG^{\alpha} + \omega^{\alpha}_{\beta} \wedge G^{\beta} = 0\,. \tag{2.13}$$

The customary restriction to a quadratic Lagrangian then reduces (2.8) and (2.12) to the equivalent statement

$$F^{\alpha}_{kl} = L_{klij} G^{\alpha ij} \tag{2.14}$$

for some L_{ijkl} that is invariant under the action of G. For example $C_{\alpha\beta}\,F^{\alpha}_{ij}\,F^{\beta}_{kl}\,g^{ik}\,g^{jl}$ is manifestly invariant under G in which case we would have

$$G^{\alpha ij} g_{ik}\,g_{jl} = F^{\alpha}_{kl}\,. \tag{2.15}$$

The reader is referred to the survey article of Yang[3] for the details of these matters, although in a slightly different notation.

Our basic problem is now evident.

Find all Yang–Mills 1-form potentials A^{α} for a semisimple Lie group G, that satisfy the field equations

$$dG^{\alpha} + \omega^{\alpha}_{\beta} \wedge G^{\beta} = 0\,,$$

where

$$F^{\alpha} = dA^{\alpha} + \omega^{\alpha}_{\beta} \wedge A^{\beta}/2, \quad \omega^{\alpha}_{\beta} = C_{\rho}{}^{\alpha}{}_{\beta}\,A^{\rho}\,, \tag{2.16}$$

and

$$L_{klij} G^{\alpha ij} = F^{\alpha}_{kl} . \tag{2.17}$$

For the present, we will not restrict the L 's except to require that they are unaffected by the action of G.

3. THE EQUIVALENT SYSTEM OF LINEAR INTEGRAL EQUATIONS

If the A^{α}'s are restricted to be antiexact [i.e., $X \rfloor A^{\alpha} = 0, A^{\alpha}(0) = 0$], there is a system of *linear* integral equations that are equivalent to the Yang–Mills free field equations (2.15)–(2.17). In order to see this, we first write (2.16) in the equivalent form

$$dA^{\alpha}_{a} = F^{\alpha} - \omega^{\alpha}_{\beta} \wedge A^{\beta}_{a} = F^{\alpha} - \tfrac{1}{2} C^{\alpha}_{\gamma\beta} A^{\gamma}_{a} \wedge A^{\beta}_{a} , \tag{3.1}$$

where we have put $A^{\alpha} = A^{\alpha}_{a}$ to emphasize the fact that the A^{α}'s are now assumed to be antiexact. Since antiexact forms form a module, $C^{\alpha}_{\gamma\beta} A^{\gamma}_{a} \wedge A^{\beta}_{a}$ is antiexact and hence belongs to ker$H$. Thus applying H to both sides of (3.1) and noting that $A^{\alpha}_{a} = H\, dA^{\alpha}_{a}$ for any antiexact form, we obtain

$$A^{\alpha}_{a} = HF^{\alpha} . \tag{3.2}$$

Written out, with the aid of (1.6)–(1.8), (3.2) gives

$$A^{\alpha}_{aj} = x^{i} \int_{0}^{1} F^{\alpha}_{ij}(\lambda x)\, \lambda \, d\lambda .$$

Thus, introducting the linear operators

$$h_{k}(f)(x) = \int_{0}^{1} f(\lambda x)\, \lambda^{k}\, d\lambda , \tag{3.3}$$

(3.2) may be written as

$$A^{\alpha}_{aj} = x^{i} h_{1}(F^{\alpha}_{ij}) . \tag{3.4}$$

It is now a simple matter to use (2.17) in order to obtain

$$A^{\alpha}_{aj} = x^{i} h_{1}(L_{ijkl} G^{\alpha kl}) . \tag{3.5}$$

We thus need to solve (2.15) for $G^{\alpha kl}$.

Let us start by decomposing G^{α} by means of the operator H

$$G^{\alpha} = d\xi^{\alpha} + G^{\alpha}_{a}, \quad \xi^{\alpha} = HG^{\alpha} . \tag{3.6}$$

When (3.6) is substituted into (2.15), we see that

$$dG^{\alpha}_{a} = -\omega^{\alpha}_{\beta} \wedge (d\xi^{\beta} + G^{\beta}_{a}),$$

and hence (2.16) gives

$$dG^{\alpha}_{a} = -C^{\alpha}_{\rho\beta} A^{\rho}_{a} \wedge (d\xi^{\beta} + G^{\beta}_{a}) . \tag{3.7}$$

However, $G^{\alpha}_{a} = H\, dG^{\alpha}_{a}$ since G^{α}_{a} is antiexact and $C^{\alpha}_{\rho\beta} A^{\rho}_{a} \wedge G^{\beta}_{a}$ belongs to ker$H$ by the module property of $\mathscr{A}$. Thus (3.7) yields

$$G^{\alpha}_{a} = -C^{\alpha}_{\rho\beta} H(A^{\rho}_{a} \wedge d\xi^{\beta}), \tag{3.8}$$

and hence (3.6) gives

$$G^{\alpha} = d\xi^{\alpha} - C^{\alpha}_{\rho\beta} H(A^{\alpha}_{a} \wedge d\xi^{\beta}) . \tag{3.9}$$

Thus, for any choice of the 1-forms ξ^{α}, (3.9) solves the field equations (2.15) for any A^{ρ}_{a}. Now

$$G^{\alpha} = \tfrac{1}{2} G^{\alpha ij} \mu_{ij}$$

and hence we need to write ξ^{α} in the form

$$\xi^{\alpha} = \frac{1}{3!} \xi^{\alpha ijk} \mu_{ijk} , \quad \xi^{\alpha(ijk)} = 0 ,$$

in which case

$$d\xi^{\alpha} = \tfrac{1}{2} \partial_{m} \xi^{\alpha mij} \mu_{ij} .$$

Further,

$$A^{\rho}_{a} \wedge d\xi^{\beta} = A^{\rho}_{ar} \partial_{m} \xi^{\alpha mrj} \mu_{j} ,$$

and hence

$$X \rfloor \tilde{A}^{\rho}_{a} \wedge d\tilde{\xi}^{\beta} = x^{i} \tilde{A}^{\rho}_{ar} \partial_{m} \tilde{\xi}^{\alpha mrj} \mu_{ij} .$$

Thus, if we put

$$\partial_{m} \xi^{\alpha mij} = B^{\alpha ij} ,$$

then (3.9) yields

$$G^{\alpha ij} = B^{\alpha ij} - C^{\alpha}_{\rho\beta} [x^{i} h_{2}(A^{\rho}_{ar} B^{\beta rj}) - x^{j} h_{2}(A^{\rho}_{ar} B^{\beta ri})] , \tag{3.10}$$

where

$$B^{\alpha ij} = \partial_{m} \xi^{\alpha mij}, \quad \xi^{\alpha(mij)} = 0 , \tag{3.11}$$

may be thought of as generating functions. It now remains only to substitute (3.10) into (3.5) in order to obtain the following linear system of Riemann–Graves integral equations for the A^{α}'s:

$$A^{\alpha}_{aj} = x^{i} h_{1}[L_{ijkl} B^{\alpha kl}] - C^{\alpha}_{\rho\beta} x^{i} h_{1} \times [L_{ijkl} x^{k} h_{2}(A^{\rho}_{ar} B^{\beta rl}) - L_{ijkl} x^{l} h_{2}(A^{\rho}_{ar} B^{\beta rk})] .$$

However, the antisymmetry of L_{ijkl} in the indices k, l allows reduction to the equivalent form

$$A^{\alpha}_{aj} = x^{i} h_{1}[L_{ijkl} B^{\alpha kl}] - 2C^{\alpha}_{\rho\beta} x^{i} h_{1}[L_{ijkl} x^{k} h_{2}(A^{\rho}_{ar} B^{\beta rl})] . \tag{3.12}$$

It is clear from the form of (3.12) that those $B^{\alpha kl}$ for which solutions exist to (3.12) serve as a system of parametrizing functions for the solutions.

When solutions exist they may be obtained by iteration of (3.12) starting with

$$A^{\alpha}_{aj(0)} = x^{i} h_{1}[L_{ijkl} B^{\alpha kl}] .$$

We draw particular note to the form that (3.12) takes when the L 's are constants

$$A^{\alpha}_{aj} = L_{ijkl} \{ x^{i} h_{1}(B^{\alpha kl}) - 2C^{\alpha}_{\rho\beta} x^{i} h_{1}[x^{k} h_{2}(A^{\rho}_{ar} B^{\beta rl})] \} . \tag{3.13}$$

The observation,

$$h_{1}(x^{k} f) = x^{k} h_{2}(f) ,$$

shows that (3.13) may also be written as

$$A^{\alpha}_{aj} = L_{ijkl} x^{i} \{ h_{1}(B^{\alpha kl}) - 2C^{\alpha}_{\rho\beta} x^{k} h_{2}[h_{2}(A^{\rho}_{ar} B^{\beta rl})] \} . \tag{3.14}$$

Thus, if we set

$$R^{\alpha}_{j} = L_{ijkl} x^{i} h_{1}(B^{\alpha kl}) , \tag{3.15}$$

and define the linear operation $Y^{\alpha r}_{j\rho}$ by

$$Y^{\alpha r}_{j\rho}(\xi^{\rho}_{r}) = -2L_{ijkl} x^{i} x^{k} C^{\alpha}_{\rho\beta} h_{2}[h_{2}(\xi^{\rho}_{r} B^{\beta rl})] , \tag{3.16}$$

then the equivalent system of linear integral equations assume the relatively simple form

$$A^{\alpha}_{aj} = R^{\alpha}_{j} + Y^{\alpha r}_{j\rho}(A^{\rho}_{ar}) . \tag{3.17}$$

Equations (3.17) are the system of linear Riemann–Graves integral equations that are satisfied by any antiexact system of 1-form potentials that satisfy the free Yang–Mills field equations. Such solutions are, in a sense, intrinsic, for an analysis of the global action of the gauge group G and its naturally associated fiber space of connection 1-forms shows that any solution of the Yang–Mills field equations can be mapped onto an antiexact solution by the action of an appropriate element of G.[4]

4. DISCUSSION

The first thing to be noted is that we are free to pick the quantities $\xi^{\alpha ijk}$, and hence the $B^{\alpha ij}$ by (3.11). Thus, if we replace $\xi^{\alpha ijk}$ by $K\xi^{\alpha ijk}$, then (3.15) and (3.16) show that we obtain the replacements

$$R^{\alpha}_{j} \rightarrow KR^{\alpha}_{j}, \quad Y^{\alpha r}_{j\rho} \rightarrow KY^{\alpha r}_{j\rho}.$$

This replaces (3.17) by

$$A^{\alpha}_{aj} = KR^{\alpha}_{j} + KY^{\alpha r}_{j\rho}(A^{\rho}_{ar}). \tag{4.1}$$

Accordingly, if the functions $\xi^{\alpha ijk}$ are chosen with sufficiently nice properties (square integrable and of compact support) then we can always pick a value for K so that iteration of (4.1) will lead to a sequence $A^{\alpha}_{aj(n)}$ that converges to a solution as n tends to infinity. There is thus no lack of solutions of (3.17) for a fairly large class of choices of the "generating functions," $\xi^{\alpha ijk}$.

The iteration process obviously starts with the choice

$$A^{\alpha}_{aj(0)} = KR^{\alpha}_{j}. \tag{4.2}$$

However, (3.16) shows that $A^{\alpha}_{a(0)} = KR^{\alpha}_{j}$ is an exact solution when we set $C^{\alpha}_{\beta\gamma} = 0$; that is, the Abelian case, namely Maxwell's equations. Accordingly $A^{\alpha}_{aj(0)} = KR^{\alpha}_{j}$ are simply r solutions of Maxwell's equations subject to the gauge condition $X \rfloor A^{\alpha}_{(0)} = 0$, that is

$$x^{j} A^{\alpha}_{j(0)} = 0, \quad A^{\alpha}_{j(0)}(0) = 0, \tag{4.3}$$

so that the $A^{\alpha}_{(0)}$'s are antiexact. We may thus view the iteration process

$$\begin{aligned} A^{\alpha}_{aj(0)} &= KR^{\alpha}_{j}, \\ A^{\alpha}_{aj(n+1)} &= KR^{\alpha}_{j} + KY^{\alpha r}_{j\rho}(A^{\rho}_{ar(n)}), \end{aligned} \tag{4.4}$$

as a sequence of corrections of r solutions of Maxwell's equations that accounts for the presence of the nonabelian gauge group with structure constants $C_{\alpha\gamma}{}^{\beta}$. This view has a merit in that it gives rise to a heightened physical intuition of this complex subject, and possible new insights through a study of the correction operators $Y^{\alpha r}_{j\rho}$.

ACKNOWLEDGMENT

This work would never have been begun without the many discussions of gauge theory with Professor Hanno Rund, which are herewith gratefully acknowledged. The author also owes a debt of thanks to the Department of Mathematics of the University of Arizona for their kind hospitality for the Spring Semester, 1980, during which this paper was written.

[1] D. G. B. Edelen, *Isovector Methods for Equations of Balance with Programs for Computer Assistance in Operator Calculations and An Exposition of Practical Topics of the Exterior Calculus* (Noordhoff, Alphen ann den Rijn, The Netherlands, 1980).

[2] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds and Physics* (North-Holland, Amsterdam, 1977).

[3] C. N. Yang, "Gauge Fields", *Proceedings of the Sixth Hawaii Topical Conference in Particle Physics*, edited by P. N. Dobson, Jr., S. Pakvasa, V. Z. Peterson, and S. F. Tvan (University of Hawaii Press, Honolulu, 1976).

[4] D. G. B. Edelen, "On the Intrinsic Structure of Yang–Mills Fields" (to be published).

On invariant group integrals in lattice QCD

K. E. Eriksson and N. Svartholm
Institute of Theoretical Physics, S-41296 Göteborg, Sweden
B. S. Skagerstam [a)]
Centre Européen de Recherches Nucléaires, Geneva, Switzerland

(Received 10 February 1981; accepted for publication 17 April 1981)

A closed expression of the SU(3) [and U(3)] one-link invariant group integral in lattice gauge theories is derived. The U(3) result is compared to the expression recently derived by Brower, Rossi, and Tan. Applications of our results are briefly touched upon.

PACS numbers: 11.10.Np, 02.60.Nm

1. INTRODUCTION

In the study of lattice gauge theories[1,2] with a gauge group G, one encounters[3] invariant group integrals of the form

$$Z(m,m^+) = \int_G dg\, e^{\mathrm{Tr}(gm^+ + g^+ m)}, \tag{1}$$

where dg is the normalized, invariant Haar measure[4] on the group G. For convenience, we have absorbed coupling constant parameters in the matrix m. Equation (1) is the generating functional for computing expectation values of any given link for lattice gauge theories or for $(G \times G)$ chiral sigma models.[5] g in (1) then belongs to the fundamental representation of the group G. Equation (1) also enters into the construction of quasicoherent states appropriate for quantum fields carrying a nonbelian charge.[6,7] In this case, one is interested in other representations of G than the fundamental one. For pion or gauge fields, one is, in fact, interested in the adjoint representation.[6,7]

Several recent papers[5,8–13] have been devoted to the actual evaluation of (1) appropriate for lattice (QCD) gauge theories, i. e., one considers $G = \mathrm{U}(N)$ [or $\mathrm{SU}(N)$]. Two very successful methods have been applied. One is the character expansion[5] of the exponential in (1). The other one is the construction of a partial differential equation which (1) satisfies due to its invariance with respect to left- and right-handed actions of the group on m.[11] In Ref. 13, a solution of this one-link Schwinger–Dyson equation of Brower–Nauenberg for U(N) was actually found with the result

$$\begin{aligned} Z(m,m^+) &\equiv Z_0(mm^+) \\ &= \Big(2^{N(N-1)/2}\prod_{k=0}^{N-1}(k!)\Big)\frac{\det\big(z_j^{i-1}I_{i-1}(z_j)\big)}{\det\big((z_j^2)^{i-1}\big)}, \end{aligned} \tag{2}$$

where $z_k \equiv 2\sqrt{x_k}$ and x_k are the eigenvalues of the matrix mm^+.

In the $1/N$ expansion[14] approach to the QCD dynamics [$G = \mathrm{SU}(3)$] one argues that N being large is a "good" approximation to $N = 3$. Since U(N) and SU(N) seem to be rather similar in nature for large N, it is justified to study U(N) instead of SU(N). This circumstance simplifies the evaluation of (1) to a large extent since, for U(N), (1) is a function only of mm^+. In the present short paper, we will present an evaluation of $Z(m,m^+)$ for $G = \mathrm{U}(N), \mathrm{SU}(N)$, where $N = 2,3$. (The expressions we obtain for $N = 2$ seem to be known to some authors. We include our method of calculating these integrals as an illustration of the technique.) For the U(N) groups we will verify the expression (2), but we obtain $Z(m,m^+) \equiv Z_0(mm^+)$ in terms of an *explicit* power expansion in mm^+. For the SU(N) groups one has, in general, a dependence on more invariants of the matrices m and m^+. For $N = 3$ we will reveal a drastic change in the nature of the integral (1) when comparing the SU(3) [U(3)] results with the SU(2) [U(2)] expressions.

Concerning the motivation for the study of the invariant, one-link group integral (1), we have at present no more to add to the remarks in Refs. 3,5–13. The series expansion of the one-link integral (1) given below can be applied to a strong coupling expansion in a standard manner. Moreover, it is well known[15,16] that two-dimensional lattice gauge theories are exactly solvable in that the solution of the theory can be reduced to the evaluation of (1) for a diagonal matrix $m = \beta\mathbb{1}$. Now, for $N = \infty$, the U(N) theory exhibits a third-order phase transition.[16] For finite N, it should be possible to see a sign of this effect on a lattice and, moreover, one can compare SU(N) with U(N) in an exact manner. Such a study is presently in progress.[17]

2. $N = 2$ INTEGRALS

The SU(2) one-link integral can easily be evaluated by using an explicit S^3 parametrization since, topologically, $\mathrm{SU}(2) \approx S^3$. Writing the group element as $g = u_0\mathbb{1} + i\mathbf{u}\cdot\boldsymbol{\sigma}$, where $\boldsymbol{\sigma}$ are the three Pauli matrices, one obtains for (1),

$$\begin{aligned} Z(m,m^+) = \frac{1}{2\pi^3}\int ds \int d^4u\, &\exp(is(u^2 - 1)) \\ &\times \exp(u_0A + u_1B + u_2C + u_3D), \end{aligned} \tag{3}$$

where $u^2 = u_0^2 + \mathbf{u}^2$ and $A = 2\mathrm{Re}(m_{11} + m_{22})$, $B = 2\mathrm{Im}(m_{22} - m_{11})$, $C = 2\mathrm{Re}(m_{21} - m_{12})$, and $D = -2\mathrm{Im}(m_{12} + m_{21})$. The integration over the u variables in (3) are just Gaussian integrals which can be carried out directly. The final integral over the s parameter is then a Fourier–Mellin transform which can easily be evaluated. Introducing the SU(2) invariant combinations $\mathrm{Tr}(mm^+)$ and $\det m + \det m^+$, the final result is

$$Z(m,m^+) = \sum_{n=0}^{\infty}\frac{(\mathrm{Tr}(mm^+) + \det m + \det m^+)^n}{n!(n+1)!}, \tag{4}$$

where we recognize the power series expansion of the first

[a)] Institute of Theoretical Physics, S-41296 Göteborg, Sweden.

modified Bessel function I_1. If we treat m and m^+ as independent variables and put $m^+ = 0$, (4) reduces to the known result given by Creutz.[3] By making use of the expression (4), it is easy to derive an expression for the one-link U(2) generating functional $Z_0(mm^+)$ by a projection technique. Writing $g = \exp(i\varphi)g'$, where g' belongs to SU(2), we obtain

$$Z_0(mm^+) = \frac{1}{2\pi}\int_0^{2\pi} d\varphi\, Z(e^{i\varphi}m,\ e^{-i\varphi}m^+)$$
$$= \sum_{n,r=0}^{\infty} \frac{[\mathrm{Tr}(mm^+)]^n[\det(mm^+)]^r}{(n+2r+1)!n!(r!)^2}. \qquad (5)$$

We have verified explicitly that (5) actually agrees with the expression (2) for $N = 2$. It is amusing to notice that the SU(2) result (4) is much simpler in structure than the U(2) form (5). This is due to the simplicity of the SU(2) group, and for SU(3) the result, which will be given below, is quite complicated in terms of SU(3) invariants.

For the sake of completeness, we also mention here the generating functional (1) in the case of the adjoint representation of SU(2). m is then a 3×3 matrix and the g's can be represented by the rotation matrices $R_{ij} = \frac{1}{2}\mathrm{Tr}(\sigma_i g \sigma_j g^+)$, where $g = u_0\mathbb{1} + i\mathbf{u}\cdot\boldsymbol{\sigma}$. Since R_{ij} is real, only the real part of m contributes in (1). Writing $M = 2\mathrm{Re}(m)$ we obtain [see Refs. 6 and 7 for details, where an integral representation of (6) can be found, and where some physical applications are discussed]

$$Z(m,m^+) = Z(M)$$
$$= \sum_{j,k,l=0}^{\infty} \frac{[2(j+k+l)-1]!!}{j!k!l!(1+2j+3k+4l)!} x^j y^k z^l, \quad (6)$$

where

$$x = \mathrm{Tr}(MM^\sim),$$
$$y = 4\det M, \qquad (7)$$
$$z = \tfrac{1}{2}[\mathrm{Tr}(MM^\sim)]^2 - \mathrm{Tr}(MM^\sim MM^\sim).$$

Since $SU(2)/Z_2 \approx SO(3)$, the generating functional (6) for the SU(2) group in adjoint representation gives the SO(3) group generating functional.

3. N = 3 INTEGRALS

In evaluating the SU(3) one-link integral, we have found it very useful to parametrize[18] the SU(3) group manifold in terms of two normalized, complex three-vectors $\mathbf{u}$ and $\mathbf{v}$. Each g in SU(3) can then be written in the form

$$g = \begin{pmatrix} u_1^* & u_2^* & u_3^* \\ v_1^* & v_2^* & v_3^* \\ w_1 & w_2 & w_3 \end{pmatrix}, \qquad (8)$$

where $w_i = \epsilon_{ijk}u_j v_k$ and $\mathbf{u}^*\cdot\mathbf{v} = 0$, which leads to eight independent variables. An integration over SU(3) can be re-expressed in terms of the $\mathbf{u}$, $\mathbf{v}$ variables in (8). For the case $G = SU(3)$ in (1) we obtain

$$\int_{SU(3)} dg\, e^{\mathrm{Tr}(gm^+ + g^+m)} = \frac{1}{2\pi^2}\int_{-\infty}^{\infty} ds\, e^{is}\int_{-\infty}^{\infty} dt\, e^{it}$$
$$\times \int d^2z \int d^6u \int d^6v\, e^{-i(s\mathbf{u}^*\cdot\mathbf{u} + t\mathbf{v}^*\cdot\mathbf{v})} e^{iz(\mathbf{u}^*\cdot\mathbf{v} + \mathbf{v}^*\cdot\mathbf{u})}$$
$$\cdot e^{\mathrm{Tr}(gm^+ + g^+m)}. \qquad (9)$$

By making use of (8) we see that the argument in the exponential of (9) is a quadratic function in $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u}^*$, $\mathbf{v}^*$. The complex Gaussian $\mathbf{u}$ and $\mathbf{v}$ integrals can be performed in a straightforward manner. [We observe that we can always write $m = g_1 D g_2$, $g_1, g_2 \in SU(3)$, where D is a diagonal matrix whose matrix elements have a common phase.] All except one of the Fourier–Mellin transformations can then be explicitly performed. As a result we obtain

$$Z(m,m^+) = \frac{i}{\pi}\int_\gamma \frac{dp}{p}\frac{1}{\sqrt{T(p)}} J_1\Big(\frac{2}{p}\sqrt{T(p)}\Big)e^{-\Delta/p}, \qquad (10)$$

where

$$T(p) = p^3 - Xp^2 + Yp + Z \equiv -\det(mm^+ - p\mathbb{1}) \quad (11)$$

and X, Y, Z, Δ are the SU(3) invariant combinations

$$X = \mathrm{Tr}(mm^+),$$
$$Y = \tfrac{1}{2}[\mathrm{Tr}(mm^+)]^2 - \mathrm{Tr}((mm^+)^2), \qquad (12)$$
$$Z = \det(mm^+),$$
$$\Delta = \det m + \det m^+.$$

The contour γ in (10) encloses the pole $p = 0$. By expanding the Bessel function in (10) and performing the contour integral, we obtain

$$Z(m,m^+) = 2\sum_{j,k,l,n=0}^{\infty} \frac{1}{(j+2k+3l+n+2)!(k+2l+n+1)!} \times \frac{X^j}{j!}\frac{Y^k}{k!}\frac{Z^l}{l!}\frac{\Delta^n}{n!}. \qquad (13)$$

For $m^+ = 0$ this result agrees with that of Creutz.[3]

The U(3) one-link integral can now be obtained by making use of the method analogous to the one used in obtaining (5). The final U(3) integral can be written in the following forms

$$Z_0(mm^+) = \frac{i}{\pi}\int_\gamma \frac{dp}{p\sqrt{T(p)}} I_0\Big(2\frac{\sqrt{Z}}{p}\Big)J_1\Big(\frac{2}{p}\sqrt{T(p)}\Big)$$
$$= 2\sum_{j,k,l=0}^{\infty} \frac{(j+2k+4l+2)!}{[(j+2k+3l+2)!]^2} \times \frac{1}{l!(k+2l+1)!}\frac{X^j}{j!}\frac{Y^k}{k!}\frac{Z^l}{l!}. \qquad (14)$$

One can verify in a straightforward manner that (14) satisfies the Brower–Nauenberg equation, which, in terms of the invariants (12) reads for U(3),

$$[(X^2 - 2Y)\partial_X^2 + 2(Y^2 - XZ)\partial_Y^2 + 3Z^2\partial_Z^2 + 2(XY - 3Z)\partial_X\partial_Y + 2XZ\partial_X\partial_Z + 4YZ\partial_Y\partial_Z + 3X\partial_X + 4Y\partial_Y + 3Z\partial_Z]Z_0(mm^+) = XZ_0(mm^+). \qquad (15)$$

We have also shown explicitly the equivalence of (14) and (2). This is an expected result, since they satisfy the same differential equation and boundary condition.[13] In the same way, (13) satisfies the SU(3) Brower–Nauenberg equation. If we write[11] $\Delta \equiv 2\cos 3\theta\sqrt{Z}$, an additional term $(-1/12)[\partial^2 Z(m,m^+)/\partial\theta^2]$ will be present on the right-hand side of Eq. (15).

Finally, we remark that in the case of matrix $m = \beta\mathbb{1}$, one can evaluate (10) directly [for U(3) and SU(3)] instead of using (13) and (14). Equation (1) can then be expressed as a power expansion [see Ref. 17 for details and for an application in two-dimensional (lattice) QCD] in β.

4. CONCLUSIONS AND REMARKS

We have given a closed expression for the SU(3) [and U(3)] one-link integrals in terms of SU(3) [U(3)] invariants. Bars has given[5] a perspective on how to make use of the one-link integrals in lattice gauge theories. A strong coupling expansion can, for example, be done for SU(3) by making use of the generating functional (13). Corrections to the Bars–Greens truncation procedure[10] can furthermore be studied systematically for SU(3) [and U(3)]. Here we only notice that in the strong coupling limit, i. e., [β is small, $\beta \lesssim \frac{1}{4}(d-1)$ where d is the dimension of the lattice] in $m=\beta J$, we obtain $Z(m,m^+) \approx 3\beta^2 \mathrm{Tr} JJ^+$. As shown by Bars and Green,[10] one can then find an approximate solution of the lattice gauge theory. It would be interesting to apply the expression (13) or (14) to a study of corrections to the Bars–Greens approximation. Polyakov[19] has furthermore shown that the one-link integral (1) enters into the mean field approximation of lattice gauge theories. The one-link integral appropriate for this approximation in the large N limit has recently been discussed by Brezin and Gross,[20] where it was shown that (1) exhibits a two-phase structure as in the two-dimensional $U(\infty)$ theory. It has been conjectured[16] that for sufficiently large N one should see a sign of this two-phase structure. In two dimensions this is what actually happens.[17] The results of the present paper can be used to study this issue in higher (space–time) dimensions.

As a final remark we notice that the steepest descent method can be applied to the integral representation (10) [and to the integral representation for U(3) easily derivable from (10)] to study the weak coupling limit (β being large). As suggested by Brower and Nauenberg,[11] one could also study this limit by considering stationary points of (1) directly. We will not, however, develop these considerations in the present note.

Note added in the proof: The one-link $U(N)$ integral has also recently been considered by V. A. Faļeev and E. Onofri (TH. 2999-CERN preprint 1980). We thank the authors for information about their results prior to publication.

ACKNOWLEDGMENTS

The authors are grateful for many discussions with A. P. Balachandran and N. Mukunda on the subject of the present paper. B. S. also wishes to thank P. Hasenfratz, M. Nauenberg, R. Stora, and G. Veneziano for discussions and R. Brower and P. Rossi for discussions on the equation (2) and for informing us about the work in Ref. 13 prior to publication.

[1]K. Wilson, Phys. Rev. D **10**, 2445 (1974).
[2]For recent reviews on lattice gauge theory see J. B. Kogut, Rev. Mod. Phys. **51**, 659 (1979); J. M. Drouffe and C. Itzykson, Phys. Rep. C **38**, 133 (1978).
[3]For an explicit construction of the Feynman rules for lattice gauge theory in terms of group variables, see, e. g., M. Cruetz, Rev. Mod. Phys. **50**, 561 (1978).
[4]See, e. g., J. D. Talman, *Special Functions—A Group Theoretic Approach* (Benjamin, New York, 1968).
[5]See, e. g., I. Bars, J. Math. Phys. **21**, 2678 (1980).
[6]K. E. Eriksson, N. Mukunda, and B. S. Skagerstam, Phys. Rev. D (in press).
[7]K. E. Eriksson, N. Mukunda, and B. S. Skagerstam (in preparation). This paper gives a quasicoherent state representation of quantum fields for arbitrary compact groups.
[8]M. Cruetz, J. Math. Phys. **19**, 2043 (1978).
[9]D. Weingarten, J. Math. Phys. **19**, 999 (1978).
[10]I. Bars and F. Green, Phys. Rev. D **20**, 3311 (1979).
[11]R. Brower and M. Nauenberg, Nucl. Phys. B **180**, 221 (1981).
[12]S. Samuels, J. Math. Phys. **21**, 2695 (1980).
[13]R. Brower, P. Rossi, and C.-I Tan, Phys. Rev. D **23**, 942 and 953 (1981).
[14]See, e. g., S. Coleman, lectures presented at the 1979 International School of Subnuclear Physics, Erice, 1980 (SLAC-PUB-2484).
[15]A. M. Migdal, Zh. Eksp. Teor. Fiz. **69**, 810 (1975) [Sov. Phys. JETP **42**, 413 (1976)].
[16]D. J. Gross and E. Witten, Phys. Rev. D **21**, 446 (1980).
[17]C. B. Lang, P. Salomonson, and B. S. Skagerstam, Phys. Lett. B **100**, 29 (1981).
[18]M. A. B. Bég and H. Ruegg, J. Math. Phys. **6**, 677 (1965).
[19]A. M. Polyakov, Phys. Lett. B **82**, 247 (1979).
[20]E. Brezin and D. J. Gross, Phys. Lett. B **97**, 120 (1980).

Minkowski space Yang–Mills fields from solutions of equations in the three-dimensional Euclidean space

E. Kyriakopoulos
Physics Laboratory A, National Technical University, Athens, Greece

(Received 6 March 1981; accepted for publication 17 April 1981)

Equations in the three-dimensional Euclidean space are derived by combining the Yang–Mills field equations with the conditions which are imposed on a Yang–Mills field in Bernreuther's method of constructing Yang–Mills fields in Minkowski space from Yang–Mills fields in Euclidean space. By a proper ansatz for the Yang–Mills fields these equations are reduced to a single differential equation. The differential equation is identical with merons' equation in Euclidean space if we consider solutions of the latter equation which are functions of the ratio t/ρ, where t is the Euclidean time and ρ is the three-dimensional radius. One such solution is the single meron solution in Euclidean space. Starting from this and applying the method we get the de Alfaro–Fubini–Furlan solution in Minkowski space. Then, a more general ansatz is considered, which leads to a system of three nonlinear differential equations.

PACS numbers: 11.10.Np, 03.50.Kk, 03.30. + p

I. INTRODUCTION

In recent years a number of exact classical solutions of the Yang–Mills field equations in Minkowski as well as in Euclidean space were found.[1–18] The best known such solutions are the instanton solutions and the de Alfaro–Fubini–Furlan solution. Other solutions have also been found, for example the elliptic solutions, etc.

The standard method of finding classical solutions is the following. We make an ansatz for the Yang–Mills fields, we introduce this expression into the field equations, and then we try to solve the resulting system of second order nonlinear differential equations. The solution of this system is not easy in general. However, a large class of real solutions of the Yang–Mills field equations in Euclidean space can be found much more easily if we take into account the fact that a self-dual field is a solution of these equations. If we substitute the ansatz for the Yang–Mills fields into the duality condition a system of first order nonlinear differential equations is obtained. But, we cannot use this method to get real solutions in Minkowski space since a self-dual field in Minkowski space is necessarily complex.

We may therefore try to invent a method of getting real solutions in Minkowski space from solutions in Euclidean space. Indeed, such a method has been proposed by Bernreuther.[10] According to this method, which is reviewed in this section, from every Euclidean Yang–Mills field which satisfies certain conditions, we can get a real Yang–Mills field in Minkowski space. In Sec. II by combining these conditions with the field equations we obtain equations in the three-dimensional Euclidean space. In Sec. III by a proper ansatz for the Yang–Mills fields these equations are reduced to a single nonlinear differential equation. The resulting equation is identical with merons' differential equation, if we consider solutions which are functions of the ratio t/ρ (t is the Euclidean time and ρ the three-dimensional radius). A singular solution in Euclidean space, and from this by the application of the method, the de Alfaro–Fubini–Furlan solution in Minkowski space are obtained. Also we find that from every solution of that type of the merons' equation in Euclidean space a solution of the Yang–Mills equations in Minkowski space can be constructed. Finally in Sec. IV we consider a more general ansatz for the Yang–Mills fields. When we introduce this ansatz into the three dimensional equations we get a system of three nonlinear differential equations.

It should be pointed out that Bernreuther's method gives nontrivial results only if the Euclidean space solution we start from is not self-dual.[19]

To describe Bernreuther's method[10] we define from the Minkowski space variables $x_0, x_j, j = 1,2,3$ the Euclidean space variables y_α, $\alpha = 1,2,...,4$ by the relations

$$y_j = x_j, \qquad y_4 = \tfrac{1}{2}(1 + x_0^2 - \mathbf{x}^2), \tag{1.1}$$

and we consider the Euclidean space fields $A^j_\alpha(y)$. From these fields with the help of the Pauli matrices σ^j we define the matrices $A_\alpha(y) = (\epsilon/2i)\sigma^j A^j_\alpha(y)$, where ϵ is the gauge coupling constant, and from them the matrices $B_0(x)$ and $B_i(x)$ as follows:

$$B_0(x) = -x_0 A_4(y), \tag{1.2}$$

$$B_i(x) = A_i(y) - x_i A_4(y). \tag{1.3}$$

Then, we can show that if the fields $A_\alpha(y)$ satisfy the relations

$$y_\alpha A_\alpha = 0, \tag{1.4}$$

$$y_\beta \frac{\partial}{\partial y_\beta} A_\alpha = -A_\alpha, \tag{1.5}$$

the quantities $A_0(x)$ and $A_i(x)$ of Eqs. (1.2) and (1.3) are Yang–Mills fields in Minkowski space. From their definition we see immediately that these fields are real.

Using Eqs. (1.2) and (1.3) we can express the field strengths $M_{\mu\nu}(x)$ in Minkowski space in terms of the field strengths $F_{\alpha\beta}(y)$ in Euclidean space. We get

$$M_{ij}(x) = F_{ij}(y) - x_i F_{4j}(y) - x_j F_{i4}(y), \tag{1.6}$$

$$M_{0i}(x) = -x_0 F_{4i}(y). \tag{1.7}$$

II. EQUATIONS IN THE THREE-DIMENSIONAL EUCLIDEAN SPACE

In this section we shall combine the Yang–Mills field equations in Euclidean space with the conditions (1.4) and (1.5) and we shall obtain three equations in the three-dimensional Euclidean space. To find the general solution of Eq. (1.5) we must solve the system

$$\frac{dy_1}{y_1} = \frac{dy_2}{y_2} = \frac{dy_3}{y_3} = \frac{dy_4}{y_4} = -\frac{dA_\alpha}{A_\alpha}. \tag{2.1}$$

From the solution of this system we find that the general solution of Eq. (1.5) is

$$A_\alpha = (1/y_4)\varphi_\alpha(z_1, z_2, z_3), \tag{2.2}$$

where

$$z_i = y_i/y_4, \quad i = 1,2,3, \tag{2.3}$$

and the φ_α are arbitrary functions of z_i. Then we find that if the fields A_α are of the form of Eqs. (2.2) the condition (1.4) is satisfied if

$$\varphi_4 = -z_i\varphi_i. \tag{2.4}$$

To express the Euclidean Yang–Mills field strengths $F_{\alpha\beta} = \partial_\beta A_\alpha - \partial_\alpha A_\beta - [A_\alpha, A_\beta]$ in terms of φ_i let us write

$$f_{ij} = \partial_j\varphi_i - \partial_i\varphi_j - [\varphi_i, \varphi_j]. \tag{2.5}$$

Thus we get

$$F_{ij} = y_4^{-2} f_{ij}, \tag{2.6}$$

$$F_{i4} = y_4^{-2} z_j f_{ji}, \tag{2.7}$$

and we can express the Euclidean Yang–Mills field equations in terms of the "field strengths" f_{ij}, the three-dimensional "fields" φ_i, and the variables z_i. We find using Eqs. (2.6) and (2.7)

$$\partial_j F_{ji} + [A_j, F_{ji}] = y_4^{-3}\Big(\frac{\partial}{\partial z_j} f_{ji} + [\varphi_j, f_{ji}]\Big), \tag{2.8}$$

$$\partial_4 F_{4i} + [A_4, F_{4i}] = y_4^{-3} z_j\Big(z_l \frac{\partial}{\partial z_l} f_{ji} + 3f_{ji} + z_l[\varphi_l, f_{ji}]\Big). \tag{2.9}$$

Therefore the field equations $\partial_\mu F_{\mu i} + [A_\mu, F_{\mu i}] = 0$ become

$$\frac{\partial}{\partial z_j} f_{ji} + [\varphi_j, f_{ji}] + z_j\Big(z_l \frac{\partial}{\partial z_l} f_{ji} + 3f_{ji} + z_l[\varphi_l, f_{ji}]\Big) = 0. \tag{2.10}$$

Also, the field equation $\partial_j F_{j4} + [\varphi_j, F_{j4}] = 0$ becomes

$$z_i\Big(\frac{\partial}{\partial z_j} f_{ij} + [\varphi_j, f_{ij}]\Big) = 0. \tag{2.11}$$

Since $f_{ji} = -f_{ij}$ we find that Eq. (2.11) is satisfied if Eqs. (2.10) are satisfied. Therefore the Euclidean Yang–Mills field equations together with the conditions (1.4) and (1.5) are reduced to the Euclidean three-dimensional equations (2.10). Using Eqs. (1.1)–(1.3) and (2.2)–(2.4) we find that from any real solution of Eqs. (2.10) we get a real solution of the Yang–Mills field equations in Minkowski space.

III. DE ALFARO–FUBINI–FURLAN SOLUTION

To find solutions of Eqs. (2.10) we make the ansatz

$$\varphi_j = -\sigma_{jl}(z_l/z)G, \tag{3.1}$$

where $G = G[(z_1^2 + z_2^2 + z_3^2)^{1/2}] = G(z)$ and the matrices σ_{jl} are generators of an SO(3) group, i.e., satisfy the relations

$$[\sigma_{jl}, \sigma_{in}] = \delta_{ji}\sigma_{ln} - \delta_{jn}\sigma_{li} - \delta_{li}\sigma_{jn} + \delta_{ln}\sigma_{ji}. \tag{3.2}$$

Then we get

$$\partial_j f_{ji} + [\varphi_j, f_{ji}] = \sigma_{ij}\frac{z_j}{z}\Big(G'' + \frac{2}{z}G' - \frac{2}{z^2}G - \frac{3}{z}(G)^2 - (G)^3\Big), \tag{3.3}$$

where the prime means differentiation with respect to z. Also since

$$z_j\varphi_j = 0, \quad z_j f_{ji} = \sigma_{ij} z_j (G' + G/z), \tag{3.4}$$

we get

$$z_j\Big(z_l\frac{\partial}{\partial z_l} f_{ji} + 3f_{ji} + [z_l\varphi_l, f_{ji}]\Big) = \sigma_{ij} z_j (zG'' + 4G' + 2G/z). \tag{3.5}$$

From Eqs. (2.10), (3.3), and (3.5) we find that the function G must satisfy the equation

$$(z^2+1)G'' + \frac{2}{z}(2z^2+1)G' + \frac{2}{z^2}(z^2-1)G - \frac{3}{z}(G)^2 - (G)^3 = 0. \tag{3.6}$$

To simplify Eq. (3.6) let us write

$$V = zG + 1. \tag{3.7}$$

Then Eq. (3.6) becomes

$$z^2(z^2+1)V'' + 2z^3V' + V - V^3 = 0. \tag{3.8}$$

Also, if we define ζ by

$$\zeta = 1/z, \tag{3.9}$$

Eq. (3.8) takes the form

$$(\zeta^2 + 1)V'' + 2\zeta V' + V - V^3 = 0, \tag{3.10}$$

where the prime denotes differentiation with respect to ζ. Finally, if we introduce η by the relation

$$\eta = \operatorname{arccot} z, \tag{3.11}$$

Eq. (3.8) becomes

$$\cos^2\eta V'' + V - V^3 = 0, \tag{3.12}$$

where now the prime means differentiation with respect to η.

Equation (3.8) or its equivalent forms (3.10) and (3.12) are related with merons' differential equation in Euclidean space[11]

$$\rho^2\Big(\frac{\partial^2}{\partial\rho^2} + \frac{\partial^2}{\partial t^2}\Big)\varphi + \varphi - \varphi^3 = 0, \tag{3.13}$$

where $\rho = (y_1^2 + y_2^2 + y_3^2)^{1/2}$ and $t = y_4$. Indeed if φ is a function of $\rho/t = z$, Eq. (3.13) becomes Eq. (3.8). Therefore, we find that every solution of merons' differential equation, which is a function of ρ/t gives a real solution of the Yang–Mills equations in Minkowski space. Such solution is the one meron solution of Eq. (3.8), which corresponds to

$$V = 1/(z^2 + 1)^{1/2}. \tag{3.14}$$

Then, from Eqs. (3.1), (3.7), and (3.14) we find

$$\varphi_i = -\sigma_{ij}(z_j/z^2)[(z^2+1)^{-1/2} - 1] \tag{3.15}$$

from which we get

$$f_{ij} = \sigma_{ij}\frac{1}{z^2+1} - (\sigma_{il}z_j - \sigma_{jl}z_i)\frac{z_l}{z^2} \times\left[\frac{1}{z^2+1} - \frac{1}{(z^2+1)^{3/2}}\right]. \tag{3.16}$$

The Yang–Mills fields in Euclidean space, which correspond to the solution (3.14), are obtained from Eqs. (2.2), (2.4), and (3.15). Expressed in terms of the Euclidean space variables y_α by the relations (2.3), they are the following:

$$A_i(y) = -\sigma_{ij}\frac{y_j}{\mathbf{y}^2}\left[\frac{y_4}{(\mathbf{y}^2+y_4^2)^{1/2}} - 1\right],\quad A_4(y) = 0, \tag{3.17}$$

where $\mathbf{y}^2 = y_1^2 + y_2^2 + y_3^2$. The corresponding field strengths are obtained from Eqs. (2.6), (2.7), and (3.16) or directly from Eqs. (3.17). We get

$$F_{ij}(y) = \sigma_{ij}\frac{1}{\mathbf{y}^2 + y_4^2} - (\sigma_{il}y_j - \sigma_{jl}y_i)\frac{y_l}{\mathbf{y}^2} \times\left[\frac{1}{\mathbf{y}^2+y_4^2} - \frac{y_4}{(\mathbf{y}^2+y_4^2)^{3/2}}\right], \tag{3.18}$$

$$F_{i4}(y) = -\sigma_{il}y_l y_4/(\mathbf{y}^2+y_4^2)^{3/2}. \tag{3.19}$$

Eqs. (3.17)–(3.19) give a singular solution in Euclidean space in the $A_4 = 0$ gauge.

The corresponding real Minkowski space solution is obtained from Eqs. (1.2), (1.3), and (3.17). Using Eqs. (1.1) to express the fields in terms of the variables x_0, x_i we get

$$B_i(x) = -\sigma_{ij}\frac{x_j}{\mathbf{x}^2}\left(\frac{1+x_0^2-\mathbf{x}^2}{[(1+t_+^2)(1+t_-^2)]^{1/2}} - 1\right),$$

$$B_0(x) = 0, \tag{3.20}$$

where $\mathbf{x}^2 = x_1^2 + x_2^2 + x_3^2$ and $t_\pm = x_0 \pm |\mathbf{x}|$. Also, from Eqs. (1.6), (1.7), (3.18) and (3.19) or directly from Eqs. (3.20) we obtain the field strengths in Minkowski space,

$$M_{ij}(x) = \sigma_{ij}\frac{4}{(1+t_+^2)(1+t_-^2)} - (\sigma_{il}x_j - \sigma_{jl}x_i) \times\frac{4x_l}{\mathbf{x}^2}\left(\frac{1}{(1+t_+^2)(1+t_-^2)} - \frac{(1+x_0^2-\mathbf{x}^2)(1+\mathbf{x}^2)}{[(1+t_+^2)(1+t_-^2)]^{3/2}}\right), \tag{3.21}$$

$$M_{0i}(x) = -\sigma_{ij}x_j\frac{4x_0(1+x_0^2-\mathbf{x}^2)}{[(1+t_+^2)(1+t_-^2)]^{3/2}}. \tag{3.22}$$

Eqs. (3.20)–(3.22) give the de Alfaro–Fubini–Furlan solution in Minkowski space in the $B_0 = 0$ gauge.[13]

Other solutions of Eq. (3.8) are the following:

$$V = 0, \pm 1. \tag{3.23}$$

Then, from Eqs. (3.7) and (3.23) we get

$$G = C/z,\quad C = -1, 0, -2, \tag{3.24}$$

and the expression (3.4b) becomes

$$z_j f_{ji} = 0. \tag{3.25}$$

Using Eqs. (1.7), (2.7), and (3.25) we find that $M_{0i} = 0$, $i = 1,2,3$. Therefore the solutions (3.23) are not interesting for our problem.

IV. MORE GENERAL ANSATZ

To look for other solutions of Eq. (2.10) we make the ansatz

$$\varphi_j = (1/2i)(\sigma_j e + z_j\sigma_l z_l\, g + \epsilon_{jkl}\sigma_k z_l h), \tag{4.1}$$

where $e = e(z)$, $g = g(z)$, and $h = h(z)$. Substituting the above expressions in Eq. (2.5) we find

$$f_{jk} = (1/2i)[(\sigma_j z_k - \sigma_k z_j)(e'/z - g - eh - z^2gh) - \epsilon_{jkl}\sigma_l(e^2 + 2h) - \epsilon_{jkl}\sigma_s z_s z_l h^2 + (\epsilon_{jls}z_k - \epsilon_{kls}z_j)\sigma_l z_s(h'/z + eg)], \tag{4.2}$$

where the prime means differentiation with respect to z. To simplfy the notation let us define A and B by the relations

$$A = (e'/z) - g - eh - z^2gh, \tag{4.3}$$

$$B = zh' + z^2eg + 2h + e^2. \tag{4.4}$$

Then, we get from Eqs. (4.1) and (4.2),

$$\frac{\partial}{\partial z_j}f_{jk} = \frac{1}{2i}\left\{-\sigma_k(zA' + 2A) + z_k\sigma_l z_l\frac{A'}{z} + \epsilon_{kjl}z_j\sigma_l\left[\frac{1}{z}(e^2+2h)' - h^2 + z\left(\frac{h'}{z} + eg\right)' + 3\left(\frac{h'}{z} + eg\right)\right]\right\}, \tag{4.5}$$

$$[\varphi_j, f_{jk}] = (1/2i)\{\sigma_k[(e + z^2g)B + z^2eh^2] - z_k\sigma_l z_l[gB + h(A + eh)] + \epsilon_{kjl}z_j\sigma_l[(e+z^2g)A - h(e^2 + 2h + z^2h^2)]\}, \tag{4.6}$$

$$z_j f_{jk} = (1/2i)[(z_k z_j\sigma_j - \sigma_k z^2)A + \epsilon_{kjl}z_j\sigma_l B], \tag{4.7}$$

$$z_l\frac{\partial}{\partial z_l}(z_j f_{jk}) = \frac{1}{2i}[(z_k z_j\sigma_j - \sigma_k z^2)(zA' + 2A) + \epsilon_{kjl}z_j\sigma_l(zB' + B)], \tag{4.8}$$

$$[z_l\varphi_l, z_j f_{jk}] = (1/2i)[-(z_k z_j\sigma_j - \sigma_k z^2)(e + z^2g)B + \epsilon_{kjl}z_j\sigma_l z^2(e+z^2g)A]. \tag{4.9}$$

Substituting the expressions (4.5)–(4.9) in Eq. (2.10) and equating to zero the coefficients of σ_k, $z_k z_j\sigma_j$, and $\epsilon_{kjl}z_j\sigma_l$ we get, respectively,

$$z(1+z^2)A' + 2(1+2z^2)A - (1+z^2)(e+z^2g)B - z^2eh^2 = 0, \tag{4.10}$$

$$z(1+z^2)A' + 4z^2A - z^2hA - z^2[e + (1+z^2)g]B - z^2eh^2 = 0, \tag{4.11}$$

$$z(1+z^2)B' + (1+3z^2)B + z^2(1+z^2)(e+z^2g)A - (1+z^2h)(e^2+2h+z^2h^2) = 0. \tag{4.12}$$

Subtracting Eq. (4.11) from Eq. (4.10) we get

$$(2 + z^2h)A - eB = 0. \tag{4.13}$$

There three unknown functions e, g, and h are determined from the system of Eqs. (4.10), (4.12), and (4.13).

To proceed further we introduce new variables l, n, and p by the relations

$$l = 2 + z^2 h, \quad n = ze, \quad p = e + z^2 g, \tag{4.14}$$

and we define the functions a and b as follows

$$A = a/z^2, \quad B = b/z. \tag{4.15}$$

Then from Eqs. (4.3), (4.4), (4.14), and (4.15) we get

$$a = n' + (1 - l)p, \tag{4.16}$$

$$b = l' + np, \tag{4.17}$$

and Eq. (4.13) becomes

$$la - nb = 0. \tag{4.18}$$

Also, Eqs. (4.10) and (4.12) become, respectively,

$$z^2(1 + z^2)(a' - pb) + 2z^3 a - n(l-2)^2 = 0, \tag{4.19}$$

$$z^2(1 + z^2)(b' + pa) + 2z^3 b - (l-1)[n^2 + l(l-2)] = 0. \tag{4.20}$$

Therefore we have to solve the system of Eqs. (4.18)–(4.20), where a and b are given by Eqs. (4.16) and (4.17).

First, we check if there are solutions of Eqs. (4.18)–(4.20) having constant l and constant n. Of course, the case $l = n = 0$ is included. We find that this can happen if

$$p = n = l = 0, \quad \text{or} \quad p = n = 0, \quad l = 2, \quad \text{or } n = 0, l = 1, p\text{:unrestricted}. \tag{4.21}$$

Then, from Eqs. (4.15)–(4.17) we get $A = B = 0$. This solution, and generally every solution which gives $A = B = 0$, is not interesting for our problem because in this case from Eqs. (1.7), (2.7), and (4.7) we get $M_{0i} = 0$, $i = 1,2,3$.

Then, we check if there are solutions having $l = 0$ but $n = n(z) \neq 0$. In this case we get from Eqs. (4.16)–(4.18), $p = 0$, $a = n'$, and $b = 0$. But then Eq. (4.20) implies $n = 0$, which means that there are no solutions having $l = 0$ and $n \neq 0$.

Also, we check if there are solutions having $n = 0$ but $l = l(z) \neq 0$. In this case Eqs. (4.16)–(4.18) imply $b = l'$, $a = 0$, and $p = 0$. Then Eq. (4.19) is satisfied and Eq. (4.20) gives

$$z^2(1 + z^2)l'' + 2z^3 l' - l(l-1)(l-2) = 0. \tag{4.22}$$

Equation (4.22) is identical with Eq. (3.8) if we make the identification $V = l - 1$. This was expected since if $n = p = 0$ which implies $e = g = 0$, and $l = 2 + zG$, the ansatzes (3.1) and (4.1) become identical.

Consider now solutions having $l \neq 0$, $n \neq 0$ and at least one of them a function of z. In this case Eqs. (4.19) and (4.20) can be replaced by simpler expressions. Indeed if we multiply Eq. (4.19) by l, Eq. (4.20) by n, subtract the resulting expressions and use Eqs. (4.16)–(4.18) we get

$$z^2(1 + z^2)pb - n[(l-1)n^2 + l(l-2)] = 0. \tag{4.23}$$

Also, adding Eqs. (4.19) and (4.23) we get

$$z^2(1 + z^2)a' + 2z^3 a - n(l-1)[n^2 + 2(l-2)] = 0. \tag{4.24}$$

Therefore, we have to solve the system of Eqs. (4.16)–(4.18), (4.23), and (4.24). From this system we can eliminate a, b, and p, in which case we get a system of two equations with two unknowns. Indeed, from Eqs. (4.16)–(4.18) we get

$$p = (ln' - nl')/(n^2 + l^2 - l), \tag{4.25}$$

which together with Eqs. (4.16) and (4.17) may be used to eliminate a, b, and p. Then, Eqs. (4.23) and (4.24) become, respectively,

$$z^2(1 + z^2)l(ln' - nl')(nn' + ll' - l') - n(n^2 + l^2 - l)^2[(l-1)n^2 + l(l-2)] = 0, \tag{4.26}$$

$$z^2\left[(1 + z^2)\frac{n(nn' + ll' - l')}{n^2 + l^2 - l}\right]' - n(l-1)[n^2 + 2(l-2)] = 0. \tag{4.27}$$

We have not found the general solution of the system of Eqs. (4.26) and (4.27).

ACKNOWLEDGMENTS

I would like to thank Professor G. Furlan for a discussion and Professor I. Kahn for discussions. Also, I would like to thank Professor Abdus Salam, Professor P. Budini, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, where part of this work was done.

[1] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85–87 (1975).
[2] V. de Alfaro, S. Fubini, and G. Furlan, Phys. Lett. B **65**, 163–166 (1976); *ibid.* **72**, 203–207 (1977); *ibid.* **73**, 463–467 (1978).
[3] E. Witten, Phys. Rev. Lett. **38**, 121–124 (1977).
[4] G. t'Hooft (unpublished); E. Corrigan and D. Fairlie, Phys. Lett. B **67**, 69–71 (1977); F. Wilczek in *Quark Confinement and Field Theory*, edited by D. Stump and D. Weingarten (Wiley, New York, 1977), pp. 211–219.
[5] J. Cervero, L. Jacobs, and C. R. Nohl, Phys. Lett. B **69**, 351–354 (1977).
[6] S. Coleman, Phys. Lett. B **70**, 59–60 (1977).
[7] E. Corrigan, D. Fairlie, S. Templeton, and P. Goddard, Nucl. Phys. B **140**, 31–53 (1978).
[8] B. M. Schechter, Phys. Rev. D **16**, 3015–3020 (1977).
[9] M. Lüscher, Phys. Lett. B **70**, 321–324 (1977).
[10] W. Bernreuther, Phys. Rev. D **16**, 3609–3611 (1977).
[11] J. Glimm and A. Jaffe, Phys. Lett. B **73**, 167–170 (1978).
[12] A. Actor, Ann. Phys. (N.Y.) **121**, 181–203 (1979); Rev. Mod. Phys. **51**, 461–525 (1979).
[13] A. Actor, Lett. Math. Phys. **3**, 345–349 (1979).
[14] R. Jackiw, C. Nohl, and C. Rebbi in *Particles and Fields*, edited by D. H. Boal and A. N. Kamal (Plenum, New York, 1977), pp. 199–258.
[15] Y. Brihaye and J. Nuyts, J. Math. Phys. **21**, 909–912 (1980).
[16] C. H. Oh and Rosy Teh, Phys. Lett. B **87**, 83–86 (1979); Phys. Rev. D **21**, 1110–1112 (1980).
[17] L. Castillejo and M. Kugler, Phys. Lett. B **93**, 61–64 (1980).
[18] E. Kyriakopoulos, Phys. Lett. B **95**, 409–412 (1980).
[19] E. Kyriakopoulos, National Technical University of Athens preprint (to be published).

Invariant connections and magnetic monopoles

A. Cant
Department of Theoretical Physics, University of St. Andrews, St. Andrews, Fife KY 16 9SS, Scotland

(Received 1 May 1980; accepted for publication 25 July 1980)

Given a Lie group K acting on a principal fiber bundle $P(M, G)$, we study the K-invariance of connections in P and in bundles associated to P. The geometry of spontaneous symmetry breaking is discussed from this point of view. The results are applied to the Wu–Yang and 't Hooft–Polyakov models of a magnetic monopole.

PACS numbers: 11.10.Np, 11.30.Qc, 14.80.Hv, 02.20. – a

1. INTRODUCTION

It is well known that classical gauge fields can be regarded as connections in principal fiber bundles.[1,2] A very useful technique in searching for explicit solutions to the Yang–Mills equations is to require that they be invariant under an appropriate group of symmetry transformations.[3,4] In differential geometric terms, we have a Lie group K acting on a principal fiber bundle $P(M, G)$, and we initially try to characterize the set of all connections in P which are invariant under K. Then we can impose further constraints if required: dynamical equations, self-duality equations, etc.

In Sec. 2 of this paper we discuss the general theory of K-invariant connections in a principal fiber bundle $P(M, G)$, following mainly Kobayashi and Nomizu,[5] from which we quote Wang's theorem on the classification of K-invariant connections when M is a homogeneous space K/J. The consequences of K-invariance for associated bundles are then considered, in particular the case of spontaneous symmetry breaking.[2,6]

This general theory is applied in Sec. 3 to the Wu–Yang model of a magnetic monopole[7]: we show this solution may be obtained as the unique SU(2)-invariant connection on the Hopf bundle $S^3(S^2, \mathrm{U}(1))$. In Sec. 4, we consider the example of the 't Hooft–Polyakov model of a monopole.[8,9] This solution, for fixed radius r, may be obtained as the unique SO(3)-invariant connection in a reduced subbundle $Q(S^2, \mathrm{SO}(2))$ of the trivial bundle $S^2\times \mathrm{SO}(3)(S^2, \mathrm{SO}(3))$. This connection is projected from a two-parameter family of SO(3)-invariant connections in $S^2\times \mathrm{SO}(3)$.

2. INVARIANT CONNECTIONS AND BROKEN SYMMETRY

Let $P(M, G)$ be a principal fiber bundle, with projection π: $P\to M$. We say that a Lie group K acts on $P(M, G)$ as a group of bundle automorphisms if for each $k\in K$ there is a transformation k_P of P which commutes with the action of the structure group G:

$$k_P(ua) = k_P(u)a \quad (\forall u\in P,\ \forall a\in G).$$

Thus k_P maps each fiber of P into another fiber of P, and so induces a transformation k_M of M.

We fix a reference point $u_0\in P$, and let J be the isotropy subgroup of K at $x_0 = \pi(u_0)$:

$$J = \{j\in K \mid j_M x_0 = x_0\}.$$

For $j\in J$, we define $\lambda(j)\in G$ by

$$j_P u_0 = u_0[\lambda(j)]. \tag{2.1}$$

Then λ: $J\to G$ is a homomorphism[5]; we also denote the induced Lie algebra homomorphism by λ: $\mathscr{j}\to\mathscr{g}$.

In applications we may initially only have a group action defined on the base space M, and we need to know about the existence and uniqueness of a lifting of this action to an action by bundle automorphisms on $P(M, G)$. This is in general a difficult problem (see, for example, the literature cited in Ref. 10). For the case where K is compact and semisimple and the structure group G is solvable and connected, then Palais and Stewart[11] have shown that such an (essentially unique) lifting exists. Another case, which will be of great importance for us, occurs when K is compact and acts transitively on M, so that M can be identified with a homogeneous space K/J. There is now actually a one-to-one correspondence between

(a) equivalence classes of principal bundles $P(M, G)$ admitting a lifting of the K action on M to an action on $P(M, G)$, and

(b) conjugacy classes of homomorphisms λ: $J\to G$.

Details may be found in Ref. 4. The lifted action of K is clearly fiber-transitive; i.e. given any two fibers of P, there is a $k\in K$ such that one fiber is mapped to the other by k_P.

Let Γ be a connection in $P(M, G)$, i.e. a smooth distribution $u\to H_u$ on P such that

$$T_u(P) = H_u \oplus V_u \tag{2.2}$$

$$H_{ua} = dR_a(H_u) \quad \forall a\in G. \tag{2.3}$$

In (2.2), V_u denotes the subspace of the tangent space $T_u(P)$ to P at u consisting of vectors tangent to the fiber through u. Vectors in H_u (V_u) are called horizontal (vertical). In (2.3), dR_a denotes the differential of the map R_a: $P\to P$ which sends $u\to ua$.

Equivalently, the connection Γ is specified by its connection form ω, which is a $\mathscr{g}$-valued one-form on P satisfying[5]

$$\omega(A^*) = A, \quad \forall A\in\mathscr{g} \tag{2.4}$$

$$\delta R_a(\omega) = \mathrm{Ad}(a^{-1})\omega \quad (\forall a\in G). \tag{2.5}$$

Here, A^* denotes the vector field on P induced by $A\in\mathscr{g}$, Ad denotes the adjoint representation of G in $\mathscr{g}$, and $\delta R_a(\omega)$ denotes the pull-back of ω by R_a. The curvature form Ω of Γ is defined to be the exterior covariant derivative of ω, i.e.

$$\Omega(X, Y) = D\omega(X, Y) = d\omega(hX, hY), \quad \text{for } X, Y\in T_u(P),$$

where hX, hY are the horizontal components of X, Y.

If $\tau = (x_t)$, $0\leqslant t\leqslant 1$, is a (piecewise smooth) curve in M, then for each $v_0 \in \pi^{-1}(x_0)$ there exists a unique horizontal lift $\tau^* = (v_t)$ starting from v_0. The induced mapping

$$p_\tau:\ \pi^{-1}(x_0)\to\pi^{-1}(x_1),\quad v_0\to v_1$$

is called parallel displacement along τ. The holonomy group $\Phi(u)$ of Γ at $u\in P$ is the set of $a\in G$ such that u can be joined to ua by a horizontal curve.

Each $k\in K$ will map the connection Γ into another connection, denoted by $k(\Gamma)$, with horizontal subspaces $H'_{k_P(u)} = dk_P(H_u)$, and connection form $\delta k_P^{-1}(\omega)$.[5] If each $k\in K$ maps Γ into Γ, i.e. $dk_P(H_u) = H_{k_P(u)}$ $(\forall u\in P)$, then we say Γ is a K-invariant connection. We remark that if K is compact, then $P(M, G)$ admits a K-invariant connection (Ref. 12, p. 282).

The following important theorm, due to Wang, and proved in Ref. 5, gives a complete algebraic classification of the set of K-invariant connections on $P(M, G)$ in the case mentioned above, where K acts fiber-transitively on $P(M,G)$.

Theorem 2.1: Let K be connected and act fiber-transitively on $P(M, G)$, and let J be the isotropy subgroup of K at $x_0 = \pi(u_0)$, with corresponding homomorphism $\lambda: J\to G$ given by (2.1). Then there is a one-to-one correspondence between the set of K-invariant connections in P and the set of linear maps $\Lambda: \mathscr{k}\to\mathscr{g}$ satisfying

$$\Lambda(X) = \lambda(X)\quad (\forall X\in\mathscr{j}) \tag{2.6}$$

$$\Lambda(\mathrm{Ad}_{\mathscr{k}}(j)(X)) = \mathrm{Ad}_{\mathscr{g}}(\lambda(j))(\Lambda(X))\quad (\forall j\in J,\ \forall X\in\mathscr{k}), \tag{2.7}$$

where, for example, $\mathrm{Ad}_{\mathscr{g}}$ denotes the adjoint representation of G in $\mathscr{g}$. The correspondence is given by

$$\Lambda(X) = \omega_{u_0}(\tilde X)\quad (X\in\mathscr{k}),$$

where $\tilde X$ is the vector field on P induced by X.

We shall apply this theorem in Secs. 3 and 4.

We now consider how these concepts, particularly that of K-invariance, carry over to associated bundles. If G acts on a manifold F on the left, we let E denote the quotient $P\times_G F$ of $P\times F$ by the G-action

$$(u,\xi)a = (ua, a^{-1}\xi)\quad (u\in P,\ \xi\in F,\ a\in G).$$

As in Ref. 5, we write $E(M, F, G, P)$ for the bundle over M with fiber F associated to $P(M, G)$. The projection $\pi_E:\ E\to M$ is that induced by the map $(u,\xi)\to\pi(u)$ of $P\times F\to M$. Thus $\pi_E(u\xi) = \pi(u)$, where $u\xi$ is the image of $(u,\xi)\in P\times F$ under the natural projection $P\times F\to E$.

An important example is the bundle $E(M, G/H, G, P)$, where H is a closed subgroup of G. In this case E may be identified with P/H, the quotient of P by the right action of H. E admits a cross-section (i.e. a map $\phi: M\to E$ such that $\pi_E\circ\phi = \mathrm{id}_M$, the identity map of M) if and only if the structure group G of $P(M, G)$ is reducible to H, i.e. there is a reduced subbundle[5] Q of P over M with structure group H.

An automorphism of $E(M,F,G,P)$ is a transformation of E which takes a fiber into another fiber. We now prove

Proposition 2.1: K acts naturally as a group of automorphisms of E, inducing the given K action on M.

Proof: If $z\in E$, put $x = \pi_E(z)$ and choose $u\in\pi^{-1}(x)$ and $\xi\in F$ such that $u\xi = z$, then define

$$k_E(z) = k_P(u)\xi\quad (\forall k\in K).$$

This is independent of the choice of u and ξ, for if we take $u'\in\pi^{-1}(x)$ with $u' = ua$, say, for $a\in G$, then clearly we must have $\xi' = a^{-1}\xi$, in order that $z = u\xi = u'\xi'$. Then $k_P(u')\xi' = k_P(ua)a^{-1}\xi = k_P(u)aa^{-1}\xi = k_P(u)\xi$.

It is clear that $k_E: E\to E$ maps fibers into fibers, and that $(kk')_E = k_E k'_E (\forall k,k'\in K)$. If $z = u\xi$, then $\pi_E\circ k_E(z) = \pi_E(k_P(u)\xi) = \pi\circ k_P(u) = k_M\circ\pi(u) = k_M\circ\pi_E(z)$, so $\pi_E\circ k_E = k_M\circ\pi_E$, i.e. k_E induces the map k_M on M. □

Let $S(E)$ denote the set of cross sections of E:

$$S(E) = \{\phi:\ M\to E\,|\,\pi_E\circ\phi = \mathrm{id}_M\}.$$

Proposition 2.2: K acts naturally on $S(E)$.

Proof: We define, for $k\in K$, $\phi\in S(E)$:

$$k\phi = k_E\circ\phi\circ k_M^{-1}.$$

Clearly

$$\pi_E\circ(k\phi) = \pi_E\circ k_E\circ\phi\circ k_M^{-1} = k_M\circ\pi_E\circ\phi\circ k_M^{-1} = k_M\circ k_M^{-1} = \mathrm{id}_M.$$

(by Prop. 2.1). Thus $k\phi\in S(E)$. Also for k, $k'\in K$:

$$(kk')\phi = (kk')_E\circ\phi\circ(kk')_M^{-1} = k_E k'_E\circ\phi\circ k'^{-1}_M k_M^{-1} = k(k'\phi).$$ □

We shall denote by $S^K(E)$ the set of K-invariant sections of E; $S^K(E) = \{\phi\in S(E)\,|\,k_E\circ\phi = \phi\circ k_M,\ \forall k\in K\}$.

Note that $S(E)$ is in one-to-one correspondence with the set $\mathscr{L}_G(P) = \{f: P\to F\,|\,f(ua) = a^{-1}f(u),\ \forall u\in P,\ \forall a\in G\}$, under the mapping

$$S(E)\to\mathscr{L}_G(P)$$
$$\phi\to\tilde\phi,$$

where $\phi(x) = u\tilde\phi(u)$ [if $u\in\pi^{-1}(x)$]. The K-action on $S(E)$ induces the following K-action on $\mathscr{L}_G(P)$:

$$k\tilde\phi = \widetilde{k\phi} = \tilde\phi\circ k_P^{-1}.$$

We denote by $\mathscr{L}_G^K(P)$ the K-invariant elements of $\mathscr{L}_G(P)$; so

$$\mathscr{L}_G^K(P) = \{\tilde\phi\in\mathscr{L}_G(P)\,|\,\tilde\phi\circ k_P = \tilde\phi,\ \ \forall k\in K\}.$$

A connection in $E(M, F, G, P)$ is defined[13] to be a smooth distribution $Q: z\to Q_z$ on E such that $T_z(E) = W_z\oplus Q_z$ [W_z being the subspace of $T_z(E)$ which is tangent at z to the fiber through z], and for each curve $\tau = (x_t)$, $0\leqslant t\leqslant 1$, in M and $z_0\in\pi_E^{-1}(x_0)$, there exists a unique one-dimensional integral manifold of Q passing through z_0 which projects onto τ.

The connection Γ induces a unique connection in E, obtained as follows.[13] If $z\in E$, with $x = \pi_E(z)$, fix $u\in\pi^{-1}(x)$. Then there is a unique $\xi\in F$ such that $u\xi = z$. So for ξ fixed we have a map

$$\eta_\xi:\ P\to E,$$
$$v\to v\xi,$$

whose differential $d\eta_\xi$ at u takes $T_u(P)\to T_z(E)$. We then define $Q_z = d\eta_\xi(H_u)$. This is independent of u and ξ, and one can show that this is a connection in E.

If $\tau = (x_t)$, $0\leqslant t\leqslant 1$, is a curve in M, then the parallel displacement in E corresponding to τ is the map

$$p_\tau^E:\ \pi_E^{-1}(x_0)\to\pi_E^{-1}(x_1)$$
$$u_0\xi\to u_1\xi = [p_\tau(u_0)]\xi,$$

$(u_0 \in \pi^{-1}(x_0))$. A cross section $\phi \in S(E)$ is said to be parallel if, for every curve $\tau = (x_t)$ in M, we have

$$p_\tau^E(\phi(x_0)) = \phi(x_1).$$

Proposition 2.3: If Γ is a K-invariant connection in $P(M,G)$, then the induced connection in $E(M,F,G,P)$ is K-invariant in the sense that

$$dk_E: \; Q_z \to Q_{k_1(z)}.$$

Proof: We observe that, for $v \in P$,

$$k_E \circ \eta_\xi(v) = k_E(v\xi) = k_P(v)\xi = \eta_\xi \circ k_P(v).$$

So $k_E \circ \eta_\xi = \eta_\xi \circ k_P$, and

$$\begin{aligned} dk_E(Q_z) &= dk_E \circ d\eta_\xi(H_u) = d(k_E \circ \eta_\xi)H_u \\ &= d(\eta_\xi \circ k_P)H_u = d\eta_\xi H_{k_P(u)} = Q_{k_1(z)}. \end{aligned}$$ □

Proposition 2.4: If $\tau = (x_t)$ is a curve in M, then in $E(M, F, G, P)$ we have

(i) $k_E \circ p_\tau^E = (p_{k_M \circ \tau}^E) \circ k_E, \quad \forall k \in K;$

(ii) If $\phi \in S(E)$ is parallel, then so is $k\phi$, $\forall k \in K$.

Proof: (i) If $u_0 \in \pi^{-1}(x_0)$ then for $\xi \in F$,

$$\begin{aligned} k_E \circ p_\tau^E(u_0\xi) &= k_E[p_\tau(u_0)\xi] \\ &= [k_P \circ p_\tau(u_0)]\xi \\ &= [(p_{k_M \circ \tau}) \circ k_P(u_0)]\xi \end{aligned}$$

[using the fact that the curve $k_P \circ \tau^*$ is the unique horizontal lift of $k_M \circ \tau$ starting at $k_P(u_0)$]

$$= p_{k_M \circ \tau}^E(k_P(u_0)\xi) = p_{k_M \circ \tau}^E \circ k_E(u_0\xi).$$

(ii) If $\phi \in S(E)$ is parallel, then

$$\begin{aligned} p_\tau^E((k\phi)(x_0)) &= p_\tau^E(k_E \circ \phi \circ k_M^{-1}(x_0)) \\ &= k_E \circ (p_{k_M^{-1} \circ \tau}^E) \circ \phi(k_M^{-1}(x_0)) \\ &= k_E \circ \phi(k_M^{-1}(x_1)) = (k\phi)(x_1). \end{aligned}$$ □

If ρ is a representation of G on a vector space V over $\mathbb{C}$, then we can form $E(M, V, G, P)$, the (complex) vector bundle associated with $P(M, G)$. All that we have said concerning general associated bundles is valid for vector bundles. In addition, we have the following properties arising from the vector space structure of V:

(1) Each fiber $\pi_E^{-1}(x)$, $x \in M$, of E has a vector space structure such that the map $u: \xi \to u\xi$ of $V \to \pi_E^{-1}(x)$ is a linear isomorphism, $\forall u \in \pi^{-1}(x)$.

(2) The set $S(E)$ of cross-sections of E is a vector space over $\mathbb{C}$, with

$$(\phi + \psi)(x) = \phi(x) + \psi(x) \quad (\lambda\phi)(x) = \lambda(\phi(x)),$$

$$(\phi, \psi \in S(E), \quad \lambda \in \mathbb{C}, \quad x \in M).$$

It is also a module over the algebra $C^\infty(M)$ of smooth real-valued functions on M, with

$$f(\phi)(x) = f(x)\phi(x) \quad (f \in C^\infty(M), \phi \in S(E), x \in M).$$

(3) If K acts on $P(M, G)$, then K acts as a group of automorphisms of E: in particular, for each $k \in K$, k_E is linear on fibers of E. Note that in general k_E is *not* a homomorphism of $C^\infty(M)$-modules; we have

$$k(f\phi) = (f \circ k_M^{-1})(k\phi) \quad (f \in C^\infty(M), k \in K, \phi \in S(E)).$$

(4) If $\phi \in S(E)$ and $X \in \chi(M)$ (the set of smooth vector fields on M), then the covariant-derivative $\nabla_X \phi \in S(E)$ of ϕ with respect to X is defined by

$$(\nabla_X \phi)(x) = \lim_{t \to 0} \frac{(p_\tau^E)_0^t \phi(x_t) - \phi(x)}{t},$$

where $\tau = (x_t)$ is the unique maximal integral curve of X starting at x_0 and $(p_\tau^E)_0^t$ denotes parallel displacement[5,13] from $\pi_E^{-1}(x_t) \to \pi_E^{-1}(x)$.

It can be shown that, under the action of K,

$$\begin{aligned} \nabla_X(k\phi) &= k_E \circ \nabla_{dk_M^{-1}(X)} \phi \\ &= [k(\nabla_{dk_M^{-1}(X)} \phi)] \circ k_M. \end{aligned}$$

We now turn to the situation where we have spontaneously broken symmetry and investigate what happens to a K-invariant connection. The geometry of broken symmetry has been clearly stated by Trautman[2] (see also Ref. 6). We consider those cross-sections $\phi \in S(E)$ such that $\tilde{\phi}(P)$ (which is invariant under the action of G on V) is actually an orbit W of G in V. Normally, W is not a subspace of V, but it is a submanifold, diffeomorphic to G/H, where $H = H_w$ is the isotropy group of $w \in W$.

Such a $\tilde{\phi}$ will then correspond to a unique cross-section $\phi' \in S(E')$, where $E'(M, G/H, G, P)$ is the bundle associated to $P(M, G)$ with fiber G/H. Thus it corresponds to a unique reduction of $P(M, G)$ to a subbundle $Q(M, H)$. We can easily see that[5]

$$\begin{aligned} Q &= \{u \in P \,|\, \mu(u) = u\tilde{\phi}(u)\} \\ &= \{u \in P \,|\, \tilde{\phi}(u) = H\}, \end{aligned}$$

where μ denotes the natural projection from P to $P/H \simeq E'$.

From Ref. 5 we quote the following:

Proposition 2.5: Suppose that the Lie algebra $\mathfrak{g}$ of G can be written as $\mathfrak{g} = \mathfrak{m} \oplus \mathfrak{h}$ (a vector space direct sum) where $\mathfrak{m}$ is a subspace of $\mathfrak{g}$ with $\mathrm{ad}_{\mathfrak{g}}(H)(\mathfrak{m}) = \mathfrak{m}$. Then, given a connection form ω in P, the $\mathfrak{h}$-component ω' of $\omega|_Q$ is a connection form in Q, called the projected connection form on $Q(M, H)$.

An important result is

Proposition 2.6: Under the above conditions, the following five statements are equivalent:

(i) ω is reducible to a connection in Q.

(ii) $\omega' = \omega|_Q$, i.e. the restriction of ω to Q is $\mathfrak{h}$-valued.

(iii) The corresponding cross-section ϕ' is parallel.

(iv) ϕ' is covariant constant: $\nabla_X \phi' = 0$ $(\forall X \in \chi(M))$.

(v) $\exists u \in P$ such that $\Phi(u) \subseteq H$, where $\Phi(u)$ denotes the holonomy group at $u \in P$.

Proof: The equivalence of (i), (ii), (iii) and (iv) is proved in Ref. 5 (pp. 83, 88, and 114); the equivalence of (iv) and (v) is shown in Ref. 14 (p. 35). □

Proposition 2.7: The map $\tilde{\phi}$ is invariant under K [i.e. $\tilde{\phi} \in \mathscr{L}_G^K(P)$] if and only if K acts on the reduced bundle $Q(M, H)$ as a group of automorphisms.

Proof: If $\tilde{\phi} \in \mathscr{L}_G^K(P)$, then $\tilde{\phi} \circ k_P = \tilde{\phi}$, $\forall k \in K$, so if $u \in Q$:

$$\tilde{\phi}(k_P(u)) = \tilde{\phi}(u) = H.$$

Thus $k_P(u) \in Q$. The fiber of Q passing through u consists of all elements of the form ua $(a \in H)$, so it is clear that K acts on

$Q(M, H)$ as a group of bundle automorphisms.

For the converse, if $K(Q)\subseteq Q$, then if $u\in Q$, $\tilde{\phi}(u) = \tilde{\phi}(k_P(u))$. Then, if $v\in P$, $v = ua$ for some $u\in Q$ and $a\in G$, so

$$\tilde{\phi}\circ k_P(v) = \tilde{\phi}(k_P(ua)) = \tilde{\phi}(k_P(u)a) = \rho(a^{-1})\tilde{\phi}\circ k_P(u) = \rho(a^{-1})\tilde{\phi}(u) = \tilde{\phi}(v), \quad \text{as required.} \qquad \square$$

Proposition 2.8: Under the conditions of Propositions 2.5 and 2.7, if ω is a K-invariant connection form in P, then the projected connection form ω' is K-invariant.

Proof: Let η be the $\mathscr{m}$-component of $\omega|_Q$, and let $u\in Q$, $X\in T_u(Q)$. Then

$$\omega(X) = \omega'(X) + \eta(X).$$

But $\omega(X) = \omega(dk_P(X)) = \omega'(dk_P(X)) + \eta(dk_P(X))$, $\forall k\in K$, [since $dk_P(X)\in T_{k_P(u)}(Q)$]. Thus $\omega'(dk_P(X)) = \omega'(X)$, i.e. $\delta k_P \cdot \omega' = \omega'$. $\square$

In the case where K acts fiber-transitively on P, the induced K-action on Q will also be fiber-transitive. If we take $u_0\in Q$, then Theorem 2.1 will also apply to $Q(M, H)$: there is a one-to-one correspondence between the set of K-invariant connections $\bar{\omega}$ in Q and the set of linear maps $\bar{\Lambda}$: $\mathscr{k}\rightarrow\mathscr{h}$ satisfying

$$\bar{\Lambda}(X) = \lambda(X) \quad (\forall X\in\mathscr{j}) \tag{2.8}$$

$$\bar{\Lambda}(\mathrm{Ad}_{\mathscr{k}}(j)(X)) = \mathrm{Ad}_{\mathscr{h}}(\lambda(j))(\bar{\Lambda}(X)) \quad (\forall j\in J,\ \forall X\in\mathscr{k}). \tag{2.9}$$

Note that J remains the same, and λ: $J\rightarrow H$. The correspondence is given by

$$\bar{\Lambda}(X) = \bar{\omega}_{u_0}(\tilde{X}|_Q) \quad (X\in\mathscr{k}).$$

Clearly the projected K-invariant connection form ω' in Q corresponds to the map

$$\Lambda': \mathscr{k}\rightarrow\mathscr{h}$$
$$X\rightarrow\mathscr{h}\text{-component of } \Lambda(X).$$

In this case, it is clear that ω is reducible to a connection in Q if and only if the corresponding map Λ maps $\mathscr{k}$ to $\mathscr{h}$.

3. THE WU–YANG MODEL OF DIRAC'S MONOPOLE

It is well known that the magnetic monopole of lowest strength $g = \frac{1}{2}(\hbar c/e)$ has a geometric interpretation as a connection in the (nontrivial) Hopf bundle $S^3(S^2, \mathrm{U}(1))$.[2,15,16] We shall discuss this model from the point of view of the theory of K-invariant connections.

We recall[15,16] the following properties of the Hopf bundle $S^3(S^2, \mathrm{U}(1))$. It is convenient to picture S^3 as the set of points $z = (z_0, z_1)\in\mathbb{C}^2$ with $|z_0|^2 + |z_1|^2 = 1$; we shall also write $z_0 = x_1 + ix_2$, $z_1 = x_3 + ix_4$. The gauge group U(1) acts on S^3 by $(z_0, z_1)a = (z_0a, z_1a)$, for $a\in\mathrm{U}(1)$. The projection π (the Hopf map) is defined to be $\pi = \eta\circ p$, where p: $S^3\rightarrow\mathbb{C}$ takes (z_0, z_1) to z_0/z_1, and η: $\mathbb{C}\rightarrow S^2$ is the stereographic map, given by

$$\eta(w) = \left(\frac{2x}{1+|w|^2}, \frac{2y}{1+|w|^2}, \frac{1-|w|^2}{1+|w|^2}\right),$$

where $w = x + iy$.

The group SU(2) acts naturally on S^3 on the left by $kz = (\alpha z_0 + \beta z_1, \gamma z_0 + \delta z_1)$, where $k = \left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\in\mathrm{SU}(2)$ and $z = (z_0, z_1)\in S^3$. It is clear that this action commutes with the U(1) action, so SU(2) acts on $S^3(S^2, \mathrm{U}(1))$ as a group of bundle automorphisms. The induced action on $\mathbb{C}$ is

$$kw = \frac{\alpha w + \beta}{\gamma w + \delta},$$

which corresponds to the natural action of SO(3) on the sphere S^2 (see, for example, Ref. 17). Since this action is transitive, Theorem 2.1 is applicable.

It will be convenient to take our reference point to be $u = (0, 1)\in S^3$; $\pi(u)$ is then the North pole $\mathbf{n} = (0, 0, 1)\in S^2$. The isotropy subgroup of SU(2) at $\mathbf{n}$ is

$$J = \left\{\begin{pmatrix} e^{i\theta} & 0\\ 0 & e^{-i\theta}\end{pmatrix}\middle|\,\theta\in\mathbb{R}\right\} \cong \mathrm{U}(1)\cong\mathrm{SO}(2)$$

and, by (2.1), the homomorphism λ: $J\rightarrow\mathrm{U}(1)$ (in this case an isomorphism) is given by

$$\lambda: \begin{pmatrix} e^{i\theta} & 0\\ 0 & e^{-i\theta}\end{pmatrix}\rightarrow e^{-i\theta},$$

which induces the Lie algebra isomorphism λ: $\mathscr{j}\rightarrow\mathrm{u}(1)$ given by

$$\lambda: \begin{pmatrix} i\theta & 0\\ 0 & -i\theta\end{pmatrix}\rightarrow -i\theta.$$

Proposition 3.1: There is exactly one SU(2)-invariant connection in $S^3(S^2, \mathrm{U}(1))$.

Proof: Since SU(2) is connected and acts fiber-transitively on $S^3(S^2, \mathrm{U}(1))$, by Theorem 2.1 there is a one-to-one correspondence between the set of SU(2)-invariant connections in S^3 and the set of linear maps Λ: su(2)$\rightarrow$u(1) satisfying (2.6) and (2.7) with $K = \mathrm{SU}(2)$, $G = \mathrm{U}(1)$. We choose the basis $\{i\sigma_k\,|\,k = 1, 2, 3\}$ for su(2), where the σ_k are the Pauli matrices, and write $\Lambda(i\sigma_k) = i\Lambda_k$ ($\Lambda_k\in\mathbb{R}$). Then (2.6) becomes simply $\Lambda_3 = -1$, while from (2.7) we find, substituting $X = i\sigma_2, i\sigma_3$, that

$$\Lambda_1 = \Lambda_1\cos2\theta + \Lambda_2\sin2\theta \quad (\forall\theta\in\mathbb{R})$$
$$\Lambda_2 = -\Lambda_1\sin2\theta + \Lambda_2\cos2\theta.$$

This says that the point $(\Lambda_1, \Lambda_2)\in\mathbb{R}^2$ is fixed under all rotations about the origin, which forces $\Lambda_1 = \Lambda_2 = 0$.

Thus there is a unique linear map Λ: su(2)$\rightarrow$u(1) satisfying (2.6) and (2.7); this proves the proposition. The map Λ is given by

$$\Lambda: \begin{pmatrix} i\alpha & \beta\\ -\bar{\beta} & -i\alpha\end{pmatrix}\rightarrow -i\alpha \quad (\alpha\in\mathbb{R},\ \beta\in\mathbb{C}). \qquad \square$$

This (unique) SU(2)-invariant connection Γ has very special properties. Since $\mathscr{k}$ can be written as $\mathscr{k} = \mathscr{j}\oplus\mathscr{m}$, where $\mathscr{m}$ is the subspace spanned by $\{i\sigma_1, i\sigma_2\}$, and $\mathrm{Ad}_{\mathscr{k}}(J)(\mathscr{m}) = \mathscr{m}$, we see that, since $\Lambda_1 = \Lambda_2 = 0$, Γ is the so-called *canonical connection* in $S^3(S^2, \mathrm{U}(1))$ (Ref. 5, p. 110). It is irreducible [the holonomy group of Γ coincides with the gauge group U(1)], and not flat (its curvature form Ω does not vanish everywhere). These statements follow immediately from Ref. 5 (Proposition 11.4 and Corollary 11.6 of Chapter II). However, the most important property for us (which we shall demonstrate shortly) is that Γ is exactly the connection corresponding to a magnetic pole of lowest strength $g = \frac{1}{2}(\hbar c/e)$. It was observed by Trautman[15] that this latter connection is SU(2)-invariant,[18] but we have shown that the strategy of looking for all possible SU(2)-invariant connections leads to a unique answer.

We now calculate the connection 1-form ω of Γ. By Theorem 2.1, at the reference point $u = (0, 1)$ we have

$$\omega_u(\widetilde{X}) = \Lambda(X) \quad (X \in \mathrm{su}(2)),$$

where $\widetilde{X}$ is the vector field on S^3 induced by X, given by

$$\widetilde{X}_u f = \frac{df}{dt}(\exp(tX)u)\Big|_{t=0} \quad (\forall f \in C^\infty(S^3)).$$

If we put $X_k = i\sigma_k$, then a simple calculation yields

$$(\widetilde{X}_1)_u = \frac{\partial}{\partial x_2}\Big|_u ; \quad (\widetilde{X}_2)_u = \frac{\partial}{\partial x_1}\Big|_u ;$$
$$(\widetilde{X}_3)_u = -\frac{\partial}{\partial x_4}\Big|_u . \tag{3.1}$$

We therefore have

$$\omega_u\left(\frac{\partial}{\partial x_1}\right) = \omega_u\left(\frac{\partial}{\partial x_2}\right) = 0, \quad \omega_u\left(\frac{\partial}{\partial x_4}\right) = i.$$

To evaluate ω at a general point $z' = (z_0', z_1') \in S^3$, we use the fact that SU(2) acts fiber-transitively on S^3. We have $u = kz'$, where

$$k = \begin{pmatrix} z_1' & -z_0' \\ \bar{z}_0' & \bar{z}_1' \end{pmatrix} \in \mathrm{SU}(2).$$

If $Y \in T_{z'}(S^3)$, then we can write

$$dk(Y) = \widetilde{X}_u + A_u^*,$$

for some $X \in \mathrm{su}(2)$ and $A \in \mathrm{u}(1)$. We then have (Ref. 5, p. 107)

$$\omega_{z'}(Y) = \Lambda(X) + A. \tag{3.2}$$

If we calculate $dk(Y)$ for $Y = \partial/\partial x_i|_{z'}$, then we obtain

$$dk\left(\frac{\partial}{\partial x_1}\Big|_{z'}\right) = x_3'(\widetilde{X}_2)_u + x_4'(\widetilde{X}_1)_u + x_2'(\widetilde{X}_3)_u,$$
$$dk\left(\frac{\partial}{\partial x_2}\Big|_{z'}\right) = -x_4'(\widetilde{X}_2)_u + x_3'(\widetilde{X}_1)_u - x_1'(\widetilde{X}_3)_u,$$
$$dk\left(\frac{\partial}{\partial x_3}\Big|_{z'}\right) = -x_1'(\widetilde{X}_2)_u - x_2'(\widetilde{X}_1)_u + x_4'(\widetilde{X}_3)_u,$$
$$dk\left(\frac{\partial}{\partial x_4}\Big|_{z'}\right) = x_2'(\widetilde{X}_2)_u - x_1'(\widetilde{X}_1)_u - x_3'(\widetilde{X}_3)_u, \tag{3.3}$$

where $z_0' = x_1' + ix_2'$, $z_1' = x_3' + ix_4'$, and we have used (3.1). Thus we have $A = 0$ in (3.2) and so, by (3.3):

$$\omega_{z'}\left(\frac{\partial}{\partial x_1}\right) = -ix_2', \quad \omega_{z'}\left(\frac{\partial}{\partial x_2}\right) = ix_1',$$
$$\omega_{z'}\left(\frac{\partial}{\partial x_3}\right) = -ix_4', \quad \omega_{z'}\left(\frac{\partial}{\partial x_4}\right) = ix_3'.$$

Therefore the connection form ω of Γ is given by

$$\omega = -i(x_2dx_1 - x_1dx_2 + x_4dx_3 - x_3dx_4), \tag{3.4}$$

which is exactly the connection given in Refs. 15 and 16. Its curvature is

$$\Omega = d\omega = 2i(dx_1 \wedge dx_2 + dx_3 \wedge dx_4) \tag{3.5}$$

which corresponds to the field of a magnetic monopole of lowest strength $g = \frac{1}{2}(\hbar c/e)$.

We conclude this section with some brief remarks on higher magnetic pole strengths. The Hopf bundle is modified as follows. We consider the lens space $L(n, 1)$, which is the quotient manifold $S^3/\mathbb{Z}_n$ of S^3 under the free discontinuous action of $\mathbb{Z}_n$ on the right given by[19]

$$(z_0, z_1)p = (z_0e^{2\pi ip/n}, z_1e^{2\pi ip/n}),$$
$$p \in \mathbb{Z}_n, \quad (z_0, z_1) \in S^3.$$

Let $\mu_n: S^3 \to S^3/\mathbb{Z}_n$ denote the projection. Since (z_0, z_1) and $(z_0, z_1)p$ always lie in the same fiber of S^3 over S^2, it is clear that μ_n is an S^2-homomorphism from $S^3(S^2, \mathrm{U}(1))$ to $S^3/\mathbb{Z}_n\,(S^2, \mathrm{U}(1))$. ($\mu_n$ induces the identity map on S^2). Note that the induced map, also denoted by μ_n, of U(1) to U(1) is the covering isomorphism $a \to a^n$. (We also denote the corresponding Lie algebra isomorphism $A \to nA$ by μ_n). This follows from the fact that the standard fiber of $S^3/\mathbb{Z}_n$ over S^2 is[20] $\mathrm{U}(1)/\mathbb{Z}_n$, which is isomorphic to U(1) under the map

$$\mathrm{U}(1)/\mathbb{Z}_n \to \mathrm{U}(1)$$
$$\mu_n(a) \to a^n \quad (a \in \mathrm{U}(1)).$$

Clearly SU(2) also acts by bundle automorphisms on $S^3/\mathbb{Z}_n\,(S^2, \mathrm{U}(1))$. The unique SU(2)-invariant connection ω in S^3 is mapped by μ_n to the SU(2)-invariant connection ω_n in $S^3/\mathbb{Z}_n$; we have

$$\delta\mu_n(\omega_n) = \mu_n \cdot \omega = n\omega. \tag{3.6}$$

[Here, as in Ref. 5, p. 82, $\mu_n \cdot \omega$ denotes the u(1)-valued 1-form on S^3 defined by $(\mu_n \cdot \omega)(X) = \mu_n(\omega(X)) = n\omega(X)$].

It follows from (3.4) and (3.6) that ω_n corresponds to a magnetic monopole of strength $g = \frac{1}{2}n(\hbar c/e)$.[15,20]

We could, of course, apply Theorem 2.1 directly to find the SU(2)-invariant connection ω_n in $S^3/\mathbb{Z}_n$. The argument is entirely analogous to the situation for S^3; choosing $\mu_n(u)$ as our reference point, we see that J is the same, while the homomorphism $\lambda: J \to \mathrm{U}(1)$ is replaced by λ^n. From conditions (2.6) and (2.7) we conclude that there is a unique SU(2)-invariant connection ω_n in $S^3/\mathbb{Z}_n$, given by the map $n\Lambda$: su(2)$\to$u(1), where Λ is defined in Proposition 3.1.

4. THE 't HOOFT–POLYAKOV MODEL OF A MONOPOLE

An interesting example of the geometric theory of the role of K-invariant connections in spontaneous symmetry breaking, described in Sec. 2, is the 't Hooft–Polyakov model of a magnetic monopole.[2,8,9,21]

We recall[2] that in this model, the gauge group G is SO(3) and the base space M can be identified with S^2 (if r is fixed). We begin with the trivial bundle $P = S^2 \times \mathrm{SO}(3)$. In this case K is also SO(3); its natural transitive action on S^2 lifts to the following action by bundle automorphisms on P:

$$k_P(\mathbf{r}, a) = (k\mathbf{r}, ka)$$
$$\forall \mathbf{r} \in S^2, \ \forall k, a \in \mathrm{SO}(3).$$

We choose $(\mathbf{n}, I)$ as our reference point in P, where $\mathbf{n} = (0, 0, 1)$ and I is the identity element of SO(3). The isotropy group J therefore consists of all rotations which leave $\mathbf{n}$ invariant, i.e. $\mathrm{SO}(2) \cong \mathrm{U}(1)$. Clearly $\lambda: J \to G$ is just the identity map $j \to j$ of SO(2).

Taking ρ to be the natural (or adjoint) representation of SO(3) acting in the space $V = \mathbb{R}^3$, we form the vector bundle

E associated with P with fiber V. The Higgs field[2] $\phi \in S(E)$ is given [in terms of the corresponding element $\tilde{\phi}$ of $\mathscr{L}_G(P)$] by $\tilde{\phi}(\mathbf{r}, a) = a^{-1}\mathbf{r}$. Clearly $\tilde{\phi}(P) = S^2 = G/H = \mathrm{SO}(3)/\mathrm{SO}(2)$, where $H = \mathrm{SO}(2)$ is again the isotropy group of $\mathbf{n}$. Thus $\tilde{\phi}$ corresponds to a unique reduction of $S^2 \times \mathrm{SO}(3)(S^2, \mathrm{SO}(3))$ to a subbundle $Q(S^2, \mathrm{SO}(2))$; we have $Q = \{(\mathbf{r}, a) \in P \mid a^{-1}\mathbf{r} = \mathbf{n}\}$.

We now find the SO(3)-invariant connections in P. By Theorem 2.1 these are in one-to-one correspondence with the set of liner maps Λ: so(3)$\to$so(3) satisfying (2.6) and (2.7). We can take $\{X_i\}$ as a basis for so(3), where

$$X_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$
$$X_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Clearly $\mathscr{J} = \mathscr{h} = \mathrm{so}(2)$ is spanned by X_3. We write $\Lambda(X_i) = \Lambda_{ki}X_k$, for $i = 1, 2, 3$ (summation over k implied). Then (2.6) becomes $\Lambda_{k3} = \delta_{k3}$; and (2.7) becomes

$$\begin{aligned} &\Lambda(X_1)\cos\theta + \Lambda(X_2)\sin\theta = j^{-1}\Lambda(X_1)j \\ &-\Lambda(X_1)\sin\theta + \Lambda(X_2)\cos\theta = j^{-1}\Lambda(X_2)j, \\ &\forall j \in \mathrm{SO}(2), \end{aligned} \tag{4.1}$$

where

$$j = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since we have

$$\begin{aligned} &j^{-1}X_1 j = X_1\cos\theta + X_2\sin\theta, \\ &j^{-1}X_2 j = -X_1\sin\theta + X_2\cos\theta, \\ &j^{-1}X_3 j = X_3, \end{aligned} \tag{4.2}$$

we find that (4.1) gives the following conditions on the Λ_{ki}:

$$\Lambda_{11} = \Lambda_{22}, \quad \Lambda_{21} = -\Lambda_{12}, \quad \Lambda_{31} = \Lambda_{32} = 0.$$

Thus we obtain a two-parameter family of linear maps Λ: so(3)$\to$so(3) which satisfy (2.6) and (2.7):

$$\Lambda:\ X = \begin{pmatrix} 0 & \alpha_3 & -\alpha_2 \\ -\alpha_3 & 0 & \alpha_1 \\ \alpha_2 & -\alpha_1 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 0 & \alpha_3 & \Lambda_{12}\alpha_1 - \Lambda_{11}\alpha_2 \\ -\alpha_3 & 0 & \Lambda_{11}\alpha_1 + \Lambda_{12}\alpha_2 \\ -\Lambda_{12}\alpha_1 + \Lambda_{11}\alpha_2 & -\Lambda_{11}\alpha_1 - \Lambda_{12}\alpha_2 & 0 \end{pmatrix} \tag{4.3}$$

$(\Lambda_{11}, \Lambda_{12} \in \mathbb{R})$

i.e. a two-parameter family of SO(3)-invariant connections in P.

Next we observe that $\mathscr{g} = \mathrm{so}(3) = \mathscr{m} \oplus \mathscr{h}$, where $\mathscr{m}$ is the subspace spanned by $\{X_1, X_2\}$, and that $\mathrm{ad}_{\mathscr{g}}(H)(\mathscr{m}) = \mathscr{m}$; thus by Proposition 2.5, if ω is a connection form in P, then the $\mathscr{h}$-component ω' of $\omega|_Q$ is a connection form in Q. Also, since $\tilde{\phi}$ is SO(3)-invariant, we see from Proposition 2.7 that SO(3) also acts on $Q(S^2, \mathrm{SO}(2))$ by bundle automorphisms. Furthermore, by Proposition 2.8, ω' is SO(3)-invariant.

Since the reference point $(\mathbf{n}, I)$ lies in Q, if ω is an SO(3)-invariant connection in P, corresponding to Λ given by (4.3), then there is a unique projected connection ω' in Q, corresponding to the map

$$\Lambda':\ \mathrm{so}(3) \to \mathscr{h},$$
$$X \mapsto \begin{pmatrix} 0 & \alpha_3 & 0 \\ -\alpha_3 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Note that ω is reducible to the connection ω' in Q if and only if Λ is already $\mathscr{h}$-valued; i.e. $\Lambda_{11} = \Lambda_{22} = 0$.

Clearly the bundle $Q(S^2, \mathrm{SO}(2))$ can be identified with the bundle $S^3/\mathbb{Z}_2(S^2, \mathrm{SO}(2))$, since Q can be identified with SO(3) under the map $(\mathbf{r}, a) \to a$, and $\mathrm{SO}(3) \simeq \mathrm{SU}(2)/\mathbb{Z}_2 \simeq S^3/\mathbb{Z}_2$. Thus we can identify the projected connection ω' in Q with the connection ω_2 in $S^3/\mathbb{Z}_2$ described in Sec. 3: ω' corresponds to a magnetic monopole of strength $g = (\hbar c/e)$.[2,15]

ACKNOWLEDGMENTS

I should like to thank Professor J. F. Cornwell for many useful discussions and for reading the manuscript. I also thank Dr. L. H. Ryder for his interest.

This work was supported by an Overseas Scholarship from the Royal Commission for the Exhibition of 1851, and by a George Murray Scholarship from the University of Adelaide.

[1]W. Drechsler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories, Lecture Notes in Physics*, Vol. 67 (Springer-Verlag, Berlin, 1977).
[2]A. Trautman, Czech. J. Phys. B **29**, 107–116 (1979).
[3]J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931–942 (1979).
[4]J. Harnad, S. Shnider, and L. Vinet, Preprint CRMA-899, Université de Montréal, 1979.
[5]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, Vol. I (Interscience, New York, 1963).
[6]J. Madore, Commun. Math. Phys. **56**, 115–123 (1977).
[7]T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845–3857 (1975).
[8]A. M. Polyakov, JETP Lett. **20**, 194–5 (1974).
[9]G. 't Hooft, Nucl. Phys. B **79**, 276–84 (1974).
[10]Y. Kosmann-Schwarzbach, Publ. Mat. l'Univ. de Paris **7**, 55–72 (1978).
[11]R. S. Palais and T. E. Stewart, Amer. J. Math. **83**, 623–644 (1961).
[12]W. Greub, S. Halperin and R. Vanstone, *Connections, Curvature, and Cohomology*, Vol. II (Academic, New York, London, 1973).
[13]Pham Mau Quan, *Introduction à la Géometrie des Variétés Différentiables*, (Monographes Universitaires de Mathématiques No. 29) (Dunod, Paris, 1969).
[14]J. H. Rawnsley, Comm. Dublin Inst. for Adv. Studies, Series A, No. 25.
[15]A. Trautman, Int. J. Theor. Phys. **16**, 561–5 (1977).
[16]L. H. Ryder, J. Phys. A **13**, 437–447 (1980).
[17]I. M. Gel'fand, R. A. Minlos and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, Oxford, 1963).
[18]In fact it is noted in Ref. 15 that this connection is invariant under the action of U(2); however, since in this case the isotropy group J is isomorphic to $\mathrm{U}(1) \times \mathrm{U}(1)$, we see that the extra U(1)-factor ultimately makes no difference when we apply Theorem 2.1, and so SU(2)-invariance is equivalent to U(2)-invariance.
[19]F. Brickell and R. S. Clark, *Differentiable Manifolds* (Van Nostrand, London, 1970).
[20]Y. S. Wu and H.-Y. Kuo, Sci. Sin. **20**, 186–196 (1977).
[21]M. Honda, Prog. Theor. Phys. **61**, 1255–59 (1979).

Symplectic formulation of relativistic quantum mechanics

Y. S. Kim

Center for Theoretical Physics, Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742

Marilyn E. Noz

Department of Radiology, New York University, New York, New York 10016

(Received 19 August 1980; accepted for publication 16 June 1981)

It is shown that the covariant harmonic oscillator formalism in the light-cone coordinate system discussed in previous papers is a realization of the symplectic group. It is shown in particular that the Lorentz transformation of the wave function along a given direction corresponds to a one-parameter subgroup of Sp(4). The diagonal form in the light-cone coordinate system is discussed in detail. The oscillator formalism is known to represent the Poincaré group for relativistic extended hadrons, while serving as a simple calculational device for basic high-energy hadronic phenomena. Likewise, the symplectic formulation given in the present paper may serve as the basic spacetime/momentum-energy symmetry for a relativistic quantum mechanics of bound-state quarks.

PACS numbers: 11.30.Cp, 03.30. + p, 03.65.Fd

I. INTRODUCTION

From a mathematical standpoint, special relativity is the physics of Lorentz transformation, and quantum mechanics is the physics of Fourier transformation.[1] It is easy to see, if not well known, that the Lorentz boost is a symplectic transformation in the plane of longitudinal and timelike coordinates. In Fourier transformation, the width of the momentum distribution is inversely proportional to that of the spatial distribution. This is also a transformation property of the symplectic group. Thus, it is not unreasonable to suspect that the natural language of relativistic quantum mechanics is the symplectic group.

There have been many attempts to construct a quantum mechanics based on wavefunctions which can accommodate special relativity. The present paper is based on our own physical prejudice that Dirac's form of relativistic dynamics has been most fruitful and is most promising.[2] Dirac's prescription is to construct spacetime representations of the Poincaré group subject to a covariant constraint condition which reduces the four-dimensional Minkowskian spacetime into a three-dimensional Euclidean space in which nonrelativistic quantum mechanics is valid.

In our previous papers,[3] we have shown that the covariant harmonic oscillator formalism can serve as a solution of Dirac's "Poisson bracket" equations for relativistic quantum mechanics. It has been shown also that the oscillator model can explain some of the basic features observed in high-energy hadronic physics, including the mass spectrum,[4] the proton form factor,[5] the parton phenomenon,[6] and the jet phenomenon.[7]

In addition, the harmonic oscillator has the basic advantage of being mathematically simple and precise in all branches of physics. For this reason, the oscillators played a decisive role in the 1920's when the present form of nonrelativistic quantum mechanics was developed. We therefore have reason to believe that the oscillator formalism will again play an important role in the development of relativistic quantum mechanics, in conjunction with our efforts to explain basic high-energy hadronic phenomena in the relativistic quark model.

In spite of what we said above, the purpose of the present paper is mathematical. We are interested in translating the basic features of the existing relativistic oscillator formalism into the language of the symplectic group. There is no shortage of papers in the literature on the symplectic group applied to nonrelativistic harmonic oscillators,[8] and it is not our intention to add anything new to the existing treatment of the nonrelativistic system. The point of this paper is to see whether special relativity can be integrated into the symplectic formulation of nonrelativistic harmonic oscillators.

In Sec. II, we present the covariant harmonic oscillator formalism in a form suitable for the mathematical development of the present paper. In Sec. III, the symplectic nature of the Lorentz boost is discussed. Section IV deals with the symplectic nature of quantum mechanics. In Sec. V, we show that the covariant harmonic oscillator formalism is essentially a realization of the Sp(4) group, and that the Lorentz transformation of the oscillator wave function corresponds to a one-parameter subgroup of Sp(4).

II. FORMULATION OF THE PROBLEM

Our starting point is Dirac's 1949 paper on forms of relativistic dynamics for atoms.[2] The word "atom" in modern language means a bound state of quarks inside a hadron. For simplicity, let us consider here a hadron consisting of two quarks. If x_1 and x_2 are the spacetime coordinates for the two quarks, the usual procedure is to use the variables

$$X = (x_1 + x_2)/2, \qquad x = (x_1 - x_2)/2\sqrt{2}, \tag{1}$$

where X is the spacetime coordinate for the hadron and x represents the spacetime separation between the quarks.

If we assume that the quarks are bound together by a harmonic oscillator force of unit strength, it is possible to construct representations of the Poincaré group for relativistic extended hadrons exhibiting the basic high-energy features.[9] The wavefunctions diagonal in the Casimir operators take the form

$$\phi(X,x) = \psi(x,P)\exp[-iP\cdot X], \tag{2}$$

where P is the 4-momentum of the hadron and $\psi(x,P)$ is the internal wave function describing the motion of the quarks inside the hadron which satisfies the differential equation

$$\tfrac{1}{2}[(\partial_\mu)^2 - x_\mu{}^2]\psi(x,P) = (n+1)\Psi(x,P), \tag{3}$$

with the subsidiary condition

$$P^\mu a_\mu{}^\dagger\psi(x,P) = 0, \tag{4}$$

where

$$a_\mu{}^\dagger = x_\mu + \partial/\partial x^\mu.$$

It was noted in Ref. 10 that the most convenient coordinate system for constructing the desired representation of the Poincaré group is the Lorentz frame which moves with the hadron. Without loss of generality, we can assume that the hadron moves along the z direction with velocity parameter β. Then the moving coordinate variables are

$$\begin{aligned} &x' = x, \quad y' = y,\\ &z' = (z - \beta t)/(1-\beta^2)^{1/2},\\ &t' = (t - \beta t)/(1-\beta^2)^{1/2}. \end{aligned} \tag{5}$$

The convenient feature of the harmonic oscillator is that the transverse variables can be separated and can be left out throughout the discussion of the Lorentz boost. Thus the only relevant factor in the wave function is

$$\psi_\beta{}^n = (\pi 2^n n!)^{-1/2} H_n(z')\exp[-(z'^2 + t'^2)/2]. \tag{6}$$

The above form does not contain excitations along the t' direction due to the subsidiary condition given in Eq. (4).

We have also discussed in Ref. 10 the possibility of writing the above form using the light-cone coordinate system.[2] The light-cone variables are

$$\begin{aligned} u &= (t+z)/\sqrt{2},\\ v &= (t-z)/\sqrt{2}. \end{aligned} \tag{7}$$

In terms of these variables, the wavefunction of Eq. (6) can be written as

$$\psi_\beta{}^n(u,v) = (\tfrac{1}{2})^n(1/\pi n!)^{1/2}\exp[-(u'^2+v'^2)/2] \times\Big[\sum_{m=0}^{n}\binom{n}{m}H_{n-m}(u')H_m(-v')\Big], \tag{8}$$

where u' and v' are derived from t' and z' according to Eq. (7). In both Eqs. (6) and (8), the Gaussian forms are diagonal. The expression in the light-cone coordinate system is more complicated than the forms given in Eq. (6), especially for excited states. In this paper, we would like to point out that this light-cone wave function serves as the starting point for a more satisfying mathematics based on the symplectic group.

The starting equation for the oscillator formalism is the hyperbolic differential equation given in Eq. (3). However, due to the constraint of Eq. (4), the physical solutions can be constructed from various elliptic differential equations with compact support. In particular, the wave function of Eq. (8) satisfies the differential equation

$$\begin{aligned} &\tfrac{1}{2}[-(\partial/\partial u')^2 - (\partial/\partial v')^2 + u'^2 + v'^2]\psi_\beta{}^n(u,v)\\ &\quad = (n+1)\Psi_\beta{}^n(u,v), \end{aligned} \tag{9}$$

with

$$\begin{aligned} \partial/\partial u' &= (1/\sqrt{2})(\partial/\partial t' + \partial/\partial z'),\\ \partial/\partial v' &= (1/\sqrt{2})(\partial/\partial t' - \partial/\partial z'). \end{aligned} \tag{10}$$

The above differential equation is a harmonic oscillator equation in a two-dimensional Euclidean space, and is separable in the u' and v' variables. The wave funtion given in Eq. (8) is a linear combination of the u' and v' solutions whose total eigenvalue is $(n+1)$. We shall use the above equation to translate the existing oscillator formalism into the language of the symplectic group.

III. SYMPLECTIC NATURE OF THE LORENTZ BOOST

We need three rotation and three boost operators in order to close the Lorentz group. However, the actual transformation is either a rotation around or a boost along a given direction. Two rotations can generate the third one. Therefore, the Lorentz transformation matrix can be constructed from two rotation operators and *one* boost operator. As in the case of Sec. II, we shall consider here the boost along the z axis, and show that this is a symplectic transformation.

The boost matrix which transforms u and v into u' and v' takes the form

$$B(\eta) = \begin{pmatrix} e^{+\eta} & 0\\ 0 & e^{-\eta}\end{pmatrix}. \tag{11}$$

It is to be applied to the column vector $\binom{u}{v}$, where

$$\sinh\eta = \beta/(1-\beta^2)^{1/2}. \tag{12}$$

With this diagonal form, it is easy to see

$$g_{ij}B_{ik}B_{jm} = g_{km}, \tag{13}$$

where

$$[g_{ij}] = \begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}. \tag{14}$$

The above $[g_{ij}]$ matrix is the metric tensor for the Sp(2) group, and commutes with the rotation matrix

$$R(\theta) = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}. \tag{15}$$

Therefore, all matrices of the form

$$B'(\eta) = R(\theta)B(\eta)R^{-1}(\theta) \tag{16}$$

will satisfy the symplectic condition given in Eq. (13). Since we obtain the light-cone coordinate system by rotating the zt Cartesian coordinate by 45°, the familiar Lorentz boost of the form given in Eq. (5) is also a symplectic transformation.

Let us next consider the quantities which remain invariant under this symplectic transformation. For this purpose, let us define the vectors

$$\mathbf{u} = \begin{pmatrix} u\\ 0\end{pmatrix} \quad\text{and}\quad \mathbf{v} = \begin{pmatrix} 0\\ v\end{pmatrix}. \tag{17}$$

Then we can consider the "cross product" of these two vectors:

$$\mathbf{A} = \mathbf{u} \times \mathbf{v}. \tag{18}$$

The above cross product remains invariant under the boost transformation given in Eq. (11). This invariance is illustrated in Fig. 1. The area of the rectangle **A** has its direction or sign depending on which side is measured first. The magnitude of the area is therefore a boost-invariant quantity:

$$A = |\mathbf{A}| + 4uv = 2(t+z)(t-z) \\ = 2(t^2 - z^2). \tag{19}$$

In the above discussion, we have been dealing with the invariant quantity formed from one spacetime coordinate (t,z). In order to explore fully the symplectic properties of the Lorentz boost, we should consider also the invariants formed from two different 4-vectors: (t_a, z_a) and (t_b, z_b), and corresponding u and v variables. Then the boost matrix of Eq. (11) assures us that both $u_a v_b$ and $v_a u_b$ remain invariant. This means that

$$J_{ab} = (u_a v_b + v_a u_b)$$

and (20)

$$K_{ab} = (u_a v_b - v_a u_b)$$

are also boost invariant. J_{ab} is symmetric under the interchange of a and b, and K_{ab} is antisymmetric. In terms of the t and z variables,

$$J_{ab} = t_a t_b - z_a z_b,$$

(21)

$$K_{ab} = t_a z_b - z_a t_b.$$

The fact that J_{ab} is boost invariant is well known. However, the invariance of K_{ab} is not yet widely known. It is one of the symplectic properties of the Lorentz boost. We can check the boost invariance of K_{ab} using the basic Lorentz transformation formulas given in Eq. (5).

IV. SYMPLECTIC NATURE OF QUANTUM MECHANICS

Symplectic properties of quantum mechanics, especially those of harmonic oscillators, have been completely and thoroughly discussed in the literature, and we are not trying to add anything new to the existing treatment.[8] The purpose of this section is to discuss some of the known symplectic properties of nonrelativistic oscillators which are relevant to the problem of making the oscillator system covariant.

The symplectic nature of quantum mechanics manifests itself in the conservation of the area of the phase space under the scale transformations of the coordinate and its conjugate momentum variable. If the coordinate variable is elongated, then its conjugate momentum variable is contracted by the same amount. As is well known, this feature is built into the Fourier transformation.

We shall discuss here the symplectic property of relativistic oscillators using the differential equation given in Eq. (9). This equation can be written in a two-dimensional quadratic form

$$\tfrac{1}{2}[q_u'^2 + q_v'^2 + u'^2 + v'^2]\psi_\beta{}^n(u,v) \\ = (n+1)\psi_\beta{}^n(u,v), \tag{22}$$

with

$$q_u' = i\partial/\partial u' = (q_0' - q_z')/\sqrt{2},$$

(23)

$$q_v' = i\partial/\partial v' = (q_0' + q_z')/\sqrt{2},$$

where

$$q_0' = i\partial/\partial t', \quad q_z' = -i\partial/\partial z'.$$

The variables q_0 and q_z represent the energy and the z-component momentum differences, respectively. If we define the momentum–energy wavefunction as

$$\phi_\beta{}^n(q_u,q_v) = (1/2\pi)\int dudv\, \psi_\beta{}^n(u,v) \exp[-i(q_u u + q_v v)], \tag{24}$$

then its mathematical form is identical to that of the spacetime wave function given in Eq. (8).

The phase space of the above oscillator system remains invariant under the following symplectic transformations:

$$\begin{pmatrix} u' \\ q_u' \end{pmatrix} = \begin{pmatrix} e^{+\eta} & 0 \\ 0 & e^{-\eta} \end{pmatrix} \begin{pmatrix} u \\ q_u \end{pmatrix},$$

(25)

$$\begin{pmatrix} v' \\ q_v' \end{pmatrix} = \begin{pmatrix} e^{-\eta} & 0 \\ 0 & e^{+\eta} \end{pmatrix} \begin{pmatrix} v \\ q_v \end{pmatrix}.$$

The parameter η in the above expression can be any real number, and can therefore be the boost parameter defined in Eq. (12). The metric in both cases takes the form given in Eq. (14).

As we rotate u and v using the matrix given in Eq. (16), q_u and q_v become rotated in the same manner. This rotation leaves the quantity

$$(q_u u + q_v v) = (q_0 t - q_z z) \tag{26}$$

invariant.

V. SYMPLECTIC NATURE OF RELATIVISTIC QUANTUM MECHANICS

In Secs. III and IV, we discussed the symplectic natures of Lorentz boosts and quantum mechanics respectively. In

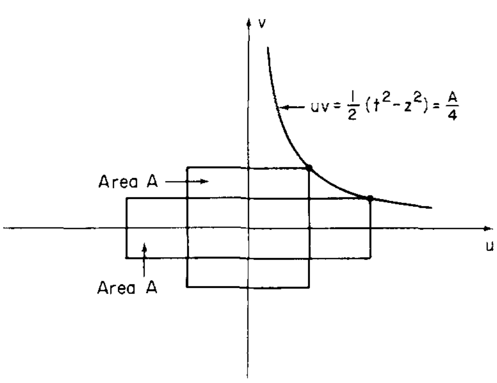


FIG. 1. The geometry of Lorentz boost. As is well known, the hyperbola tells us that $(t^2 - z^2)$ is boost-invariant. This condition tells us also that the area of the rectangle inscribed by this hyperbola remains invariant under the boost.

order to give a symplectic formulation of the combined effect of quantum mechanics and relativity, we consider the four-dimensional space (u, v, q_u, q_v). Then the Lorentz boost matrix becomes

$$B(\eta) = \begin{pmatrix} e^{+\eta} & 0 & 0 & 0 \\ 0 & e^{-\eta} & 0 & 0 \\ 0 & 0 & e^{-\eta} & 0 \\ 0 & 0 & 0 & e^{+\eta} \end{pmatrix}. \tag{27}$$

This transformation leaves invariant the symplectic metric for Lorentz boost:

$$[g_{ij}] = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{28}$$

The transformation matrix of Eq. (27) also leaves invariant the symplectic metric for Fourier transformation:

$$[h_{ij}] = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}. \tag{29}$$

We can illustrate these transformation properties using spacetime and momentum-energy diagrams. When the hadron is at rest, the wave function of Eq. (6) or (8) is localized within a circle around the origin in the *zt* or *uv* plane. This localization region becomes Lorentz-deformed according to the geometry given in Fig. 1. The momentum-energy wave-function has the same deformation property as that of the spacetime wavefunction. Figure 2 describes these deformation properties.

The physics of Fig. 2 has been extensively discussed in the literature.[11] It has been shown[6] in particular that the Lorentz deformation property described in this figure leads to an explanation of the peculiarities in Feynman's parton picture,[12] and to an accurate calculation of the proton structure function. This deformation property explains also the proton form factor behavior[5] and the origin of the hadronic jet phenomenon.[7] As is well known, the hadronic mass spectra can be understood in terms of the SU(6)⊗O(3) scheme.[4] There seems to be a misunderstanding that the O(3) in this scheme is inherently nonrelativistic. This O(3) group is the little group of the Poincaré group for massive hadrons,[9] and is therefore relativistic.

As for the mathematics, it is important to note that the area of the ellipse remains invariant under Lorentz boosts. This has been translated into the matrix language of Eq. (27). It is also interesting to note that the major axis of the spacetime ellipse is conjugate to the minor axis of the momentum-energy ellipse, according to Eq. (23), and vice versa. For this reason, the volume of the phase space remains Lorentz-invariant. This symplectic feature is represented by Eq. (25).

When we make a Lorentz boost along the *z* direction, the transformation matrix should satisfy the metric conditions for both $[g_{ij}]$ and $[h_{ij}]$. This corresponds to a one-parameter subgroup of the Sp(4) group.

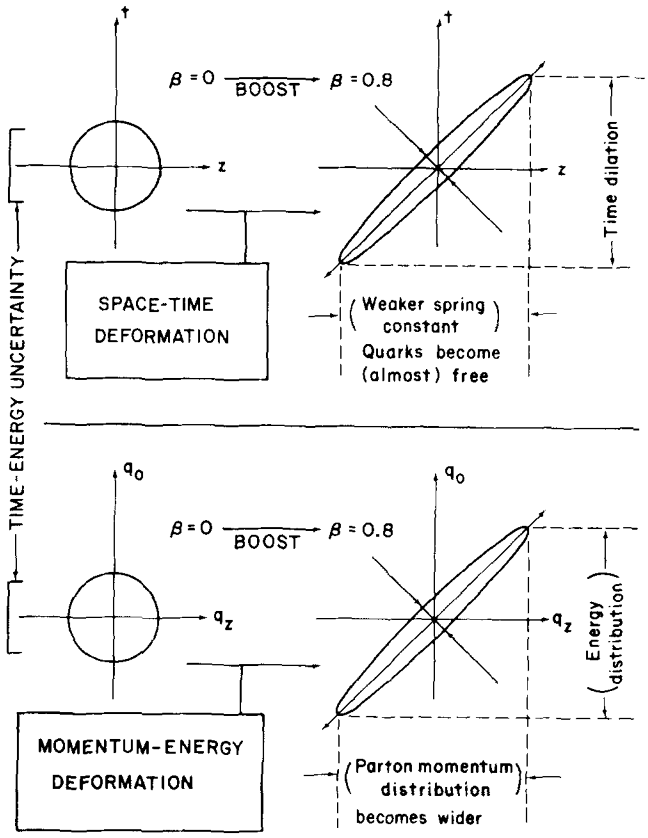


FIG. 2. The symplectic transformation in which the hadron is Lorentz-deformed according to the spacetime geometry given in Fig. 1. As the hadron moves very rapidly, both the spatial and momentum distributions become spread wide along their respective longitudinal axes. It is shown in Refs. 6 and 11 that this mechanism produces the pecularities of Feynman's parton picture[12] which is universally observed in high-energy laboratories.

VI. CONCLUDING REMARKS

We have discussed in this paper the possibility of using the language of the symplectic group for describing covariantly localized quantum mechanical wave functions. It is shown that the Lorentz boost of quantum mechanics can be described simply by *one matrix* given in Eq. (27), which is a representation of a one-parameter subgroup of Sp(4).

It is not uncommon for new physical theories to be preceded by a specific solution such as that of the harmonic oscillator. If the development of quantum mechanics in the 1920's is a lesson to us, the generalization of the specific solution takes the form of translating it into a matrix or group theoretical language. We have translated here the covariant oscillator formalism, which is compatible both with special relativity and with quantum mechanics, into a language of the symplectic group.

Since physics is an experimental science, the value of the mathematics presented in this Journal is not necessarily dictated by its length or complexity. It is determined by its ability to explain what we see in the real world. The matrix given in Eq. (27) is simple enough for us to suspect that the basic spacetime/momentum-energy symmetry governing the hadronic phenomena discussed in Refs. 4–7 is that of the symplectic group.

Our discussion in this paper was limited to Lorentz transformations along one given axis. A more general discussion of the spacetime symmetry should include the ten independent transformations associated with the generators of the Poincaré group, and this discussion is given in Ref. 9. The present paper deals with the specific additional properties associated with the Lorentz boost along a given direction.[13]

ACKNOWLEDGMENTS

We would like to thank K. B. Wolf and P. Winternitz for pointing out the importance of the symplectic group in dealing with harmonic oscillator problems. We benefited also from discussions with A. Dragt and D. Fivel.

[1]This remark was made by E. P. Wigner in one of his lectures on scattering theory given at Princeton University during the academic year 1958–59.
[2]P. A. M. Dirac, Rev. Mod. Phys. **21**, 392 (1949).
[3]For an interpretation of Dirac's "instant form" quantum mechanics in terms of the covariant oscillator formalism, see D. Han and Y. S. Kim, Prog. Theor. Phys. **64**, 1852 (1980); Am J. Phys. (to be published).
[4]For a review paper on the hadronic mass spectra, see P. E. Hussar, Y. S. Kim, and M. E. Noz, Am. J. Phys. **48**, 1043 (1980).
[5]K. Fujimura, T. Kobayashi, M. Namiki, Prog. Theor. Phys. **43**, 73 (1970); R. G. Lipes, Phys. Rev. D **5**, 2849 (1972); Y. S. Kim and M. E. Noz, Phys. Rev. D **8**, 3521 (1973).
[6]Y. S. Kim and M. E. Noz, Phys. Rev. D **15**, 335 (1977); P. E. Hussar, Phys. Rev. D **23**, 1781 (1981).
[7]T. Kitazoe and S. Hama, Phys. Rev. D **19**, 2006 (1979); Y. S. Kim, M. E. Noz, and S. H. Oh, Found. Phys. **9**, 947 (1979); T. Kitazoe and T. Morii, Phys. Rev. D **21**, 685 (1980) and Nucl. Phys. B **164**, 76 (1980); S. H. Oh, Y. S. Kim, and M. E. Noz, Found. Phys. **10**, 635 (1980).
[8]M. Moshinsky and P. Winternitz, J. Math. Phys. **21**, 1667 (1980). This paper contains a complete list of references on this subject.
[9]Y. S. Kim, M. E. Noz, and S. H. Oh, J. Math. Phys. **20**, 1341 (1979).
[10]Y. S. Kim, M. E. Noz, and S. H. Oh, J. Math. Phys. **21**, 1224 (1980).
[11]For a detailed physical interpretation of Fig. 2, see Y. S. Kim and M. E. Noz, Found. Phys. **9**, 375 (1979). See also Ref. 6.
[12]R. P. Feynman, in *High Energy Collisions*, Proceedings of the Third International Conference, Stony Brook, edited by C. N. Yang *et al.* (Gordon and Breach, New York, 1969). See also J. D. Bjorken and E. A. Paschos, Phys. Rev. **185**, 1975 (1969).
[13]There are many papers in the literature on the Lorentz, Poincaré, and symplectic groups for spacetime symmetries. However, unlike the present paper, most of the existing papers deal with larger groups from which the basic spacetime symmetry is derivable. See, for instance, L. P. Staunton, J. Math. Phys. **19**, 1471 (1979) or H. M. Ruck, Phys. Rev. D **20**, 2027 (1979).

Boson, para-boson, and boson-fermion representations of some graded Lie algebras

R. Jagannathan[a)]
Department of Physics, St. Joseph's College, Tiruchirapalli-620 002, India
R. Vasudevan[b)]
Institute of Mathematical Sciences, "Matscience", Madras-600 020, India
Physics Department, Southern Illinois University, Carbondale, Illinois 62901

(Received 15 June 1979; accepted for publication 18 July 1980)

Some graded Lie algebras with the Lie algebras so(3), so(2,1), so(4), so(3,1), and so(2,2) as their Bose sectors are realized in terms of bosons, para-bosons and certain bilinear combinations of bosons and fermions.

PACS numbers: 11.30.Pb, 02.20.Sv

1. INTRODUCTION

This paper studies first certain types of boson and para-boson representations of some graded Lie algebras (GLA's) with the Lie algebras (LA's), so(3) and so(2,1) as their Bose sectors in a unified manner and then considers similarly the same types of representations of GLA's with so(4), so(3,1), and so(2,2) as their Bose sectors. Further it is shown that a method of description of general linear GLA's in terms of a mixture of bosonic and fermionic operators due to Freund and Kaplansky[1] can be utilized for a similar representation of any GLA. This fact is used for constructing a boson-fermion representation of the GLA, Gsu(2) with su(2) $\sim$ so(3) as Bose sector, studied in detail by Pais and Rittenberg.[2] Based on this result such representations in terms of bilinear combinations of bosons and fermions are constructed also for other GLA's, studied in the paper, with so(2,1), so(4), so(3,1) and so(2,2) as their Bose sectors. All the GLA's in this paper are to be considered over real numbers and they belong to the class of orthosymplectic GLA's after complexification.[1]

The para-Bose creation and annihilation operators $\{\beta_j^{(p)\dagger}, \beta_j^{(p)} | j = 1,2,\ldots, n\}$ belonging to statistics of order p obey the relations given by Green[3] as

$$[\beta_j^{(p)}, \{\beta_k^{(p)\dagger}, \beta_l^{(p)}\}] = 2\delta_{jk}\, \beta_l^{(p)},$$
$$[\beta_j^{(p)}, \{\beta_k^{(p)}, \beta_l^{(p)}\}] = 0,$$
$$\beta_j^{(p)}\beta_k^{(p)\dagger}|0\rangle = p\delta_{jk}|0\rangle, j,k,l = 1,2,\ldots,n. \tag{1}$$

Jordan *et al.*[4] have given explicitly the matrix representation of single para-Bose operator β obeying

$$[\beta, \{\beta^\dagger, \beta\}] = [\beta^2, \beta^\dagger] = 2\beta \tag{2}$$

using the representation of the generators of the LA so(2,1) in terms of β and $\beta^\dagger$. Using such a para-boson realization of so(2,1) it has been shown by us[5] earlier that $(\beta, \beta^\dagger)$ and the generators of so(2,1) expressed in terms of $(\beta, \beta^\dagger)$ generate a GLA very similar to the GLA, Gsu(2) of Pais and Rittenberg.[2] This paper extends such a study of para-Bose description also to certain other GLA's with the LA's so(3), so(4), so(3,1), and so(2,2) as their Bose sectors.

We have shown earlier[5] that using the creation and annihilation operators $\{b_j^+, b_j | j = 1,2,\ldots,np + [p/2]\}$ belonging to a single Bose field and satisfying the relations

$$[b_j, b_k^+] = \delta_{jk}, \quad [b_j, b_k] = [b_j^+, b_k^+] = 0,$$
$$j,k = 1,2,\ldots,np + [p/2], \tag{3}$$

one can construct a set of operators

$$\begin{aligned}\beta_j^{(p)} &= \Bigg\{\sum_{r=1}^{[p/2]} i^{r-1}\Bigg[\prod_{s=1}^{r-1} \exp\Big\{\Big(\frac{\pi}{2}\Big)^{1/2}[(1+i)b_s^+ - (1-i)\,b_s]\Big\}\Bigg] \\ &\times\Bigg[\Big\{\exp\Big[i\Big(\frac{\pi}{2}\Big)^{1/2}(b_r^+ + b_r)\Big]\, b_{[p/2]+(2r-2)n+j}\Big\} \\ &+ \Big\{\exp\Big[\Big(\frac{\pi}{2}\Big)^{1/2}(b_r^+ - b_r)\Big]\, b_{[p/2]+(2r-1)n+j}\Big\}\Bigg]\Bigg\} \\ &+ \tfrac{1}{2}i^{[p/2]}[1 - (\exp(ip\,\pi)], \\ &\times\Bigg[\prod_{s=1}^{[p/2]} \exp\Big\{\Big(\frac{\pi}{2}\Big)^{1/2}[(1+i)b_s^+ - (1-i)\,b_s]\Big\}\Bigg] \\ &\times b_{(2n+1)[p/2]+j}, j = 1,2,..,n,\end{aligned} \tag{4}$$

where $[p/2]$ stands for the integer part of $p/2$. In view of this representation of a para-Bose field in terms of a single Bose field and also due to the fact that para-bosons of order $p = 1$ are just bosons, the para-boson representations of the GLA's considered below provide also a class of boson representations.

In Sec. 2 the general structure of a GLA is recalled. The para-boson representations of the above mentioned GLA's are attained in Secs. 3 and 4. In Sec. 5 the method of description of general linear GLA's in terms of bilinear expressions of bosonic and fermionic operators due to Freund and Kaplansky[1] is utilized for constructing the boson-fermion representation of the GLA's considered in earlier sections. The paper concludes in Sec. 6 with some remarks including an observation on the C-theorem of Pais and Rittenberg.[2]

2. STRUCTURE OF A GLA

In the specification of structure of GLA's let us follow Pais and Rittenberg[2] and Scheunert *et al.*[6] Then for any GLA the basic commutation and anticommutation relations among the even generators $\{Q_m | m = 1,2,\ldots, D\}$ and the odd generators $\{V_\alpha | \alpha = 1,2,\ldots,d\}$ are given by

$$[Q_m, Q_n] = f^p_{mn}\, Q_p, \tag{5a}$$
$$[Q_m, V_\beta] = F^\alpha_{m\beta}\, V_\alpha, \tag{5b}$$

[a)]Present address: Institute of Mathematical Sciences, "Matscience," Madras-600 020, India.
[b)]Permanent address.

$$\{V_\alpha, V_\beta\} = A^{\,m}_{\alpha\beta}\, Q_m, \tag{5c}$$
$$m,n,p = 1,2,\ldots,D, \quad \alpha,\beta = 1,2,\ldots,d.$$

Throughout the paper the normal convention of summation over repeated indices is assumed and it will be clear from the context as in Eqs. (5a)–(5c) above. The even generators $\{Q_m\}$ spanning a LA constitute the so-called Bose sector of the GLA. GLA's over complex numbers having the LA's $\mathrm{sp}(2p) \oplus \mathrm{o}(m)$ and $\mathrm{sl}(m) \oplus \mathrm{sl}(n) \oplus \mathrm{gl}(1)$ as their Bose sectors are called orthosymplectic GLA and special linear GLA and denoted as $\mathrm{osp}(2p,m)$ and $\mathrm{spl}(m,n)$ respectively. As Freund and Kaplansky[1] have stated, the GLA's thus far encountered in physics are, after complexification, of either orthosymplectic type or special linear type or else they are Inönu–Wigner contractions of these GLA's. All the GLA's considered in this paper are of orthosymplectic type after complexification. The special linear and the orthosymplectic GLA's have been shown to be simple in the sense that they have no invariant subalgebras. More details on the mathematics of GLA's and the associated supersymmetries of physics can be found in Refs. 1,2 and 6–13.

3. GLA'S WITH so(3) AND so(2,1) AS BOSE SECTORS

Here we shall realize in terms of para-bosons certain GLA's with the LA's so(3) and so(2,1) as their Bose sectors. Since complexification of so(3) and so(2,1) lead to the same LA over complex numbers[14,15] isomorphic to sp(2), these GLA's with so(3) and so(2,1) as Bose sectors are of the type osp(2,1) in their complex forms.

If we define

$$\begin{aligned} J_1 &= \tfrac{1}{4}(\beta\beta - \beta^+\beta^+), \\ J_2 &= \tfrac{1}{4}i(\beta\beta + \beta^+\beta^+), \\ J_3 &= \tfrac{1}{4}(\beta\beta^+ + \beta^+\beta), \end{aligned} \tag{6}$$

where β and β^+ are boson operators or para-Bose operators of any order, obeying Eq. (2), then we have

$$[J_j, J_k] = i\epsilon_{jkl}J_l, \quad j,k,l = 1,2,3. \tag{7}$$

Now it is straigtforward to see that the elements defined by

$$\begin{aligned} Q_m &= t_m J_m, \quad m = 1,2,3, \\ V_1 &= s_1\beta^+, \quad V_2 = s_2\beta, \end{aligned} \tag{8}$$

generate a GLA as per Eqs. (5a)–(5c). The structure constants of this GLA are given by

$$f^{\,p}_{mn} = i\epsilon_{mnp}\,(t_m t_n/t_p); \tag{9}$$

$$\begin{aligned} [F^{\alpha}_{1\beta}] &= (t_1/4s_1s_2)\{(s_1^2 + s_2^2)\,\tau_1 - i(s_1^2 - s_2^2)\,\tau_2\}, \\ [F^{\alpha}_{2\beta}] &= (it_2/4s_1s_2)\{(s_1^2 - s_2^2)\,\tau_1 - i(s_1^2 + s_2^2)\,\tau_2\}, \\ [F^{\alpha}_{3\beta}] &= (t_3/2)\,\tau_3; \end{aligned} \tag{10}$$

$$\begin{aligned} [A^{\,1}_{\alpha\beta}] &= (-i8s_1s_2/t_1^2)\,\tau_2\,F_1, \\ [A^{\,2}_{\alpha\beta}] &= (-i8s_1s_2/t_2^2)\,\tau_2 F_2, \\ [A^{\,3}_{\alpha\beta}] &= (-i8s_1s_2/t_3^2)\,\tau_2\,F_3, \end{aligned} \tag{11}$$

$$m,n,p = 1,2,3, \quad \alpha,\beta = 1,2,$$

where

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{12}$$

Here and everywhere below, the rows and columns of the matrices are labelled by the upper or the first index and the lower or the second index respectively, e.g., α and β respectively in Eqs. (10) and (11).

Following Pais and Rittenberg[2] let the metric tensors $\{h_{mn} | m,n = 1,2,\ldots,D\}$ of the underlying LA and $\{g_{kl} | k,l = 1,2,\ldots,D+d\}$ of the entire GLA be defined by

$$\begin{aligned} h_{mn} &= h_{nm} = f^{\,p}_{mq} f^{\,q}_{np}, \\ g_{mn} &= g_{nm} = h_{nm} - F^{\alpha}_{m\beta} F^{\beta}_{n\alpha}, \\ g_{D+\alpha,D+\beta} &= -g_{D+\beta,D+\alpha} \\ &= F^{\lambda}_{m\alpha} A^{\,m}_{\beta\lambda} - F^{\lambda}_{m\beta} A^{\,m}_{\alpha\lambda}, \\ g_{D+d,m} &= g_{m,D+d} = 0, \\ m,n,p,q &= 1,2,\ldots,D, \quad \alpha,\beta,\lambda = 1,2,\ldots,d. \end{aligned} \tag{13}$$

Then in the present case we have that

$$[h_{mn}] = 2\begin{pmatrix} t_1^2 & 0 & 0 \\ 0 & t_2^2 & 0 \\ 0 & 0 & t_3^2 \end{pmatrix}, \quad m,n, = 1,2,3, \tag{14}$$

and

$$[g_{kl}] = \frac{3}{2}\begin{pmatrix} t_1^2 & 0 & 0 & 0 & 0 \\ 0 & t_2^2 & 0 & 0 & 0 \\ 0 & 0 & t_3^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8s_1s_2 \\ 0 & 0 & 0 & -8s_1s_2 & 0 \end{pmatrix}, \quad k,l = 1,2,3,4,5. \tag{15}$$

When $\{t_1,t_2,t_3,t_4,t_5\}$ are all nonzero, as they should be to get the above GLA structure we always have

$$\det\|h\| \neq 0, \quad \det\|g\| \neq 0. \tag{16}$$

It should be noted from Eqs. (10) and (11) that if we have the condition

$$t_1^2 = t_2^2 = t_3^2 = t^2 \tag{17}$$

irrespective of the choice of s_1 and s_2 then the matrices $\{F\}$ and $\{A\}$ of Eqs. (10) and (11) respectively are related by

$$\begin{aligned} A^{\,m}_{\alpha\beta} &= C_{\alpha\lambda}F^{\lambda}_{m\beta}, \quad C_{\alpha\lambda} = -C_{\lambda\alpha}, \\ m &= 1,2,\ldots,D, \quad \alpha,\beta,\lambda = 1,2,\ldots,d, \end{aligned} \tag{18}$$

with $D = 3$ and $d = 2$, and

$$[C_{\alpha\lambda}] = (-i8s_1s_2/t^2)\,\tau_2, \quad \alpha,\lambda = 1,2. \tag{19}$$

If we choose

$$\begin{aligned} t_1 &= (G_{22}G_{33})^{1/2}, \quad t_2 = (G_{33}G_{11})^{1/2}, \quad t_3 = (G_{11}G_{22})^{1/2} \\ s_1 &= s_2 = i/2\sqrt{2}, \end{aligned} \tag{20}$$

then using Eqs. (9)–(11), it is found easily that the elements defined by Eqs. (8) and (20) generate a GLA with structure constants given by

$$f^{p}_{mn} = i\epsilon_{mnp} G_{pp}, \tag{21a}$$

$$F_m = \tfrac{1}{2}(G_{nn} G_{pp})^{1/2} \tau_m, \tag{21b}$$

$$A^{m} = (i/G_{nn} G_{pp}) \tau_2 F_m, \tag{21c}$$

where $m,n,p, = 1,2,3$ cyclically.

When $G_{11} = G_{22} = G_{33} = 1$, the above GLA has the LA so(3) as its Bose sector and is the same as Gsu(2) ~ Gso(3) discussed in detail by Pais and Rittenberg.[2] When we take $G_{11} = G_{22} = -G_{33} = 1$ the above GLA has so(2,1) as its Bose sector. It is evident from Eq. (20) that the condition in Eq. (17) can be satisfied in the case of so(3) and not in the case of so(2,1). Hence, as can be seen directly from Eq. (21c), while the relation in Eq. (18) exists for the GLA with the compact LA so(3) as Bose sector as is known already,[2] such a relation does not exist in the case of the GLA with the noncompact LA so(2,1) as Bose sector. The graded so(2,1) considered by us earlier[5] corresponds to the choice $\{t_1 = t_2 = -i, t_3 = 1, s_1 = s_2 = 1\}$ in the above scheme.

Thus if β is a para-Bose operator of any order, the elements

$$\begin{aligned}
Q_1 &= \tfrac{1}{4}(G_{22}G_{33})^{1/2}(\beta\beta - \beta^{+}\beta^{+}),\\
Q_2 &= \tfrac{1}{4}i(G_{33}G_{11})^{1/2}(\beta\beta + \beta^{+}\beta^{+}),\\
Q_3 &= \tfrac{1}{4}(G_{11}G_{22})^{1/2}(\beta\beta^{+} + \beta^{+}\beta),\\
V_1 &= i\beta^{+}/2\sqrt{2}, \quad V_2 = i\beta/2\sqrt{2},
\end{aligned} \tag{22}$$

represent the generators of certain GLA's with so(3) and so(2,1) as Bose sectors respectively when

$G_{11} = G_{22} = G_{33} = 1$ and $G_{11} = G_{22} = -G_{33} = 1$.

4. GLA'S WITH so(4), so(3,1) AND so(2,2) AS BOSE SECTORS

The GLA's with so(4), so(3,1), and so(2,2) as Bose sectors belong to the class of GLA's characterized as osp(2,3) after complexification since the complex extensions of all the underlying LA's in these cases lead to the same LA over complex numbers[14,15] isomorphic to sp(2) ⊕ o(3).

Now we shall construct the para-Bose representations of certain GLA's with so(4), so(3,1), and so(2,2) as their Bose sectors in a unified manner following the unified description of GLA's associated with so(3) and so(2,1) given above. To this end let us start with two commuting para-Bose operators β and β' of the same order or different orders, each obeying Eq. (2). For example, using Eq. (4) and six pairs of boson operators $\{b_i, b_i^{+} | i = 1,2,..,6\}$ we can construct two such commuting para-Bose operators, each of order 2, as

$$\begin{aligned}
\beta^{(2)} &= \{\exp[i(\pi/2)^{1/2}(b_1^{+} + b_1)]b_2\}\\
&\quad + \{\exp[(\pi/2)^{1/2}(b_1^{+} - b_1)]b_3\},\\
\beta'^{(2)} &= \{\exp[i(\pi/2)^{1/2}(b_4^{+} + b_4)]b_5\}\\
&\quad + \{\exp[(\pi/2)^{1/2}(b_4^{+} - b_4)]b_6\}.
\end{aligned} \tag{23}$$

Let us define

$$\begin{aligned}
J_1 &= \tfrac{1}{4}(\beta\beta - \beta^{+}\beta^{+}),\\
J_2 &= \tfrac{1}{4}i(\beta\beta + \beta^{+}\beta^{+}),\\
J_3 &= \tfrac{1}{4}(\beta\beta^{+} + \beta^{+}\beta),
\end{aligned} \tag{24}$$

and

$$\begin{aligned}
J_1' &= \tfrac{1}{4}(\beta'\beta' - \beta'^{+}\beta'^{+}),\\
J_2' &= \tfrac{1}{4}i(\beta'\beta' + \beta'^{+}\beta'^{+}),\\
J_3' &= \tfrac{1}{4}(\beta'\beta'^{+} + \beta'^{+}\beta'),
\end{aligned} \tag{25}$$

such

$$\begin{aligned}
[J_j, J_k] &= i\epsilon_{jkl} J_l,\\
[J_j', J_k'] &= i\epsilon_{jkl} J_l',\\
[J_j', J_k] &= 0, \quad j,k,l = 1,2,3.
\end{aligned} \tag{26}$$

Then in analogy with the usual procedure[15] of constructing the generators of so(4) from the generators of so(3) let us write

$$\begin{aligned}
M_j &= \mathscr{L}_{kl} = -\mathscr{L}_{lk} = (G_{kk}G_{ll})^{1/2}(J_j + J_j'),\\
N_j &= \mathscr{L}_{0j} = -\mathscr{L}_{j0} = (G_{00}/G_{jj})^{1/2}(G_{kk}G_{ll})\\
&\quad \times (J_j - J_j'),
\end{aligned} \tag{27}$$

where $j,k,l = 1,2,3$ cyclically. Now it is easily seen that the elements defined by Eq. (27) obey the commutation relations

$$\begin{aligned}
[\mathscr{L}_{\mu\nu}, \mathscr{L}_{\rho\sigma}] = i(&G_{\mu\rho}\mathscr{L}_{\nu\sigma} + G_{\nu\sigma}\mathscr{L}_{\mu\rho}\\
&- G_{\mu\sigma}\mathscr{L}_{\nu\rho} - G_{\nu\rho}\mathscr{L}_{\mu\sigma}),
\end{aligned} \tag{28}$$

with

$$G_{\mu\nu} = G_{\mu\mu}\delta_{\mu\nu} \tag{29}$$

for $\mu,\nu,\rho,\sigma = 0,1,2,3$ or

$$\begin{aligned}
[M_j, M_k] &= i\epsilon_{jkl} G_{ll} M_l,\\
[M_j, N_k] &= i\epsilon_{jkl} G_{kk} N_l,\\
[N_j, N_k] &= i\epsilon_{jkl} G_{00} M_l,\\
j,k,l &= 1,2,3.
\end{aligned} \tag{30}$$

It follows easily from Eqs. (24)–(30) that the elements

$$\begin{aligned}
Q_m &= \begin{cases} M_m; & m = 1,2,3\\ M_{m-3}; & m = 4,5,6, \end{cases}\\
V_1 &= (i/2\sqrt{2})\beta^{+}, \quad V_2 = (i/2\sqrt{2})\beta,
\end{aligned} \tag{31}$$

generate a GLA with the structure constants given as follows:

$$f^{p}_{mn} = \begin{cases} iG_{pp}\epsilon_{mnp}; \ m,n,p = 1,2,3,\\ -f^{p}_{nm} = iG_{n-3,n-3}\epsilon_{n,n-3,p-3}; \ m = 1,2,3, \ n,p = 4,5,6,\\ iG_{00}\epsilon_{m-3,n-3,p}; \ m,n = 4,5,6, \ p = 1,2,3. \end{cases} \tag{32}$$

$$\begin{aligned}
[F^{\alpha}_{j\beta}] &= \tfrac{1}{2}(G_{kk}G_{ll})^{1/2}\tau_j,\\
[F^{\alpha}_{j+3,\beta}] &= \tfrac{1}{2}(G_{00}/G_{jj})^{1/2}(G_{kk}G_{ll})\tau_j,
\end{aligned} \tag{33}$$

$$\begin{aligned}
[A^{j}_{\alpha\beta}] &= (i/2\, G_{kk}G_{ll})\tau_2 F_j,\\
[A^{j+3}_{\alpha\beta}] &= (iG_{jj}/2G_{00})\tau_2 F_{j+3},\\
j,k,l &= 1,2,3 \quad \text{(cyclically)}, \quad \alpha,\beta = 1,2.
\end{aligned} \tag{34}$$

The elements of the metric tensors h and g in this case become

$$h_{mm} = \begin{cases} 4G_{nn}G_{pp}; & m,n,p = 1,2,3 \text{ (cyclically)}, \\ 4G_{00}G_{m-3,\,m-3}; & m = 4,5,6, \end{cases}$$

$$h_{mn} = 0; \quad m \neq n, \quad m,n = 1,2,\ldots,6, \tag{35}$$

and

$$g_{kk} = \begin{cases} \tfrac{7}{2}G_{ll}G_{mm}; \quad k,l,m = 1,2,3 \quad \text{(cyclically)}, \\ G_{00}G_{k-3,k-3}\{4 - \tfrac{1}{2}(G_{l-3,l-3}G_{m-3,m-3}/G_{k-3,k-3})^{1/2}\}; \\ \quad k,l,m = 4,5,6, \text{ (cyclically)} \end{cases}$$

$$g_{78} = -g_{87} = -\tfrac{1}{4}\{3 + G_{11}^2G_{22}^2 + G_{22}^2G_{33}^2 + G_{33}^2G_{11}^2\},$$

$$g_{77} = g_{88} = g_{k7} = g_{k8} = g_{7k} = g_{8k} = 0, \quad g_{kl} = 0,$$

$$k \neq l, \quad k,l = 1,2,\ldots,6. \tag{36}$$

It is evident from Eqs. (28)–(30) that when we take

$$G_{00} = G_{11} = G_{22} = G_{33} = 1, \tag{37a}$$

$$-G_{00} = G_{11} = G_{22} = G_{33} = 1, \tag{37b}$$

$$-G_{00} = G_{11} = G_{22} = -G_{33} = 1, \tag{37c}$$

the elements $\{M_j, N_j \mid j = 1,2,3\}$ defined through Eqs. (24)–(27) generate respectively the LA's so(4), so(3,1) and so(2,2).[16] Thus if β and β' are two commuting para-Bose operators of the same order or different orders the elements

$$Q_1 = \tfrac{1}{4}(G_{22}G_{33})^{1/2}(\beta\beta - \beta^+\beta^+ + \beta'\beta' - \beta'^+\beta'^+),$$

$$Q_2 = \tfrac{1}{4}i(G_{33}G_{11})^{1/2}(\beta\beta + \beta^+\beta^+ + \beta'\beta' - \beta'^+\beta'^+),$$

$$Q_3 = \tfrac{1}{4}(G_{11}G_{22})^{1/2}(\beta\beta^+ + \beta^+\beta + \beta'\beta'^+ + \beta'^+\beta'),$$

$$Q_4 = \tfrac{1}{4}(G_{00}/G_{11})^{1/2}(G_{22}G_{33})(\beta\beta - \beta^+\beta^+ - \beta'\beta' + \beta'^+\beta'^+),$$

$$Q_5 = \tfrac{1}{4}i(G_{00}/G_{22})^{1/2}(G_{33}G_{11})(\beta\beta + \beta^+\beta^+ - \beta'\beta' - \beta'^+\beta'^+),$$

$$Q_6 = \tfrac{1}{4}(G_{00}/G_{33})^{1/2}(G_{11}G_{22})(\beta\beta^+ + \beta^+\beta - \beta'\beta'^+ - \beta'^+\beta'),$$

$$V_1 = (i/2\sqrt{2})\beta^+, \quad V_2 = (i/2\sqrt{2})\beta, \tag{38}$$

represent the generators of certain GLA's with the LA's so(4), so(3,1) and so(2,2) as their Bose sectors when (G_{00}, G_{11}, G_{22}, G_{33}) take values as (1,1,1,1), ($-$1,1,1,1), and ($-$1,1,1,$-$1) respectively and the corresponding structure constants are given by Eqs. (32)–(34). If we replace β by β' in the definitions of V_1 and V_2 in Eq. (38) similar results are obtained.

In all three cases of the above GLA's associated with so(4), so(3,1), and so(2,2) the metric tensors h and g are such that

$$\det\|h\| \neq 0, \quad \det\|g\| \neq 0, \tag{39}$$

as it is seen from Eqs. (35) and (36). Here again it is interesting to note from Eqs. (34) and (37) that while in the case of the above GLA with the compact LA so(4) as Bose sector the relation of the type in Eq. (18) is satisfied with

$$A^m_{\alpha\beta} = C'_{\alpha\sigma}F^\sigma_{m\beta}, \quad C'_{\alpha\sigma} = -C'_{\sigma\alpha},$$

$$[C'_{\alpha\sigma}] = (i/2)\,\tau_2, \tag{40}$$

$$m = 1,2,\ldots,6., \quad \alpha,\beta,\sigma = 1,2,$$

there does not exist a similar relation in the cases of the above GLA's with the noncompact LA's so(3,1) and so(2,2) as Bose sectors.

5. REPRESENTATIONS IN TERMS OF BILINEAR EXPRESSIONS OF BOSONS AND FERMIONS

Let us take m pairs of fermionic creation and annihilation operators, $\{a_j^+(+), a_j(+) \mid j = 1,2,\ldots,m\}$ and n pairs of bosonic creation and annihilation operators $\{a_k^+(-), a_k(-) \mid k = 1,2\ldots,n\}$ or a set of m pairs of bosonic operators, $\{a_j^+(-), a_j(-) \mid j = 1,2,\ldots,m\}$ and n pairs of fermionic operators, $\{a_k^+(+), a_k(+) \mid k = 1,2,\ldots,n\}$. These operators are such that

$$[a_j(\pm)a_k^+(\pm) \pm a_k^+(\pm)a_j(\pm)] = \delta_{jk},$$

$$[a_j(\pm)a_k(\pm) \pm a_k(\pm)a_j(\pm)] = 0 \tag{41}$$

$$\forall j,k.$$

Following Freund and Kaplansky[1] we shall define

$$R_{ij} = \begin{cases} a_i^+(\pm)\,a_j(\pm); & i,j = 1,2,\ldots,m, \\ a_i^+(\pm)\,a_{j-m}(\mp); & i = 1,2,\ldots,m, \quad j-m = 1,2,\ldots,n, \\ a_{i-m}^+(\mp)a_j(\pm); & i-m = 1,2,\ldots,n, \quad j = 1,2,\ldots m, \\ a_{i-m}^+(\mp)a_{j-m}(\mp); & i-m, \quad j-m = 1,2,\ldots,n \end{cases} \tag{42}$$

Freund and Kaplansky[1] have shown that in the case of a set of m fermions and n bosons corresponding to the choice of upper signs in all the brackets in Eq. (42) above, the operators

$$\{R_{ij}, R_{m+k,m+l} \mid i,j = 1,2,\ldots,m,$$
$$k,l = 1,2,\ldots,n\}$$

and

$$\{R_{i,m+k}, R_{m+l,j} \mid i,j = 1,2,\ldots,m,$$
$$k,l = 1,2,\ldots,n\} \text{ act}$$

respectively as even and odd generators of a GLA and under the corresponding commutation and anticommutation relations the set of all elements $\{R_{ij} \mid i,j = 1,2\ldots,m+n\}$ behave exactly like the matrices $\{E_{ij} \mid i,j = 1,2,\ldots,m+n\}$ with the correspondence $\{R_{ij} \longleftrightarrow E_{ij}\}$, where E_{ij} is the matrix with "1" in the (ij)–position and "0" everywhere else. Now it is easy to see from Eqs. (41) and (42) that in general the set of operators $\{R_{ij} \mid i,j = 1,2,\ldots,m+n\}$ has this property for the choice of either upper signs or lower signs in all the brackets in Eq. (42). From this it is obvious that if a GLA has an $(m+n)$–dimensional matrix representation Γ with even generators $\{Q_k \mid k = 1,2,..,D\}$ and odd generators $\{V_\alpha \mid \alpha = 1,2,\ldots,d\}$ represented by matrices of the type

$$\Gamma(Q_k) = \begin{pmatrix} A_k & 0 \\ 0 & B_k \end{pmatrix}, \quad k = 1,2,\ldots,D, \tag{43}$$

and

$$\Gamma(V_\alpha)=\begin{pmatrix}0 & C_\alpha\\ D_\alpha & 0\end{pmatrix},\quad \alpha=1,2,\ldots,d., \tag{44}$$

where A 's are $m\times m$ matrices, B 's are $n\times n$ matrices, C 's are $m\times n$ matrices, D 's are $n\times m$ matrices and 0's are null matrices of suitable dimensions, then the operators

$$\begin{aligned}
Q_k &= \sum_{i,j=1}^{m+n}\Gamma(Q_k)_{ij}R_{ij},\\
V_\alpha &= \sum_{i,j=1}^{m+n}\Gamma(V_\alpha)_{ij}R_{ij},\\
k &= 1,2,\ldots,D,\quad \alpha=1,2,\ldots,d,
\end{aligned} \tag{45}$$

form a representation of the generators of the given GLA.

For example let us consider the case of the GLA osp(2,1), with so(3) as its Bose sector, considered in Sec. 3. A three-dimensional matrix representation of the generators $\{Q_1,Q_2,Q_3,V_1,V_2\}$ of this GLA can be written as[2,7,13]

$$\begin{aligned}
&\Gamma(Q_1)=\frac{1}{2}\begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix},\ \Gamma(Q_2)=\frac{i}{2}\begin{pmatrix}0&-1&0\\1&0&0\\0&0&0\end{pmatrix},\\
&\Gamma(Q_3)=\frac{1}{2}\begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix},\\
&\Gamma(V_1)=\frac{1}{2}\begin{pmatrix}0&0&1\\0&0&0\\0&1&0\end{pmatrix},\ \Gamma(V_2)=\frac{1}{2}\begin{pmatrix}0&0&0\\0&0&1\\-1&0&0\end{pmatrix}.
\end{aligned} \tag{46}$$

Then it follows that the representations

$$\begin{aligned}
\Gamma'(Q_1)&=(G_{22}G_{33})^{1/2}\Gamma(Q_1),\\
\Gamma'(Q_2)&=(G_{33}G_{11})^{1/2}\Gamma(Q_2),\\
\Gamma'(Q_3)&=(G_{11}G_{22})^{1/2}\Gamma(Q_3),\\
\Gamma'(V_1)&=\Gamma(V_1),\quad \Gamma'(V_2)=\Gamma(V_2)
\end{aligned} \tag{47}$$

correspond to the generators of the GLA with structure constants given by Eqs. (21a)–(21c). Hence employing the above procedure we can conclude that the operators

$$\begin{aligned}
Q_1&=\tfrac{1}{2}(G_{22}G_{33})^{1/2}\\
&\quad\times[a_1^+(\pm)a_2(\pm)+a_2^+(\pm)a_1(\pm)],\\
Q_2&=\tfrac{1}{2}i(G_{33}G_{11})^{1/2}\\
&\quad\times[a_2^+(\pm)a_1(\pm)-a_1^+(\pm)a_2(\pm)],\\
Q_3&=\tfrac{1}{2}(G_{11}G_{22})^{1/2}\\
&\quad\times[a_1^+(\pm)a_1(\pm)-a_2^+(\pm)a_2(\pm)],\\
V_1&=\tfrac{1}{2}[a_1^+(\pm)a_1(\mp)+a_1^+(\mp)a_2(\pm)],\\
V_2&=\tfrac{1}{2}[a_2^+(\pm)a_1(\mp)-a_1^+(\mp)a_1(\pm)],
\end{aligned} \tag{48}$$

represent the generators of the GLA's with so(3) and so(2,1) as Bose sectors respectively when $G_{11}=G_{22}=G_{33}=1$ and $G_{11}=G_{22}=-G_{33}=1$.

Using a procedure similar to that adopted in Sec. 4 for obtaining the representations of the GLA's with so(4), so(3,1) and so(2,2) as Bose sectors from the representations of the GLA's with so(3) and so(2,1) as Bose sectors we can conclude as follows. If we take two commuting sets of operators $\{a_i^+(\pm), a_i(\pm), a_1^+(\mp), a_1(\mp)|i=1,2\}$ and $\{a_k'^+(\pm), a_k'(\pm), a_1'^+(\mp), a_1'(\mp)|k=1,2\}$, such that

$$\begin{aligned}
&[a_i(\pm)a_j^+(\pm)\pm a_j^+(\pm)a_i(\pm)]=\delta_{ij},\\
&[a_i(\pm)a_j(\pm)\pm a_j(\pm)a_i(\pm)]=0,\\
&[a_k'(\pm)a_l'^+(\pm)\pm a_l'^+(\pm)a_k'(\pm)]=\delta_{kl},\\
&[a_k'(\pm)a_l'(\pm)\pm a_l'(\pm)a_k'(\pm)]=0,\\
&[a_i(\pm)a_k'(\pm)-a_k'(\pm)a_i(\pm)]=0,\\
&[a_i^+(\pm)a_k'(\pm)-a_k'(\pm)a_i^+(\pm)]=0,\\
&[a_i(\pm)a_k'(\mp)-a_k'(\mp)a_i(\pm)]=0,\\
&[a_i^+(\pm)a_k'(\mp)-a_k'(\mp)a_i^+(\pm)]=0,\\
&i,j,k,l=1,2,
\end{aligned} \tag{49}$$

then the operators defined by

$$\begin{aligned}
Q_1&=\tfrac{1}{2}(G_{22}G_{33})^{1/2}[a_1^+(\pm)a_2(\pm)+a_2^+(\pm)a_1(\pm)\\
&\quad+a_1'^+(\pm)a_2'(\pm)+a_2'^+(\pm)a_1'(\pm)],\\
Q_2&=\tfrac{1}{2}i(G_{33}G_{11})^{1/2}[a_2^+(\pm)a_1(\pm)-a_1^+(\pm)a_2(\pm)\\
&\quad+a_2'^+(\pm)a_1'(\pm)-a_1'^+(\pm)a_2'(\pm)],\\
Q_3&=\tfrac{1}{2}(G_{11}G_{22})^{1/2}[a_1^+(\pm)a_1(\pm)-a_2^+(\pm)a_2(\pm)\\
&\quad+a_1'^+(\pm)a_1'(\pm)-a_2'^+(\pm)a_2'(\pm)],\\
Q_4&=\tfrac{1}{2}(G_{00}/G_{11})^{1/2}(G_{22}G_{33})[a_1^+(\pm)a_2(\pm)+a_2^+(\pm)a_1(\pm)\\
&\quad-a_1'^+(\pm)a_2'(\pm)-a_2'^+(\pm)a_1'(\pm)],\\
Q_5&=\tfrac{1}{2}i(G_{00}/G_{22})^{1/2}(G_{33}G_{11})[a_2^+(\pm)a_1(\pm)-a_1^+(\pm)a_2(\pm)\\
&\quad-a_1'^+(\pm)a_1'(\pm)+a_1'^+(\pm)a_2'(\pm)],\\
Q_6&=\tfrac{1}{2}(G_{00}/G_{33})^{1/2}(G_{11}G_{22})[a_1^+(\pm)a_1(\pm)-a_2^+(\pm)a_2(\pm)\\
&\quad-a_1'^+(\pm)a_1'(\pm)+a_2'^+(\pm)a_2'(\pm)],\\
V_1&=\tfrac{1}{2}[a_1^+(\pm)a_1(\mp)+a_1^+(\mp)a_2(\pm)],\\
V_2&=\tfrac{1}{2}[a_2^+(\pm)a_1(\mp)-a_1^+(\mp)a_1(\pm)],
\end{aligned} \tag{50}$$

generate a GLA with structure constants given by Eqs. (32)–(34) corresponding to GLA's with so(4), so(3,1) and so(2,2) as Bose sectors when $(G_{00}=G_{11}=G_{22}=G_{33}=1)$ $(-G_{00}=G_{11}=G_{22}=G_{33}=1)$ and $(-G_{00}=G_{11}=G_{22}=-G_{33}=1)$. In Eq. (50) we may replace $\{a_i'(\pm)|i=1,2\}$ by $\{a_i'(\mp)|i=1,2\}$.

It should be noted that even if we have a matrix representation of the generators of a GLA given as $\{\Gamma'(Q_m), \Gamma'(V_\alpha)|m=1,2,\ldots,D.,\alpha=1,2,\ldots,d.\}$ which are not of the block structure prescribed in Eqs. (43) and (44), we can obtain an operator representation in terms of a mixture of bosons and fermions by using the matrix representation

$$\Gamma(Q_m)=\begin{pmatrix}\Gamma'(Q_m) & 0\\ 0 & \Gamma'(Q_m)\end{pmatrix},\quad m=1,2,\ldots,D., \tag{51}$$

$$\Gamma(V_\alpha)=\begin{pmatrix}0 & \Gamma'(V_\alpha)\\ \Gamma'(V_\alpha) & 0\end{pmatrix},\quad \alpha=1,2,\ldots,d., \tag{52}$$

in Eq. (45) with suitable pairs of fermions and bosons.

6. CONCLUSION

In conclusion let us observe the following:

(i) From Eqs. (5b) and (5c) we have

$$[\{V_\alpha, V_\beta\}, V_\mu] = A^{\ m}_{\alpha\beta} F^{\sigma}_{m\mu} V_\sigma, \tag{53}$$

$$\begin{aligned}[\{V_\alpha, V_\beta\}, \{V_\mu, V_\lambda\}] &= \{[\{V_\alpha, V_\beta\}, V_\mu], V_\lambda\} \\ &\quad + \{[\{V_\alpha, V_\beta\}, V_\lambda], V_\mu\} \\ &= A^{\ m}_{\alpha\beta} F^{\sigma}_{m\mu}\{V_\sigma, V_\lambda\} + A^{\ m}_{\alpha\beta}F^{\nu}_{m\lambda}\{V_\nu, V_\mu\}, \\ m = 1,2,\ldots,D.,&\quad \alpha,\beta,\lambda,\mu,\sigma,\nu = 1,2,\ldots,d.\end{aligned} \tag{54}$$

Now if $\{Q_m | m = 1,2,..,D\}$ can be represented as linear combinations of $(\{V_\alpha, V_\beta\} | \alpha, \beta = 1,2,\ldots,d)$ as a result of Eq. (5c), then Eq. (5a) should be a direct consequence of Eq. (54) and hence of Eq. (53). Thus if $\{Q_m | m = 1,2,\ldots,D\}$ have realizations as linear combinations of $(\{V_\alpha, V_\beta\} | \alpha,\beta = 1,2,\ldots,d)$, then the consistency of Eqs. (5a)–(5c) implies that the representations of $\{V_\alpha | \alpha = 1,2,\ldots,d\}$ and hence of $\{Q_m | m = 1,2,\ldots,D\}$ can be obtained by solving Eq. (53) alone. Let us apply this consideration to the case of the GLA osp($2n$,1) described by Bednar and Sachl.[13] This GLA, denoted also as (sp($2n$); $2n$) in Ref. 13, is generated by the even elements $\{X_{ij} | i,j = -n,\ldots,-1,1,\ldots,n\}$ and the odd elements $\{V_k | k = -n,\ldots,-1,1,\ldots,n\}$ obeying the basic commutation and anticommutation relations given as

$$\begin{aligned}[X_{ij}, X_{kl}] &= \delta_{kj}X_{il} - \delta_{il}X_{kj} \\ &\quad + \epsilon_i\epsilon_j\delta_{-l,j}X_{k,-i} + \epsilon_j\epsilon_k\delta_{-i,k}X_{-j,l},\end{aligned} \tag{55a}$$

$$[X_{ij}, V_k] = \delta_{kj}V_i - \epsilon_i\epsilon_j\delta_{-i,k}V_j, \tag{55b}$$

$$\{V_k, V_l\} = 2\epsilon_k X_{l,-k} = 2\epsilon_l X_{k,-l}, \tag{55c}$$

$$i,j,k,l = -n,\ldots,-1,1,\ldots,n$$

with $\epsilon_i = 1$ for $i > 0$ and $\epsilon_i = -1$ for $i < 0$. In this case the relation among $\{V_k | k = -n,\ldots,n\}$ corresponding to Eq. (53) becomes

$$[\{V_j, V_k\}, V_l] = 2\epsilon_j\delta_{-j,l}V_k + 2\epsilon_k\delta_{-k,l}V_j, \tag{56}$$

$$j,k,l = -n,\ldots,-1,1,\ldots,n.$$

Now a comparison of Eqs. (1) and (56) reveals immediately that we can represent V's by para-Bose operators of any order as

$$V_k = \beta_k, \quad V_{-k} = \beta_k^+, \tag{57}$$

$$k = 1,2,\ldots,n.$$

Then $\{X_{ij} | i,j = -n,\ldots,-1,1,\ldots,n\}$ can be represented in terms of $\{\beta_i, \beta_i^+ | i = 1,2,\ldots,n\}$ through Eq. (55c) This fact has been responsible for the above realizations of the GLA's with so(3), so(2,1), so(4), so(3,1) and so(2,2) as Bose sectors in terms of para-Bose operators.

(ii) From Secs. 3 and 4 the following is clear. For both the GLA's with the compact LA's so(3) and so(4) as Bose sectors the relation in Eq. (18) is obeyed while so(3) and so(4) are simple and nonsimple respectively. In the case of all the GLA's with noncompact LA's, so(2,1), so(3,1) and so(2,2) as Bose sectors the relation in Eq. (18) is not obeyed while so(2,1) is simple and so(3,1) and so(2,2) are nonsimple. But all these GLA's obey Eq. (16). This pattern seems to suggest some connection between the compactness properties of the Bose sector of a GLA and the existence of a relation of the type in Eq. (18) between the structure constant matrices $\{F\}$ and $\{A\}$ of the GLA. Hence it should be worthwhile to study in more detail the above examples of GLA's from the view point of the connection between the structure of a GLA and the existence of a relation of the type in Eq. (18) as given by the so called C–theorem of Pais and Rittenberg[2] later generalized by Scheunert *et al.*[6]

ACKNOWLEDGMENTS

We wish to thank Professor Alladi Ramakrishnan for his kind encouragement. We thank for useful discussions Professor N. R. Ranganathan and Miss S. N. Uma, who is studying the above examples of GLA's in further detail with regard to the C– Theorem. One of the authors (R. V.) thanks Professor B. Gruber and his colleagues for kind hospitality at the Physics Department, Southern Illinois University, Carbondale, Illinois.

[1] P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).
[2] A. Pais and V. Rittenberg, J. Math. Phys. **16**, 2062 (1975); **17**, 598E (1976).
[3] H. S. Green, Phys. Rev. **90**, 270 (1953).
[4] T. F. Jordan, N. Mukunda, and S. V. Pepper, J. Math. Phys. **4**, 1089 (1963).
[5] R. Jagannathan and R. Vasudevan, J. Math. Phys **20**, 390 (1979); **20**, 2372 (E) (1979).
[6] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **17**, 1626, 1640 (1976); **18**, 146, 155 (1977).
[7] V. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709 (1978).
[8] M. Scheunert, J. Math. Phys. **20**, 712 (1979).
[9] V. G. Kac, Adv. Math. **26**, 8 (1977).
[10] L. Corwin, Y. Ne'eman and S. Sternberg, Rev. Mod. Phys. **47**, 573 (1975).
[11] P. Fayet and S. Ferrara, Phys. Rep. **32C**, 249 (1977).
[12] N. Backhouse, J. Math. Phys. **19**, 2400 (1978).
[13] M. Bednar and V. F. Sachl, J. Math. Phys. **19**, 1487 (1978); **20**, 367 (1979).
[14] A. O. Barut and R. Raczka, Proc. Roy. Soc. (London) A **287**, 519 (1965).
[15] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[16] P. Cordero and G. C. Ghirardi, Fortschr. Phys. **20**, 105 (1972).

Plane waves and topology

D. K. Ross
Physics Department, Iowa State University, Ames, Iowa 50011

(Received 3 October 1980; accepted for publication 5 June 1981)

We investigate particles which can be described in terms of plane waves and find that the number of topologically distinct such particles is the same as the number of disconnected pieces of the gauge group or structure group of the appropriate fiber bundle.

PACS numbers: 14.60.Gh, 02.40. + m

I.INTRODUCTION

The different neutrinos (ν_e, ν_μ, ν_γ) are perplexing since they are all apparently neutral, spin 1/2, left handed, apparently massless (experiments[1] suggesting neutrino oscillations and a nonzero neutrino mass have not been confirmed and have come under increasing criticism) particles which are nonetheless distinct in their interactions.[2] With all these attributes in common, one is hard put to imagine structurally how they differ from one another. Similar remarks can be made about other families of closely associated particles such as the charged leptons or the various quarks. We can, of course, describe such a situation through a separate flavor group or a "horizontal" group unified with the other gauge interactions.[3] A more exciting possibility, however, which we address in this paper, is that in some cases such particles may be represented through topologically inequivalent fiber bundles. Thus the difference between ν_e, ν_μ, and ν_γ may be in the inequivalent topological twists of some appropriate bundle. Rather than analyze some particular gauge theory in detail, we will leave the gauge group G as general as possible and delimit the situation by focusing our attention on particles which can be represented as plane waves. For any physical class of plane-wave particles such as neutrinos, we eventually will conclude that the number of topologically inequivalent such particles, related to G through an associated fiber bundle, is the same as the number of disconnected pieces in G. The results of this paper arose out of an attempt to establish a more detailed one-to-one correspondence between specific neutrinos and various twists of a principle fiber bundle with a *connected* gauge group. This interesting possibility died on the rocky ground of Steenrod's classification theorem used later, when the various configurations were found to be topologically equivalent.

In Sec. II, we discuss the fiber bundle description of gauge fields and the concept of particles as cross sections of associated vector bundles. We particularly concentrate on particles which can be represented as plane waves. In Sec. III, we enumerate the topological varieties of fiber bundles associated with such particles.

II. FIBER BUNDLES AND GAUGE FIELDS

Our starting point will be a principal fiber bundle $\underline{P}$, with structure group G, projection π, and with Minkowski space, M as the base space. π: $\underline{P}\rightarrow B$ is a C^∞-surjection of $\underline{P}$ onto M. Gauge potentials associated with the four basic interactions of physics are then connections in $\underline{P}$ and gauge fields are curvature 2-forms of $\underline{P}$.[4] This is just the usual description of a gauge theory[5] in terms of a fiber bundle. $\underline{P}$ can be viewed as a generalized and possibly nontrivial topological product of G and M. (see Fig. 1.) Notice that for the base space we explicitly use Minkowski space M rather than R^4 or R^4-{ball around origin} as more commonly done. Using M is clearly preferred physically and will be important below. G might be $SU_2\times U_1\times$ color SU_3, SU_5, SU_9, etc.

A more explicit definition of a fiber bundle is given in the lovely book by Choquet-Bruhat, DeWitt-Morette, and Dillard-Bleick.[6] A fiber bundle ($\underline{P}, B, \Pi, G$) is two topological spaces $\underline{P}$ and B (the base space $\equiv M$ in our case) with a continuous surjective mapping Π: $\underline{P}\rightarrow B$, a typical fiber F with a topological group G of homeomorphisms of F onto itself, and a covering of B by a family of open sets $\{U_j; j\in J\subseteq N\}$ such that (1) Locally $\Pi^{-1}(U_j)$ is homeomorphic to the topological product $U_j\times F$ for all $j\in J$. This homeomorphism $\Phi_j:\Pi^{-1}(U_j)\rightarrow U_j\times F$ has the form $\Phi_j(p) = (\Pi(p),\overset{\Delta}{\Phi}_j(p))$. If $x\in U_j$, $\overset{\Delta}{\Phi}_{j,x}$ is a homeomorphism from F_x (the fiber over x) onto F. (2) The homeomorphism $\overset{\Delta}{\Phi}_{k,x}\circ\overset{\Delta}{\Phi}{}_{j,x}^{-1}:F\rightarrow F$ is an element of the structural group G for all $x\in U_j\cap U_k$ and all $j,k\in J$. (3) The induced mapping $g_{ik}: U_j\cap U_k \rightarrow G$ by $x\rightarrow g_{jk}(x) = \overset{\Delta}{\Phi}_{k,x}\circ\overset{\Delta}{\Phi}{}_{j,x}^{-1}$ is continuous. A principal fiber bundle is one in which the typical fiber F and the structural group G are identical and in which G acts on F by left translation.

In the usual gauge theory, particle fields such as electrons or neutrinos are sections of a vector bundle associated to the principal fibration. From Mayer,[5] we can define a

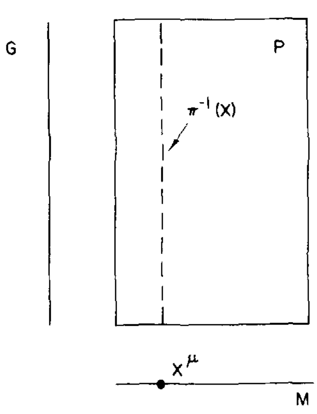


FIG. 1. Principal fiber bundle P with structure group G, projection Π, and base space M Minkowski space.

vector space V on which the group G acts on the left by a representation r: $G \rightarrow \mathrm{GL}(V)$, where

$$r:(g,v) \rightarrow r(g)v, \quad r(g_1 g_2) = r(g_1) r(g_2), \tag{1}$$

$$r(e) = I, \quad r(g^{-1}) = [r(g)]^{-1} \quad \text{with } g, g_1, g_2 \in G, \tag{2}$$

and $v \in V$.

On the product of the principal bundle $\underline{P}$ with V, we define the right action of G as

$$(x,v)\cdot g = (x \cdot g, r(g^{-1})v). \tag{3}$$

The associated vector bundle is then the orbit space of this action, $(\underline{P} \times V)/G$, with typical fiber V, base space M, and projection W where for each point of $(\underline{P} \times V)/G$ we denote by $W(z) = W(x,v)$ the point of M equal to $\pi(x)$ for all $(x,v) \in (\underline{P} \times V)/G$. More generally, any manifold E for which there is a morphism ρ: $\underline{P} \times V \rightarrow E$ such that $\rho(xg, r(g^{-1})v) = \rho(x,v)$ and such that the quotient map $\bar{\rho}:(\underline{P} \times V)/G \rightarrow E$ is a diffeomorphism is called a vector bundle associated to the principal fibration.

Let us be a little more explicit now in regards to the gauge group or structure group G and the vector space V. We will be interested in a class of similar particles below such as neutrinos. These particles must be cross sections of the associated vector bundle E. Thus G must be chosen in such a way that the various relevant properties of the particles which distinguish the given class can be encompassed and defined by G. G may be something like SU_5 or it may have to include gravitation also, through some version of extended supergravity,[7] depending on the specific application. In this paper we want to see if we can get anything interesting *without* specifying G in detail, realizing that G contains most of the physics. The space V is a multidimensional vector space which is associated with G by means of a representation as above.

The associated vector bundle E, defined above, is too general for our purposes. Sections of E describe a general particle field over M, as mentioned above (see Fig. 2). Let us now further focus our attention on freely propagating particles which can be described by plane wave solutions of the form

$$\Psi = N e^{-i(Et - \mathbf{p}\cdot\mathbf{x})}. \tag{4}$$

N is a normalizing factor and may contain spinors and various group indices. Although no interacting particle can ever be truly represented by a plane wave, nonetheless, for space-time regions far removed from previous or subsequent interactions, plane waves are a very good approximation. By going to simple wave functions of this type, we hope to extract physical results which would otherwise remain obscure.

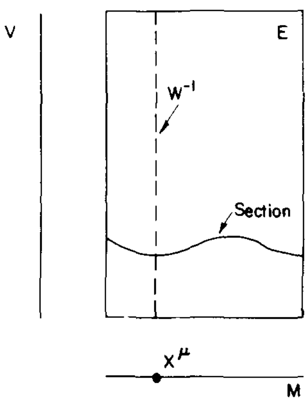


FIG. 2. Associated vector bundle E with typical fiber V, projection W, and base space M Minkowski space. A section of E is a general particle field.

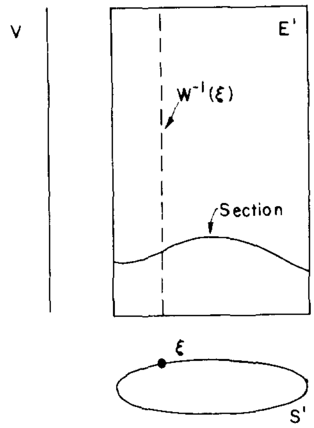


FIG. 3. Vector bundle E' with typical fiber V, projection W, and base space the circle S^1. A section of E' is a particle which can be described by a plane wave.

Limiting ourselves to particles which can be represented by plane waves leads to strong mathematical constraints. Thus, we are interested not in E but in the vector bundle E' whose sections are of the form (4). Sections of E' are not general functions of $\mathbf{x},t$ in Minkowski space but functions of $\xi \equiv Et - \mathbf{p}\cdot\mathbf{x}$ only. Also Ψ is periodic in the one-dimensional variable ξ with period 2π so that $\xi \pm 2\pi n = \xi$ as far as the particle wave function (4) is concerned. Thus we can take E' to be a vector bundle with typical fiber the multidimensional vector space V and base space the *one-dimensional circle* S^1 representing the variable ξ (see Fig. 3).

Sections of the bundle E' describe particles which can be described by free plane-wave wavefunctions. Locally these sections look like sections of simple products of V and S^1 (fiber bundles are always locally trivial). Globally, however, E' need not be trivial. As a simple example, a cylinder is a trivial vector bundle with a one-dimensional interval $I \subset R$ as typical fiber and S^1 as the base space. Using the same typical fiber and base space we can also construct the Möbius strip which is a nontrivial bundle. The twisting of the Möbius strip is tied up in the relationship between the homeomorphisms $\overset{\Delta}{\Phi}_{j,x}$ and $\overset{\Delta}{\Phi}_{k,x}$ mentioned above in the definition of a fiber bundle. Locally, the Möbius strip and the cylinder are equivalent, but they are not equivalent globally.

III. CLASSIFICATION OF THE FIBER BUNDLES

We have now distilled the problem down to enumerating the number of possible global topological varieties of the associated bundle E'. This should tell us the number of topologically distinct particles (in a given class specified by G) which can be represented by a plane wave. The problem of classifying fiber bundles over N-dimensional spheres has been solved. We have the following theorem from Steenrod[8]:

"The equivalence classes of bundles over S^n with group G are in one–one correspondence with equivalence classes of elements of $\Pi_{n-1}(G)$ under the operations of $\Pi_0(G)$. Such a correspondence is provided by $\mathscr{B} \rightarrow \chi(\alpha)$, where α is a generator of $\Pi_n(S^n)$ and $\chi: \Pi_n(S^n) \rightarrow \Pi_{n-1}(G)$ is a characteristic homeomorphism of $\mathscr{B}$." For our present application, we are interested in equivalence classes of bundles over S^1 and hence equivalence classes of elements of $\Pi_0(G)$ under the operations of $\Pi_0(G)$ (an abelian group), where G is the original gauge group or structure group we started with. The zeroth homotopy group Π_0 measures connectedness.[5] Thus $\Pi_0(\mathrm{O}(n)) = Z_2$ for $n \geqslant 1$ since $\mathrm{O}(n)$ consists of two disconnected pieces with positive and negative determinants. On the other hand, $\Pi_0(\mathrm{SO}(n)) = \Pi_0(\mathrm{U}(n)) = \Pi_0(\mathrm{SU}(n)) = 0$ for $n \geqslant 1$ since these groups are all connected.

We thus conclude that the number of inequivalent topological varieties of our associated bundle E' is simply the number of disconnected pieces in the original structure group or gauge group G that we started with. Thus the number of topologically distinct particles which are associated with the gauge group G as above and which can be represented as a plane wave is the same as the number of disconnected pieces in G. As an example of this result, if G is chosen such that the cross sections of the associated vector bundle E' represent neutrino plane waves, then the number of topologically distinct neutrinos is just the number of pieces in G. If the appropriate G is disconnected, this paper suggests that neutrino flavor may have a topological origin. The hard part, of course, is to find the G that nature has chosen.

[1]V. Barger *et al.*, Phys. Lett. B **93**, 194 (1980); A. De Rújula *et al.*, Nucl. Phys. B **168**, 54 (1980); F. Reines, H. W. Sobel, and E. Pasierb, Phys. Rev. Lett. **45**, 1307 (1980).

[2]G. Danby, J. M. Gaillard, K. Goulianos, L. M. Lederman, N. Mistry, M. Schwartz, and J. Steinberger, Phys. Rev. Lett. **9**, 36 (1962).

[3]P. Langacker, SLAC Preprint PUB-2544 (to appear in Phys. Rep.), contains numerous references.

[4]A. Trautman, Rep. Math. Phys. **1**, 29 (1970).

[5]M. E. Mayer, in *Fiber Bundle Techniques in Gauge Theories* (Springer, Berlin, 1977).

[6]Y. Choquet-Bruhat, C. De Witt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics* (North-Holland, Amsterdam, 1977).

[7]D. Z. Freedman and B. de Witt, Nucl. Phys. B **130**, 105 (1977); D. Z. Freedman, Phys. Rev. Lett. **38**, 105 (1977); J. Sherk, E. Cremmer, and S. Ferrara, Phys. Lett. B **74**, 61 (1978); E. Cremmer, J. Scherk, and J. Schwarz, Phys. Lett. B **84**, 83 (1979).

[8]N. Steenrod, *The Topology of Fibre Bundles* (Princeton U. P., Princeton, N. J., 1951).

A new method of constructing the symmetry coordinates of molecular vibrations based on the correspondence theorem

Shoon K. Kim
Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122

(Received 24 March 1981; accepted for publication 18 May 1981)

A general method of constructing the symmetry coordinates of molecular vibrations is presented via the correspondence theorem which describes the parallelism between a SALC (symmetry adapted linear combination) and an elementary basis function belonging to the same irreducible representation of the symmetry group of the molecule.

PACS numbers: 33.10.Gx

I. INTRODUCTION

In the theory of molecular vibrations of a polyatomic molecule belonging to a point group G, the primary step is to construct the symmetry coordinates of molecular vibrations belonging to the irreducible representations (irreps) of the group G. These are the symmetry adapted linear combinations (SALC's) of the external or internal vibrational coordinates of the molecule.[1] Ordinary, these are constructed by means of the so-called projection operator method.[1,2] This method is very general and powerful but it is often extremely laborious to use.

Recently, the author has introduced a general method[3,4] of constructing the SALC's of equivalent basis functions of point groups. It is based on the correspondence theorem which describes the parallelism between a SALC and an elementary basis function belonging to the same irrep of G. Here, the latter means a basis function given by a homogeneous polynomial in the spatial coordinates x,y,z. More specifically, the linear coefficients of the SALC's are determined from the property of the elementary basis set on the equivalent points with respect to the group G. It does not require the explicit matrix representations of G, unlike the projection operator method. The effectiveness of this method has been demonstrated by actual constructions of the symmetry adapted LCAO–MO's, hybrid AO's, and lattice harmonics, etc.[3,4] The present work applies this method to the construction of the symmetry coordinates of molecular vibrations, restricting the vibrations to be small.

For this purpose, it is necessary to characterize the transformation property of a set of equivalent vibrational coordinates of a molecule under consideration by a set of equivalent basis functions located at equilibrium points defined on the molecule. For example, the transformation property of an infinitesimal atomic displacement can be characterized by a set of p orbitals (p_x, p_y, p_z) located at the position of the atom. Such characterizations are also possible for internal vibrational coordinates, even though they are given by the linear combinations of two or more atomic displacements. This then enables us to write down the general expression for the internal or external symmetry coordinates in terms of the elementary basis functions or basis operators of the group G.

In Sec. 2, the correspondence theorem will be discussed to the extent which will be needed in the present work. Then, the external symmetry coordinates are expressed as the SALC's of equivalent atomic displacements. For the internal coordinates, we shall first show how to characterize a set of equivalent ones by a set of equivalent basis functions, then establish the rule of characterization for each type: The types of internal coordinates which will be considered are the bond stretchings, valence-angle-bending, bond-plane angle changes (the angle between a bond and a plane defined by two bonds), and bond twistings (or torsions). The accurate definitions of these coordinates are given in the classic work of Wilson *et al.*[1] It will be shown that the general expression for the internal symmetry coordinates takes a particularly simple form since the intrinsic symmetry property of any internal coordinate is one dimensional. Simple illustrative examples are given in Sec. 3. Here, we shall characterize the irreps of G by their elementary basis functions.

II. THE GENERAL EXPRESSIONS FOR THE SYMMETRY COORDINATES

We shall first discuss the correspondence theorem in a form which is most suitable to the present problem. Let $S^{(n)} = \{r_\nu^0; \nu = 1,2...,n\}$ be a set of symmetrically equivalent points on a molecule belonging to a point group G. Then, the set $S^{(n)}$ constitutes a transitive G space such that the transformation of $S^{(n)}$ define a $n\times n$ permutation representation $\Delta^{(n)}(R)$,

$$R\, r_\nu^0 = \sum_{\sigma=1}^{n} r_\sigma^0 \Delta_{\sigma\nu}{}^{(n)}(R) \quad \nu = 1,2...,n \tag{2.1}$$

for all 3×3 orthogonal transformations $R \in G$. Let $\{f_s^A(r)\}$ be a basis set of the space vector r belonging to a unitary representation $D^A(R)$ (of dimension d_A) of G, then, the set $\{f_s^A(r - r_\nu^0)\}$ defines a set of equivalent basis functions belonging to the direct product representation $\Delta^{(n)}(R)\times D^A(R)$,

$$f_s^A(R^{-1}r - r_\nu^0) = \sum_{k=1}^{d_A} \sum_{\sigma=1}^{n} f_k^A(r - r_\sigma^0)\, \Delta_{\sigma\nu}^{(n)}(R)\, D_{ks}^A(R). \tag{2.2}$$

Accordingly, all SALC's of these equivalent basis functions are classified by the irreps contained in $\Delta^{(n)}(R)\times D^A(R)$. Let $D^\gamma(R)$ be one of these irreps, then a set of SALC's belonging to $D^\gamma(R)$ is given by

$$\psi_i^\gamma(r) = \sum_{s=1}^{d_A} \sum_{\nu=1}^{n} [\mathring{T}_i^\gamma(r)\, g_s^A(r)^*]_{r=r_\nu^0} f_s^A(r - r_\nu^0), \tag{2.3}$$

provided that the set is not null. Here, $\mathring{T}_k^\gamma(r)$ is a basis set of operators[4] belonging to $D^\gamma(R)$ acting on a basis set $\{g_\nu^A(r)\} \in D^A(R)$ and the asterisk denotes the complex conjugate. The basis operators $\mathring{T}_i^\gamma(r)$ are easily constructed by replacing part of the Cartesian coordinates in the corresponding elementary basis functions $T_i^\gamma(r)$ by their differential operators. It can be shown[4] that if $D^\gamma(R)$ is contained in $\Delta^{(n)}(R)\times D^A(R)$ n_γ times, one can construct n_γ linearly independent sets of SALC's by means of (2.3). This expression represents a special case of the correspondence theorem[4] in the sense that $\psi_i^\gamma(r)$ transforms like $\mathring{T}_i^\gamma(r)$. It is noted that the set of equivalent points $S^{(n)}$ plays a crucial role in determining the linear coefficients in (2.3). Thus, it is essential for the present work to characterize a set of equivalent vibrational coordinates by a set of equivalent basis functions defined on equivalent points on the molecule. In the following, this will be achieved for the external coordinates first then for the internal coordinates under the asumption that all vibrations considered are small.

A. The external vibrational coordinates

The external coordinates are defined by the atomic displacements in the Cartesian coordinates fixed in the molecular frame at equilibrium. Their transformation properties are characterized as follows.

The Cartesian displacement δr_ν of the νth equivalent atom transforms like a vector quantity $\rho_\nu(r) = r - r_\nu^0$, where r_ν^0 is the equilibrium position of the νth atom.

Let $D^{(3)}(R)$ be the three dimensional representation of the orthogonal transformation $R \in G$. Then $\{\rho_\nu(r)\}$ provides a basis for $\Delta^{(n)}(R)\times D^{(3)}(R)$. Let $D^\gamma(R)$ be an irrep contained in this direct product representation, then a set of external symmetry coordinates belonging to $D^\gamma(R)$ is given by

$$S_i^\gamma = \sum_{\nu=1}^{n} \mathring{T}_i^\gamma(r)(r\cdot\delta r_\nu)|_{r=r_\nu^0}. \tag{2.4}$$

We shall not discuss this case any further, since it is the same problem as the problem of constructing the SALC's of the equivalent p orbitals, which are fully discussed in the previous work.[4]

B. The internal vibrational coordinates

We shall first show how to characterize, in general, a set of equivalent internal coordinates of a molecule by a set of equivalent basis functions defined on the molecule. Let $\{\xi_\nu\}$ be a set of n-equivalent internal coordinates. Then under the symmetry operations R's of G, the set transforms according to a direct product representation $\Delta^{(n)}(R)\times D^\xi(R)$, where $D^\xi(R)$ is a one dimensional representation of G intrinsic to the type of the internal coordinates ξ_ν and $\Delta^{(n)}(R)$ is the $n\times n$ matrix representation of $R\in G$, which describes the permutation of the absolute values $|\xi_\nu|$ of the coordinates. Then, this definition of $\Delta^{(n)}(R)$ determines the corresponding set of equivalent points $S^{(n)} = \{r_\nu^0;\nu = 1,2...,n\}$ on the molecule: It is given by a set of n-equivalent points generated by all $R\in G$ from a single point which is invariant under the operations which leave one of $|\xi_\nu|$ invariant. It is also evident that there always exists a basis function $\xi(r)$ belonging to a one dimensional representation $D^\xi(R)$ of the point group G. Accordingly, the set $\{\xi(r - r_\nu^0)\}$ defined on $S^{(n)}$ gives a required set of equivalent basis functions which characterizes the transformation property of the set $\{\xi_\nu\}$. The following are the rules which define the basis set $\{\xi(r - r_\nu^0)\}$ corresponding to each type of the internal coordinates $\{\xi_\nu\}$ most frequently encountered in molecular vibrations:

(1) A bond-stretching transforms like a scalar function located at the midpoint of the bond.

(2) A valence-angle-bending (the change in the angle between two valence bonds) transforms like a scalar function located at a point on the line which bisects the angle in the plane of two bonds. Evidently, the actual point on the line must be chosen in accordance with the rest of the equivalent angles.

It should be noted here that depending on the molecular symmetries it is possible to choose alternative sets of locations $\{r_\nu^0\}$ for the scalar functions in (1) and (2).

(3) Let α_ν be the bond-plane angle defined by three bonds meeting at the νth equivalent atom, the three bonds being initially on a plane perpendicular to the z axis. Then, the change of the angle α_ν between one bond and the plane of the remaining two bonds transforms like $z_\nu = z - z_\nu^0$; i.e., it transforms like the p_z atomic orbital located at the νth atom.

This rule is easily understood from the observation that the angle change α_ν can be regarded as an infinitesimal lifting of the vertex atom out of the plane leaving the rest of the three atoms in the plane. Consequently, this rule holds also for lifting a vertex atom where more than three bonds in a plane meet at the vertex atom.

(4) A bond-twisting (or torsion) transforms like a pseudoscalar with respect to the point group G located at the midpoint of the bond.

Here, a pseudoscalar with respect to a point group G means a function of the space vector which is invariant under all proper rotations of G and changes its sign under all improper rotations of G if these are contained in G. Simple examples of the pseudoscalar functions for O_h and T_d include $xyz\tilde{x}\tilde{y}\tilde{z}$ and $\tilde{x}\tilde{y}\tilde{z}$, respectively. Here $\tilde{z} = xy' - yx'$, etc., and the coordinate systems are taken as given by Koster *et al.*[5] Note that T_d is a subgroup of O_h and xyz is invariant under all operations of T_d. For the case of D_{nh}, an elementary basis $z\tilde{z}$ is a pseudoscalar of the group, the z axis being along the prinicipal axis.

Finally, we shall write down the general expression for the symmetry coordinates of $\{\xi_\nu\}\in\Delta^{(n)}(R)\times D^\xi(R)$. Let $D^\alpha(R)$ be an irrep contained in $\Delta^{(n)}(R)$ and let $\{U_i^\alpha(r)\}$ be an elementary basis set of $D^\alpha(R)$. Then a set of symmetry coordinates belonging to $D^\alpha(R)\times D^\xi(R)$ is given by

$$\xi_i^\alpha = \sum_{\nu=1}^{n} U_i^\alpha(r_\nu^0)\xi_\nu;\ i = 1,2,...,d_\alpha \tag{2.5}$$

provided that the set $\{U_i(r_\nu^0)\}$ is not null. Such a basis set is called proper on $S^{(n)}$. It has been shown[3] that if $D^\alpha(R)$ is contained in $\Delta^{(n)}(R)$ n_α times, there exist n_α linearly independent proper basis sets on $S^{(n)}$. According to (2.5), for the construction of the symmetry coordinate of ξ_ν it is most convenient to classify the irreps of G by the direct product $D^{\alpha}(R)\times D^{\xi}(R)$; this is always possible since the intrinsic representation $D^{\xi}(R)$ is one dimensional. Application of (2.5) will be discussed in the next section.

3. ILLUSTRATIVE EXAMPLES

With use of the general expression (2.5) and the rules of characterization (1)–(4), the problem of constructing the internal symmetry coordinates for small molecular vibrations becomes almost trivial. Nevertheless, it seems worthwhile to discuss some illustrative examples and make additional remarks through examples.

We shall first consider the small vibrations of the benzene molecule belonging to D_{6h}. This problem has been throughly discussed in the treatise given by Wilson *et al.*[1] We shall discuss here only the construction of the internal symmetry coordinates. It serves as an excellent example for comparison with the ordinary method as well as for the use of the rules (1)–(4) since it requires all of them. Let us place the origin of a Cartesian coordinate system at the center of the benzene molecule and let the z axis be perpendicular to the molecular plane and the y axis pass through a carbon atom. The irreps of D_{6h} are characterized by the elementary basis functions as follows:

$$\begin{array}{ll}
\Gamma_1^+;1 & \Gamma_1^-;z\tilde{z}\\
\Gamma_2^+;\tilde{z}\text{ or }\Gamma_3^-\times\Gamma_4^- & \Gamma_2^-;z\\
\Gamma_3^+;z(x^3-3xy^2) & \Gamma_3^-;y^3-3yx^2\\
\Gamma_4^+;z(y^3-3yx^2) & \Gamma_4^-;x^3-3xy^2\\
\Gamma_5^+;(\tilde{x}\tilde{y})\text{ or }z(y,-x) & \Gamma_5^-;(x,y)\text{ or }\tilde{z}(y,-x)\\
\Gamma_6^+;(x^2-y^2,2xy)\text{ or} & \Gamma_6^-;z(x^2-y^2,2xy)\text{ or}\\
\quad\tilde{z}(2xy,-(x^2-y^2)) & \quad z\tilde{z}(2xy,-(x^2-y^2)).
\end{array}\tag{3.1}$$

Here, the notations Γ's for the irreps are those of Koster *et al.*[5] and $\tilde{x}=yz'-zy'$, etc. For later use, we have introduced extra basis sets for some of the irreps with the correct orientations for degenerate cases: for example, two basis sets $(\tilde{x}\tilde{y})$ and $z(y,-x)$ in Γ_5^+ tell us that Γ_5^+ is equivalent to $\Gamma_2^-\times\Gamma_5^-$ and that $(\tilde{x},\tilde{y})$ and $(zy,-zx)$ belong to the same irrep up to a phase factor.

Let $S^{(6)}$ be the set of the equilibrium positions $\{r_\nu^0\}$ of the six carbon atoms and their x,y coordinates be as follows, in the order $r_1^0\sim r_6^0$,

$$\begin{aligned}&(0,1),\ (\tfrac12\sqrt{3},\tfrac12),\ \ (\tfrac12\sqrt{3},-\tfrac12),\\ &(0,-1),\ (-\tfrac12\sqrt{3},-\tfrac12),\ (-\tfrac12\sqrt{3},\tfrac12).\end{aligned}\tag{3.2}$$

Let $\bar{S}^{(6)}$ be the set of the midpoints of the six C–C bonds given by

$$\bar{r}_\nu^0=\tfrac12(r_\nu^0+r_{\nu+1}^0);\ r_7^0=r_1^0,\ \nu=1,2,\ldots,6.\tag{3.3}$$

It will be seen that these two sets define all sets of equivalent internal coordinates of the benzene molecule. Let $\Delta^{(6)}(R)$ and $\bar{\Delta}^{(6)}(R)$ be the permutation representation via $S^{(6)}$ and $\bar{S}^{(6)}$, respectively. Then, the irreps contained in these permutations are given as follows.

$$\begin{array}{lllll}\Delta^{(6)}(R)=\Gamma_1^+ & +\Gamma_3^- & +\Gamma_5^- & +\Gamma_6^+, & (3.4a)\\ \qquad\qquad 1 & (y^3-3yx^2) & (x,y) & (x^2-y^2,2xy) &\end{array}$$

$$\begin{array}{l}\bar{\Delta}^{(6)}(R)=\Gamma_1^++\Gamma_4^-+\Gamma_5^-+\Gamma_6^+,\\ \qquad\qquad\qquad (x^3-3xy^2)\end{array}\tag{3.4b}$$

where we have also given the respective elementary bases sets which are proper and hence will be used for $U_i^{\alpha}(r)$ of (2.5).

It is most convenient to introduce the following notations for the symmetry coordinates belonging to $D^{\alpha}(R)\times D^{\xi}(R)$,

$$\begin{aligned}\xi_{U_i^\alpha}&=\sum_\nu U_i^\alpha(r_\nu^0)\xi_\nu,\quad \text{if } D^\alpha(R)\in\Delta^{(6)}(R),\\ \bar{\xi}_{U_i^\alpha}&=\sum_\nu U_i^\alpha(\bar{r}_\nu^0)\xi_\nu,\quad \text{if } D^\alpha(R)\in\bar{\Delta}^{6}(R),\end{aligned}\tag{3.5}$$

where the difference occurs only in the arguments of the linear coefficients.

Since the benzene molecule is planar, we shall discuss the in-plane and the out-of-plane modes separately.

A. The in-plane modes

From the rule (1), the six C–H stretches s_ν and the six C–C stretches t_ν belong simply to their respective permutation representations,

$$\{s_\nu\}\in\Delta^{(6)}(R);\quad \{t_\nu\}\in\bar{\Delta}^{(6)}(R).\tag{3.6}$$

Thus, one can immediately write down their SALC's using (3.4) and (3.5). The 12 equivalent C–C–H valence angle bendings (in-plane) are regarded as six equivalent subsets $\{\theta_{\nu1},\theta_{\nu2}\}$, $\nu=1,2\ldots6$; each subset belonging to the subgroup C_{2v} of D_{6h}. Then, the induced representation onto D_{6h} via the irreps of C_{2v} lead to the following two basis sets on $S^{(6)}$,

$$\begin{aligned}\alpha_\nu&=\theta_{\nu1}+\theta_{\nu2}\in\Delta^{(6)}(R),\\ \beta_\nu&=\theta_{\nu1}-\theta_{\nu2}\in\Delta^{(6)}(R)\times\Gamma_2^+.\end{aligned}\tag{3.7}$$

Again, one can write down the corresponding SALC's using (3.4) and (3.5). To classify the SALC's of β_ν's in terms of the original irreps given in (3.1) one may need

$$\Delta^{(6)}(R)\times\Gamma_2^+=\Gamma_2^++\Gamma_4^-+\Gamma_5^-+\Gamma_6^+.\tag{3.8}$$

Collecting the results obtained thus far we have the following sets of the symmetry coordinates belonging to the in-plane modes of vibrations,

$$\begin{aligned}&\Gamma_1^+;s_1,\bar{t}_1,\alpha_1\\ &\Gamma_2^+;\beta_1\\ &\Gamma_3^-;s_{(y^3-3xy^2)},\alpha_{(y^3-3yx^2)}\\ &\Gamma_4^-;\bar{t}_{(x^3-3xy^2)},\beta_{(y^3-3yx^2)}\\ &\Gamma_5^-;(s_x,s_y),(\bar{t}_x,\bar{t}_y),(\alpha_x,\alpha_y),(\beta_y,-\beta_x)\\ &\Gamma_6^+;(s_{x^2-y^2},s_{2xy}),(\bar{t}_{x^2-y^2},\bar{t}_{2xy}),(\alpha_{x^2-y^2},\alpha_{2xy}),\\ &\qquad(\beta_{2xy},-\beta_{x^2-y^2}).\end{aligned}\tag{3.9}$$

Here, it is noted that $\alpha_1=0$ and one of (α_x,α_y) and $(\bar{t}_x,\bar{t}_y)$ are redundant. The above results are consistent with the direct decomposition of the in-plane mode Γ_{in},

$$\begin{aligned}\Gamma_{\text{in}}=2\Gamma_1^++\Gamma_2^++2\Gamma_3^-+2\Gamma_4^-+3\Gamma_5^-\\ +4\Gamma_6^+.\end{aligned}\tag{3.10}$$

B. The out-of-plane mode

Let $\{\gamma_\nu\}$ be the bond-plane angle bending at the six carbon atoms and $\{\tau_\nu\}$ be the torsions of the six C–C bonds. According to the rules (3) and (4) these are represented by $\{z_\nu\}$ on $S^{(6)} = \{r_\nu^0\}$ and $\{(z\bar{z})_\nu\}$ on $\bar{S}^{(6)} = \{\bar{r}_\nu^0\}$, respectively. Thus, from (3.1) and (3.4) we have

$$\{\gamma_\nu\} \in \Delta^{(6)} \times \Gamma_2^- = \Gamma_2^- + \Gamma_4^+ + \Gamma_5^+ + \Gamma_6^-, \quad (3.11)$$

$$\{\tau_\nu\} \in \bar{\Delta}^{(6)} \times \Gamma_1^- = \Gamma_1^- + \Gamma_4^- + \Gamma_5^+ + \Gamma_6^-. \quad (3.12)$$

The corresponding symmetry coordinates are again written down with use of (3.4) and (3.5). The final results for the out-of-plane modes are summarized as follows:

$$\begin{aligned}
&\Gamma_1^-;\ \bar{\tau}_1\\
&\Gamma_2^-;\ \gamma_1\\
&\Gamma_4^+;\ \gamma_{(y^3-3yx^2)},\ \bar{\tau}_{(x^3-3xy^2)}\\
&\Gamma_5^+;(\gamma_x,\gamma_y),\ (\bar{\tau}_y,\ -\bar{\tau}_x)\\
&\Gamma_6^-;\ (\gamma_{x^2-y^2},\gamma_{2xy}),\ (\bar{\tau}_{2xy},\ -\bar{\tau}_{x^2-y^2}).
\end{aligned} \quad (3.13)$$

Here, one of the two bases belonging to Γ_5^+ is redundant and also $\bar{\tau}_1 = 0$, on account of the ring structure of the benzene molecule. The above results are again consistent with the direct decomposition of the out-of-plane modes given by

$$\Gamma_{\text{out}} = \Gamma_2^- + 2\Gamma_4^+ + \Gamma_5^+ + 2\Gamma_6^-. \quad (3.14)$$

We shall not write down the final explicit forms of these symmetry coordinates, which are trivally obtained with use of (3.2) and (3.3). The explicit results thus obtained are consistent with the well known results given by Wilson *et al.*

Evidently, we can treat any planar molecule in the analogous manner as given above. Next, we consider the vibrational coordinates of the methane molecule $CH_4 \in T_d$. Here, it is necessary to construct the symmetry coordinate for the four C–H stretches and six valence angle bending of H–C–H. Since these are represented by scalar functions, the problem is reduced ot that of constructing the SALC's of the s orbitals located at the four equivalent positions $\{r_\nu^0, \nu = 1,2,3,4\}$ of the H atoms and at the six equivalent positions $\{\frac{1}{2}(r_\nu^0 + r_\mu^0); \nu \neq \mu\}$ given by the midpoints of two different H atoms. In the analogous manner one can construct the internal symmetry coordinates of a molecule $XY_6 \in O_h$.

The symmetry coordinates thus far discussed are not the normal coordinates which completely diagonalize the respective Hamiltonian, in the case when there exist more than one set of symmetry coordinates belonging to an irrep. In such a case it is necessary to solve the respective secular equation for each irrep. It is noted here that there exists a simple algebraic method of matrix diagonalization introduced recently by the author.[6] According to this method, one can write down the transformation matrix explicitly from the reduced characteristic equation of the matrix which is to be diagonalized.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Haw-Won Kim, Department of Chemistry, Temple University, for reading the manuscript and giving helpful suggestions.

[1] E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, New York, 1955), p. 135.

[2] E. P. Wigner, *Group Theory and its Applications to the Quantum Mechanics of Atomic Spectra* (Academic, New York, 1959).

[3] Shoon K. Kim, Physica A, **105A**, 521 (1981).

[4] Shoon K. Kim, Physica A, **106A**, 577 (1981)

[5] G. H. Koster, J. O. Dimmock, R. G. Wheeler, and H. Statz, *Properties of the Thirty-two Point Groups* (M.I.T. Press, Cambridge, MA, 1963).

[6] Shoon K. Kim, J. Math. Phys. **20**, 2153 (1979); **20**, 2159 (1979).

2307,2316

Quantum field theory of particles of indefinite mass. II. An electromagnetic model

Levere Hostler
Physics Department, Wilkes College, Wilkes-Barre, Pennsylvania 18766

(Received 14 January 1981; accepted for publication 24 April 1981)

A new model of the electromagnetic field is postulated. The model is designed within the conceptual framework of the quantum field theory of particles of indefinite mass. The model is built around an interaction Lagrangian of the type $\mathscr{L}_I = \frac{1}{2}\rho V - J^\mu A_\mu$, where $(J^\mu;\rho)$ is a conserved c-number 5-current. A gauge invariance principle is thereby built into the theory. The object $\frac{1}{2}V$ is a new 5th component of the vector potential. Due to the use of the new evolution parameter τ, second quantization in the Lorentz gauge $\partial^\mu A_\mu = 0$ parallels Coulomb gauge quantization in ordinary quantum electrodynamics. The Hamiltonian of the free electromagnetic field is negative definite and has the physical interpretation of minus half the sum of the squares of the masses of the photons present. The inner product in Hilbert space is not positive definite, spacelike states with timelike polarization having negative norm.

PACS numbers: 41.10. – j, 11.10. – z

I. INTRODUCTION

A number of authors[1–7] have investigated new forms of relativistic quantum theory based on Fock type wave equations. Such theories are a form of "proper time theory" in which the evolution of the system is described by a parameter independent of the space–time coordinates. The wave functions in these theories are true space–time amplitudes subject to the normalization condition $\int d^4x|\phi|^2 = 1$, the integration being over all space–time.

Here we continue the work of Ref. 7, in which such a theory was built up from rigorous correspondence arguments, starting with a classical theory of particles of indefinite mass. A quantum field theory based upon this classical foundation describes particles of indefinite mass in which spin zero bosons obey a second order Fock equation, our Eq. (2.9)

In Sec. II we attempt to model the electromagnetic field within the framework of this formalism. Our new theory of electromagnetism is built around the interaction Lagrangian (2.1), in which the conserved 5-current (2.2) characteristic of indefinite mass theory is coupled to a 5-potential

$(A^\mu; V/2)$.

A principle of gauge invariance is thereby built into the theory. In the complete Lagrangian (2.10) of the electromagnetic field, the free field part is chosen to be a gauge invariant structure that leads to wave equations that are second order Fock equations. The theory presented here is essentially 5-dimensional in nature. This is reflected in the fact that the potentials acquire a new 5th component, $V/2$, not present in conventional electrodynamics. The new electromagnetic equations are presented in 3-dimensional form in Eqs. (2.23)–(2.29), and in 4-dimensional form in Eqs. (2.30)–(2.34). It is recognized that the model embodied in the new equations may be provisional, since much more will need to be investigated before it becomes clear whether the new equations can be a correct law of nature and whether they can be equivalent to the usual electromagnetic theory in a suitable limiting case. With this proviso, we proceed to take our model seriously, and to investigate its consequences.

Quantization of the new equations is carried out in Sec. III. Because our evolution parameter is independent of the space–time coordinates, the Lorentz gauge condition $\partial^\mu A_\mu = 0$ can here be treated like the Coulomb gauge of conventional electrodynamics. Accordingly, our quantization procedure parallels Coulomb gauge quantization in ordinary electrodynamics. Generalized coordinates $q^a(\mathbf{k},\tau)$ are introduced (Sec. III A) in terms of which the potential has an expansion (3.13) in which the $e_a(\mathbf{k})_\mu$ are polarization 4-vectors, and the $f(\mathbf{k},\mathbf{x})$ are the real basis functions (3.9). Conjugate momenta (3.23) can be defined in the usual way (Sec. III B) and simple canonical quantum conditions (3.24) can be imposed, thereby defining the operator structure of the theory. The implied quantum conditions in coordinate space, Eq. (3.43) of Sec. III C, have a form familiar from conventional Coulomb gauge quantum theory.[8] An unpleasant result of our quantization procedure is the use of an indefinite metric in Hilbert space, spacelike states with the timelike polarization vectors having negative norm. As a kind of compensation for going over to an indefinite metric, the free photon Hamiltonian (3.42) turns out to be negative definite for all states, including the "bad" states mentioned above. In Sec. IV the photon propagator is calculated. This has a form [see Eq. (4.5)] containing additional terms of a type familiar from conventional Coulomb gauge quantum theory. Previous experience with conventional Coulomb gauge field theory suggests the possibility of removing the additional terms in Feynman integrals by use of current conservation and suitable integrations by parts. This procedure is illustrated with the example of the vacuum–vacuum amplitude and leads to a simplified effective photon propagator, Eq. (4.16), for use in Feynman integrals.

In Appendix A we explore the physical interpretation of the 5-current conservation law, Eq. (2.2). We find that this conservation law can be understood in simple classical terms, assuming the 4-current density J_μ to be only a partial current density due to some but not all of the particles present. The ideas developed in Appendix A are proven out in Appendix B, where we calculate cross sections for scattering, pair production, and pair annihilation, using the indefinite mass formalism. The model considered [see Eq. (B1)] is that of a particle obeying the first order Fock equation and interacting with a conventional arbitrary external c-number electromagnetic field. Results in agreement with the usual quantum theory of particles of definite mass are obtained, and this agreement holds to all orders in the interaction.

We continue to use the same notation as in Ref. 7: Heavyside–Lorentz units are used, with $\hbar = c = 1$; and the metric in Minkowski space has $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$. The Einstein summation convention is assumed throughout.

II. GENERALIZATION OF MAXWELL'S EQUATIONS

We will here attempt within the framework of the indefinite mass formalism of Ref. 7 to model the electromagnetic field interacting with a prescribed c-number source. We shall assume that a principle of gauge invariance remains meaningful in the new formalism. Accordingly we expect a gauge invariant interaction Lagrangian in which a potential is coupled to a conserved current. Now in indefinite mass theory we have no conserved 4-current, but we do have a conserved 5-current. Therefore, the postulated interaction Lagrangian will take the form

$$\mathscr{L}_I = -\rho\tfrac{1}{2}V - A_\mu J^\mu, \tag{2.1}$$

in which ρ and J^μ form a conserved 5-current:

$$\frac{\partial\rho}{\partial\tau} + \partial_\mu J^\mu = 0. \tag{2.2}$$

Our model of electromagnetism will be built around the interaction Lagrangian (2.1). It is evident that the potential of the electromagnetic field has acquired a new degree of freedom, $V/2$, not suggested by Lorentz invariance alone; and that our theory will be five-dimensional.

The electromagnetic field tensor F_{AB} is identified with the curl of the five vector potential,

$$\Omega_B \equiv (A_\mu ; \tfrac{1}{2}V), \tag{2.3}$$

$$B = 0, 1, 2, 3, 4,$$

$$F_{AB} \equiv d_A \Omega_B - d_B \Omega_A, \tag{2.4}$$

$$A, B = 0, 1, 2, 3, 4,$$

formed using the five gradient operator:

$$d_A \equiv (\partial_\mu ; \partial_4), \tag{2.5}$$

$$A = 0, 1, 2, 3, 4; \quad \partial_4 \equiv \partial/\partial\tau.$$

Equation (2.4) is equivalent to the pair of equations

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.6}$$

and

$$F_{4\nu} = \dot{A}_\nu - \partial_\nu \tfrac{1}{2}V, \tag{2.7}$$

in which the dot signifies differentiation with respect to evolution time. From Eq. (2.4), the generalized homogeneous Maxwell's equations

$$d_A F_{BC} + d_B F_{CA} + d_C F_{AB} = 0 \tag{2.8}$$

are immediate.

Next we consider the inhomogeneous Maxwell's equations. The option of writing down an electromagnetic theory fully covariant under a 5-dimensional Lorentz group has been suggested earlier by Katayama, Sawada, and Tagagi;[3] and leads to five-dimensional wave equations of the Klein–Gordon type. Because of the rigorous correspondence limit established in Ref. 7 for Fock type indefinite mass equations; we here wish to investigate the possibility of generalizing Maxwell's equations in such a way that the wave equations of the new theory will be of the Fock type. Since we want to describe a real field obeying Bose–Einstein statistics, the second order Fock equation [Eq. (4.21) of Ref. 7, repeated here for convenience]

$$\left\{\left(\frac{\Box^2}{2}\right)^2 + \frac{\partial^2}{\partial\tau^2}\right\}\phi = 0 \tag{2.9}$$

is appropriate. We thus expect a wave equation in our new electromagnetic theory of the general form $O\Omega = \Gamma$, in which Γ is a source term and O is an operator which is second order in the evolution time but fourth order in ∂_μ. A gauge invariant Lagrangian that meets these requirements is

$$\begin{aligned}\mathscr{L} = {} & \tfrac{1}{2}(\dot{A}_\mu - \tfrac{1}{2}\partial_\mu V)(\dot{A}^\mu - \tfrac{1}{2}\partial^\mu V) \\ & - \tfrac{1}{8}\partial^\mu F_{\mu\nu}\partial_\lambda F^{\lambda\nu} - \rho\tfrac{1}{2}V - A_\mu J^\mu.\end{aligned} \tag{2.10}$$

The implied Lagrangian equations of motion are

$$\left\{\left(\frac{\Box^2}{2}\right)^2 + \frac{\partial^2}{\partial\tau^2}\right\}A_\mu = -J_\mu + \tfrac{1}{2}\partial_\mu(\tfrac{1}{2}\Box^2\Box\cdot\mathbf{A} + \dot{V}) \tag{2.11}$$

and

$$\tfrac{1}{2}\Box^2 V = -\rho + \Box\cdot\dot{\mathbf{A}}. \tag{2.12}$$

In a "Lorentz type" gauge,

$$\tfrac{1}{2}\Box^2\Box\cdot\mathbf{A} + \dot{V} = 0, \tag{2.13}$$

the equations of motion (2.11) and (2.12) go over into simple inhomogeneous second order Fock equations

$$\left\{\left(\frac{\Box^2}{2}\right)^2 + \frac{\partial^2}{\partial\tau^2}\right\}A_\mu = -J_\mu \tag{2.14}$$

and

$$\left\{\left(\frac{\Box^2}{2}\right)^2 + \frac{\partial^2}{\partial\tau^2}\right\}V = -\frac{\Box^2}{2}\rho, \tag{2.15}$$

thereby justifying our choice (2.10) of the Lagrangian density. The form of the inhomogeneous equations for the generalized Maxwell field tensor F_{AB} is implicit in Eqs. (2.11) and (2.12). They are quite simply expressed through the use of a "contravariant" analog

$$d^A \equiv ((\Box^2/4)\partial^\mu ; \partial_4), \tag{2.16}$$

$$A = 0, 1, 2, 3, 4,$$

of the five-dimensional gradient operator (2.5). In terms of d^A we can write the inhomogeneous Maxwell's equations as

$$d^A F_{AB} = -\Gamma_B, \tag{2.17}$$

in which

$$\Gamma_B \equiv (J_\mu ; (\Box^2/4)\rho), \tag{2.18}$$

$$B = 0, 1, 2, 3, 4,$$

are the "covariant" components of the five current density. By application of the operator d^B on both sides of Eq. (2.17) and contracting on B the five current conservation law

$$d^B\Gamma_B = 0 \tag{2.19}$$

can be derived [written out in component form Eq. (2.19) reads $(\Box^2/4)(\dot{\rho} + \partial_\mu J^\mu) = 0$]. By working with the operators d_A and d^A we are thus able to create the illusion of imitating the familiar Maxwell theory while staying within the framework of Fock type equations. Of course there is no underlying five-dimensional manifold that could give our terms "covariant" and "contravariant" their usual meaning in the sense of Riemannian geometry.[9] That the terms covariant and contravariant seem apt however, is illustrated again in the identity

$$d^A d_A = \left\{\left(\frac{\Box^2}{2}\right)^2 + \frac{\partial^2}{\partial\tau^2}\right\}, \tag{2.20}$$

showing that the "Laplacian" of our theory is precisely the relevant wave operator of our theory; i.e., the wave operator of the second order Fock equation.

Next we write out our generalized Maxwell's equations in three-dimensional form. We define electric and magnetic field vectors $\mathscr{E}_a$ and $\mathscr{B}_a$ through

$$\mathscr{E}_a \equiv F_{0a}, \quad \mathscr{B}_a \equiv -\tfrac{1}{2}\epsilon_{abc}F_{bc}, \tag{2.21}$$

$$a, b, c = 1, 2, 3.$$

The new components $F_{4\mu}$ of the generalized field tensor are written in terms of a three vector N_a and a scalar W as follows:

$$F_{40} \equiv W, \quad F_{4a} \equiv -N_a, \tag{2.22}$$

$$a = 1, 2, 3.$$

In terms of the fields (2.21) and (2.22) the homogeneous Maxwell's equations (2.8) take the form

$$\nabla\times\mathscr{E} = -\frac{\partial\mathscr{B}}{\partial t}, \tag{2.23}$$

$$\nabla\cdot\mathscr{B} = 0, \tag{2.24}$$

$$\dot{\mathscr{E}} = -\nabla W - \frac{\partial\mathbf{N}}{\partial t}, \tag{2.25}$$

$$\dot{\mathscr{B}} = \nabla\times\mathbf{N}, \tag{2.26}$$

in which the dot signifies differentiation with respect to evolution time.

The new vector $\mathbf{N}$ and scalar W are seen to be a vector and scalar potential describing the change of $\mathscr{E}$ and $\mathscr{B}$ with evolution time. Note, however, that in contrast to the original five potentials (2.3) the potentials $\mathbf{N}$ and W are physical gauge invariant objects. From the inhomogeneous Maxwell's equations (2.17) we get

$$\nabla\cdot\mathbf{N} = \rho - \partial w/\partial t, \tag{2.27}$$

$$\tfrac{1}{4}\Box^2\nabla\cdot\mathscr{E} = -J_0 - \dot{W}, \tag{2.28}$$

and

$$\tfrac{1}{4}\Box^2\left(\nabla\times\mathscr{B} - \frac{\partial\mathscr{E}}{\partial t}\right) = -(\mathbf{J} + \dot{\mathbf{N}}). \tag{2.29}$$

In 4-dimensional form the new Maxwell's equations read

$$\dot{F}_{\mu\nu} = \partial_\mu N_\nu - \partial_\nu N_\mu, \tag{2.30}$$

$$\partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu} = 0, \tag{2.31}$$

$$\partial^\mu N_\mu = \rho, \tag{2.32}$$

and

$$\tfrac{1}{4}\Box^2\partial^\lambda F_{\lambda\mu} = -(J_\mu + \dot{N}_\mu), \tag{2.33}$$

where

$$N_\mu \equiv F_{4\mu}. \tag{2.34}$$

III. SECOND QUANTIZATION

A. Generalized coordinates

We will work in the Lorentz gauge, for which

$$\partial^\mu A_\mu = 0. \tag{3.1}$$

Because of the use of our new evolution parameter τ, the constraint equation (3.1) is holonomic. Accordingly, independent generalized coordinates of the field can be found that obey the constraint equations identically. In this respect quantization in the Lorentz gauge is for us like Coulomb gauge quantization in ordinary quantum electrodynamics. As indicated in the introduction, however, the parallel is not perfect, since we will require the use of an indefinite metric, in contrast to the case of ordinary Coulomb gauge quantization. We use a special symbol $\overline{O}$ to denote the adjoint of an operator O acting on a Hilbert space with indefinite metric. Thus the reality of the 5-potential is expessed in the form

$$\overline{A}_\mu = A_\mu, \quad \tfrac{1}{2}\overline{V} = \tfrac{1}{2}V. \tag{3.2}$$

To find generalized coordinates of the field we go over to momentum space, defining

$$A_\mu = \int \frac{d^4k}{(2\pi)^4|k^2|^{1/2}} \alpha_\mu(\mathbf{k},\tau)e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{3.3}$$

The Fourier amplitudes $\alpha_\mu(\mathbf{k},\tau)$ obey the constraint equations

$$\alpha_\mu(\mathbf{k},\tau) = \overline{\alpha}_\mu(-\mathbf{k},\tau) \tag{3.4}$$

expressing the fact that the field A_μ is real. This constraint equation is equivalent to the two relations

$$\alpha_\mu(\mathbf{k},\tau)_1 = \alpha_\mu(-\mathbf{k},\tau)_1 \tag{3.5a}$$

and

$$\alpha_\mu(\mathbf{k},\tau)_2 = -\alpha_\mu(-\mathbf{k},\tau)_2 \tag{3.5b}$$

on the real and imaginary parts of α_μ:

$$(\alpha_\mu)_1 \equiv \tfrac{1}{2}(\alpha_\mu + \overline{\alpha}_\mu), \quad (\alpha_\mu)_2 \equiv \tfrac{1}{2}i(\overline{\alpha}_\mu - \alpha_\mu).$$

The relations (3.5a) and (3.5b) express the fact that the real and imaginary parts of the Fourier amplitude are even and odd, respectively, in $\mathbf{k}$. This suggests expressing the real and imaginary parts of α_μ as the even and odd part, respectively, of a general real function $Q_\mu(\mathbf{k},\tau)$:

$$\alpha_\mu(\mathbf{k},\tau)_1 \equiv \tfrac{1}{2}(Q_\mu(\mathbf{k},\tau) + Q_\mu(-\mathbf{k},\tau)), \tag{3.6}$$

$$\alpha_\mu(\mathbf{k},\tau)_2 \equiv \tfrac{1}{2}(Q_\mu(\mathbf{k},\tau) - Q_\mu(-\mathbf{k},\tau)). \tag{3.7}$$

The constraint equations (3.5a) and (3.5b) are automatically obeyed if we work in terms of the coordinates Q_μ. In terms of the Q_μ the representation (3.3) becomes

$$A_\mu = \int \frac{d^4k}{(2\pi)^4|k^2|^{1/2}} Q_\mu(\mathbf{k},\tau)f(\mathbf{k},\mathbf{x}), \tag{3.8}$$

in which the functions $f(\mathbf{k},\mathbf{x})$ are

$$f(\mathbf{k},\mathbf{x}) \equiv \tfrac{1}{2}(1+i)e^{-i\mathbf{k}\cdot\mathbf{x}} + \tfrac{1}{2}(1-i)e^{i\mathbf{k}\cdot\mathbf{x}}. \tag{3.9}$$

The real functions $f(\mathbf{k},\mathbf{x})$ form an orthonormal set:

$$\int d^4x f(\mathbf{k}_2,\mathbf{x}) f(\mathbf{k}_1,\mathbf{x}) = (2\pi)^4 \delta^4(\mathbf{k}_2 - \mathbf{k}_1). \tag{3.10}$$

The representation (3.8) is a general representation of a real field, A_μ. For the case at hand we require the further relation

$$k^\mu Q_\mu(\mathbf{k},\tau) = 0, \tag{3.11}$$

expressing the Lorentz gauge condition (3.1). To integrate the constraint equation (3.11) we introduce three real orthonormal basis vectors $\mathbf{e}_a(\mathbf{k})$ ($a = 1, 2, 3$ for $k^2 > 0$; $a = 0, 1, 2$ for $k^2 < 0$) orthogonal to $\mathbf{k}$, and expand Q_μ as

$$Q_\mu(\mathbf{k},\tau) \equiv e_a(\mathbf{k})_\mu q^a(\mathbf{k},\tau). \tag{3.12}$$

The $q^a(\mathbf{k},\tau)$ thereby introduced are true generalized coordinates of the electromagnetic field: the $q^a(\mathbf{k},\tau)$ are real and independent, and the potential

$$A_\mu = \int \frac{d^4k}{(2\pi)^4 |k^2|^{1/2}} e_a(\mathbf{k})_\mu q^a(\mathbf{k},\tau) f(\mathbf{k},\mathbf{x}) \tag{3.13}$$

obeys all the needed constraint equations identically. For future reference we note the relations

$$\mathbf{e}_{a_2}(\mathbf{k})\cdot\mathbf{e}_{a_1}(\mathbf{k}) = \begin{array}{c} \quad a_1 \to \\ \begin{array}{c} a_2 \\ \downarrow \end{array} \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \end{array}$$

$$\equiv \mathfrak{g}_{a_2,a_1}(k^2), \tag{3.14}$$

$k^2 > 0$

and

$$\mathbf{e}_{a_2}(\mathbf{k})\cdot\mathbf{e}_{a_1}(\mathbf{k}) = \begin{array}{c} \quad a_1 \to \\ \begin{array}{c} a_2 \\ \downarrow \end{array} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \end{array}$$

$$\equiv \mathfrak{g}_{a_2,a_1}(k^2), \tag{3.15}$$

$k^2 < 0.$

For timelike $\mathbf{k}$ the $\mathbf{e}_a(\mathbf{k})$ are three spacelike vectors; but for spacelike $\mathbf{k}$ one unit vector, chosen to be $\mathbf{e}_0(\mathbf{k})$, must be timelike.

B. Lagrangian and quantum conditions

It is a straightforward matter to substitute the representation (3.13) into the Lagrangian (2.10) and integrate over all space–time to obtain the total Lagrangian in the form

$$L = \sum \frac{d^4k}{(2\pi)^4|k^2|} \Big\{ \mathfrak{g}_{ab}\Big(\tfrac{1}{2}\dot{q}^a\dot{q}^b - \frac{|k^2|^2}{8} q^a q^b\Big) - q^a \mathbf{e}_a\cdot\mathbf{J}(\mathbf{k},\tau) + \tfrac{1}{2}\frac{k^2}{4}[V(\mathbf{k},\tau)]^2 - \tfrac{1}{2}\rho(\mathbf{k},\tau)V(\mathbf{k},\tau)\Big\}, \tag{3.16}$$

in which $V(\mathbf{k},\tau)$, $\mathbf{J}(\mathbf{k},\tau)$, and $\rho(\mathbf{k},\tau)$ are defined through the equations

$$V(\mathbf{x},\tau) \equiv \int \frac{d^4k}{(2\pi)^4|k^2|^{1/2}} V(\mathbf{k},\tau) f(-\mathbf{k},\mathbf{x}), \tag{3.17}$$

$$\mathbf{J}(\mathbf{x},\tau) \equiv \int \frac{d^4k}{(2\pi)^4|k^2|^{1/2}} \mathbf{J}(\mathbf{k},\tau) f(\mathbf{k},\mathbf{x}), \tag{3.18}$$

and

$$\rho(\mathbf{x},\tau) \equiv \int \frac{d^4k}{(2\pi)^4|k^2|^{1/2}} \rho(\mathbf{k},\tau) f(-\mathbf{k},\mathbf{x}). \tag{3.19}$$

To obtain Eq. (3.16) the relation

$$\partial_\mu f(\mathbf{k},\mathbf{x}) = k_\mu f(-\mathbf{k},\mathbf{x}) \tag{3.20}$$

was required.

Since the velocity $\dot{V}(\mathbf{k},\tau)$ does not appear in the Lagrangian (3.16), the Lagrange equation for V degenerates into a constraint equation. This constraint equation, $k^2\frac{1}{2}V = \rho$, can be integrated to give

$$V(\mathbf{k},\tau) = (2/k^2)\rho(\mathbf{k},\tau). \tag{3.21}$$

Of course, in the Minkowski metric the quantity $1/k^2$ is singular. We must therefore provide a prescription defining the singularity. Since we require $1/k^2$ in Eq. (3.21) to carry a real function into a real function, we assume a prescription such as

$$\frac{1}{k^2} \equiv \frac{1}{2}\Big(\frac{1}{k^2 + i\epsilon} + \frac{1}{k^2 - i\epsilon}\Big). \tag{3.22}$$

This prescription will be implicitly assumed in subsequent occurrences of $1/k^2$.

The conjugate momenta $P_a \equiv \partial L/\partial\dot{q}^a$, where L is the Lagrangian (3.16), are found to be

$$P_a = \frac{d^4k}{(2\pi)^4|k^2|} \mathfrak{g}_{ab}\dot{q}^b. \tag{3.23}$$

Canonical quantum conditions can be imposed in a straightforward way, the only nonzero commutators being

$$[q^a(\mathbf{k}_2,\tau); P_b(\mathbf{k}_1,\tau)] = i\delta^a{}_b \delta_{\mathbf{k}_2,\mathbf{k}_1}. \tag{3.24}$$

Substituting the expression (3.23) for P_a into (3.24), and passing to the continuum limit, gives

$$[q^a(\mathbf{k}_2,\tau); \dot{q}^b(\mathbf{k}_1,\tau)] = i|k^2{}_1|\mathfrak{g}^{ab}(2\pi)^4\delta^4(\mathbf{k}_2 - \mathbf{k}_1) \tag{3.25}$$

$$\mathfrak{g}^{ab} \equiv \mathfrak{g}_{ab}.$$

The Hamiltonian $H \equiv \Sigma\dot{q}^a\partial L/\partial\dot{q}^a - L$, where L is the Lagrangian (3.16) has the form

$$H = H_0 + H_{\rm INT} \tag{3.26}$$

in which

$$H_0 = \int \frac{d^4k}{(2\pi)^4|k^2|} \mathfrak{g}_{ab}\Big(\tfrac{1}{2}\dot{q}^a\dot{q}^b + \frac{|k^2|^2}{8} q^a q^b\Big) \tag{3.27}$$

is the free photon Hamiltonian, and

$$H_{\rm INT} = \int \frac{d^4k}{(2\pi)^4|k^2|} \Big(q^a e_a\cdot J + \frac{\rho^2}{2k^2}\Big) \tag{3.28}$$

is the interaction Hamiltonian.

We introduce raising and lowering operators that provide a factorization of the free photon Hamiltonian (3.27):

$$\mathfrak{a}^a(\mathbf{k},\tau) \equiv \tfrac{1}{2}q^a(\mathbf{k},\tau) - (i/|k^2|)\dot{q}^a(\mathbf{k},\tau), \tag{3.29}$$

$$\bar{\mathfrak{a}}^a(\mathbf{k},\tau) = \tfrac{1}{2}q^a(\mathbf{k},\tau) + (i/|k^2|)\dot{q}^a(\mathbf{k},\tau). \tag{3.30}$$

The commutation relations for the creation and annihilation operators are

$$[\mathfrak{a}^a(\mathbf{k}_2,\tau); \bar{\mathfrak{a}}^b(\mathbf{k}_1,\tau)] = -\mathfrak{g}^{ab}(2\pi)^4\delta^4(\mathbf{k}_2 - \mathbf{k}_1). \tag{3.31}$$

In view of the relation $\mathfrak{g}^{ab} = \mathfrak{g}_{ab}$ and Eqs. (3.14) and (3.15), it is apparent that the sign of the commutator in Eq. (3.31) is positive except for spacelike 4-momenta with timelike polarization vector $\mathbf{e}_0$. Therefore the operators $\mathfrak{a}^a,\bar{\mathfrak{a}}^a$ are clearly lowering and raising operators, respectively; except for the "bad" states mentioned above having spacelike 4-momentum and timelike polarization vectors $\mathbf{e}_0$. Now in order to obtain a Lorentz invariant propagator, we find it necessary to continue to treat $\mathfrak{a}^0$ as an annihilation operator. It is at this point that we encounter the need for a Hilbert space with an indefinite metric. In order to continue to treat $\mathfrak{a}^0$ as an annihilation operator in the face of the commutation relation

$$[\mathfrak{a}^0(\mathbf{k}_2,\tau);\bar{\mathfrak{a}}^0(\mathbf{k}_1,\tau)] = -(2\pi)^4\delta^4(\mathbf{k}_2-\mathbf{k}_1), \qquad (3.32)$$

$$k^2{}_2<0,\ k^2{}_1<0;$$

we assume an indefinite metric in our Hilbert space, along the lines of Gupta and Bleuler. For our present application we need not enter into great mathematical detail on the indefinite metric,[10] but it is worth pointing out that in this Gupta–Bleuler scheme it is not $\bar{\mathfrak{a}}^0\mathfrak{a}^0$ but its negative, $-\bar{\mathfrak{a}}^0\mathfrak{a}^0$, that is a positive definite operator. As a result all states, including the bad states, contribute $-\frac{1}{2}|k^2|$ to the free photon Hamiltonian (3.27),

$$H_0=\int\frac{d^4k}{(2\pi)^4}\frac{|k^2|}{2}\,\mathfrak{g}_{ab}\bar{\mathfrak{a}}^a\mathfrak{a}^b. \qquad (3.33)$$

The free photon Hamiltonian in this scheme is thus a negative definite operator.

C. Traveling wave representation, return to coordinate space

The canonical variables having served their purpose of determining the basic commutator structure of the theory; we now make a canonical transformation to a more convenient representation. We solve Eqs. (3.29) and (3.30) for q^a and $\dot{q}^a$ in the form

$$q^a=\bar{\mathfrak{a}}^a+\mathfrak{a}^a \qquad (3.34)$$

and

$$\dot{q}^a=\tfrac{1}{2}i|k^2|(\mathfrak{a}^a-\bar{\mathfrak{a}}^a). \qquad (3.35)$$

Using the representation (3.13) of A_μ in terms of the canonical variables, we find the following traveling wave representations of A_μ and $\dot{A}_\mu$:

$$A_\mu=\int\frac{d^4k}{(2\pi)^4|k^2|^{1/2}}\,e_a(\mathbf{k})_\mu\{\alpha^a(\mathbf{k},\tau)e^{-i\mathbf{k}\cdot\mathbf{x}}+\bar{\alpha}^a(\mathbf{k},\tau)e^{i\mathbf{k}\cdot\mathbf{x}}\} \qquad (3.36)$$

and

$$\dot{A}_\mu=\int\frac{d^4k}{(2\pi)^4|k^2|^{1/2}}\,e_a(\mathbf{k})_\mu\tfrac{1}{2}i|k^2| \times\{\alpha^a(\mathbf{k},\tau)e^{-i\mathbf{k}\cdot\mathbf{x}}-\bar{\alpha}^a(\mathbf{k},\tau)e^{i\mathbf{k}\cdot\mathbf{x}}\}; \qquad (3.37)$$

in which

$$\alpha^a(\mathbf{k},\tau)\equiv\mathfrak{a}^a(\mathbf{k},\tau)\tfrac{1}{2}(1+i)+\epsilon_a(k^2)\mathfrak{a}^a(-\mathbf{k},\tau)\tfrac{1}{2}(1-i) \qquad (3.38)$$

and

$$\bar{\alpha}^a(\mathbf{k},\tau)=\bar{\mathfrak{a}}^a(\mathbf{k},\tau)\tfrac{1}{2}(1-i)+\epsilon_a(k^2)\bar{\mathfrak{a}}^a(-\mathbf{k},\tau)\tfrac{1}{2}(1+i). \qquad (3.39)$$

In these equations $\epsilon_a(k^2)$ is the parity of $\mathbf{e}_a(\mathbf{k})$:

$$e_a(-\mathbf{k})\equiv\epsilon_a(k^2)\mathbf{e}_a(\mathbf{k}). \qquad (3.40)$$

It can be verified that the new operators introduced through Eqs. (3.38) and (3.39) are again a set of raising and lowering operators:

$$[\alpha^a(\mathbf{k}_2,\tau);\bar{\alpha}^b(\mathbf{k}_1,\tau)]=-\mathfrak{g}^{ab}(2\pi)^4\delta^4(\mathbf{k}_2-\mathbf{k}_1). \qquad (3.41)$$

The free particle Hamiltonian (3.33) goes over into a similar structure,

$$H_0=\int\frac{d^4k}{(2\pi)^4}\frac{|k^2|}{2}\,\mathfrak{g}_{ab}\bar{\alpha}^a\alpha^b, \qquad (3.42)$$

when expressed in the new representation.

The one-evolution-time coordinate space commutation relations,

$$[A_\mu(2);\dot{A}_\nu(1)]=i\left(g_{\mu\nu}-\frac{\partial_{2\mu}\partial_{2\nu}}{\Box^2{}_2}\right)\delta^4(2,1), \qquad (3.43)$$

can be derived quite simply by use of the traveling wave representations (3.36), (3.37), and the commutation relations (3.41) for the raising and lowering operators. We have here returned to the coordinate space picture with which we started. For completeness we also present the free photon Hamiltonian (3.27) and the interaction Hamiltonian (3.28) in coordinate space:

$$H_0=\int d^4x\{\tfrac{1}{2}\dot{A}_\mu\dot{A}^\mu+\tfrac{1}{8}(\Box^2A_\mu)\Box^2A^\mu\} \qquad (3.44)$$

and

$$H_{\rm INT}=\int d^4x\left\{J^\mu A_\mu-\tfrac{1}{2}\rho\frac{1}{\Box^2}\rho\right\}. \qquad (3.45)$$

IV. VACUUM-VACUUM AMPLITUDE AND FREE FIELD PROPAGATOR

It is straightforward to obtain the equation of motion $\dot{\alpha}^a=i\frac{1}{2}|k^2|\alpha^a$ of the lowering operator in the free field case. The evolution-time development of the free field lowering operator is thus

$$\alpha^a(\mathbf{k},\tau)=\alpha^a(\mathbf{k},0)e^{i|k^2|\tau/2}, \qquad (4.1)$$

in which the initial values $\alpha^a(\mathbf{k},0),\bar{\alpha}^a(\mathbf{k},0)$ are a set of evolution-time independent raising and lowering operators obeying the commutation relations (3.41). The representation (3.36) can now be rewritten in the form

$$A_\mu=\int\frac{d^4k}{(2\pi)^4|k^2|^{1/2}}\,e_a(\mathbf{k})_\mu\{\alpha^a(\mathbf{k},0)e^{-i\mathbf{k}\cdot\mathbf{x}+i|k^2|\tau/2} + \bar{\alpha}^a(\mathbf{k},0)e^{i\mathbf{k}\cdot\mathbf{x}-i|k^2|\tau/2}\}, \qquad (4.2)$$

which exhibits the evolution-time dependence for free fields explicitly. Next we consider the propagator

$$\begin{aligned}iD_F(2,1)_{\mu\nu}&\equiv\langle0|T(A_\mu(2)A_\nu(1))|0\rangle\\&\equiv\theta(\tau_2-\tau_1)\langle0|A_\mu(2)A_\nu(1)|0\rangle\\&\quad+\theta(\tau_1-\tau_2)\langle0|A_\nu(1)A_\mu(2)|0\rangle,\end{aligned} \qquad (4.3)$$

in which A_μ is given by (4.2). Since the calculation of a vacuum expectation value such as (4.3) is quite standard,[8] we omit the details. The end result is most simply expressed in momentum space. We define

$$iD_F(2,1)_{\mu\nu} \equiv \int \frac{d^4k}{(2\pi)^4} \times \int \frac{dN}{4\pi} iD_F(\mathbf{k},N)_{\mu\nu} e^{-i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)+iN(\tau_2-\tau_1)/2}; \tag{4.4}$$

and find

$$iD_F(\mathbf{k},N)_{\mu\nu} = (-g_{\mu\nu} + k_\mu k_\nu / k^2) iD_F(\mathbf{k},N), \tag{4.5}$$

in which $iD_F(\mathbf{k},N)$ (without Lorentz indices) is the scalar propagator found in Ref.7,[11]

$$iD_F(\mathbf{k},N) = \frac{i}{(k^2/2)^2 - (N/2)^2 + i\epsilon}. \tag{4.6}$$

In coordinate space this scalar propagator is a solution of the inhomogeneous second order Fock equation

$$\left[(\tfrac{1}{2}\Box^2{}_2)^2 + \frac{\partial^2}{\partial\tau^2{}_2}\right] D_F(2,1) = \delta^4(\mathbf{x}_2 - \mathbf{x}_1)\delta(\tau_2 - \tau_1). \tag{4.7}$$

The additional term, proportional to $k_\mu k_\nu$ in the expression (4.5) for the full tensor photon propagator is the analog of an additional $k_i k_j$ term in the transverse photon propagator of ordinary Coulomb gauge quantum electrodynamics. In ordinary Coulomb gauge quantum electrodynamics the additional term mars the manifest Lorentz invariance of the theory. However, as is known, manifest invariance can be restored in Feynman integrals, since the additional term can be eliminated by use of current conservation and suitable integrations by parts. For the case at hand the additional $k_\mu k_\nu$ terms pose no threat to manifest Lorentz invariance. On the other hand, it is expected that in Feynman integrals the additional terms could still be eliminated, with some resultant simplification of the propagator. In the following we illustrate this for the simple example of the vacuum–vacuum amplitude in the theory with c-number source terms restored.

In order to calculate the amplitude for the vacuum to remain a vacuum in the original theory with c-number source terms present, it is necessary to go over to an interaction representation. Our interaction picture parallels exactly the interaction picture familiar from ordinary field theory, and differs from the latter only in our use of the new evolution parameter τ. The vacuum–vacuum amplitude thus has the familiar structure

$$\langle 0,\text{out}|0,\text{in}\rangle = \langle 0,\text{in}|S|0,\text{in}\rangle, \tag{4.8}$$

in which

$$S \equiv T\left(\exp\left(-i\textstyle\int_{-\infty}^{\infty} d\tau H_{int}\right)\right) \tag{4.9}$$

is the S matrix. If we substitute the expression (3.45) for H_{int}, Eq. (4.8) becomes

$$\langle 0,\text{out}|0,\text{in}\rangle = \langle 0,\text{in}|T\left\{\exp\left[-i\textstyle\int_{-\infty}^{\infty}(d\tau)\left(\mathrm{J}_\mu \mathrm{A}^\mu{}_{\text{in}} - \tfrac{1}{2}\rho\frac{1}{\Box^2}\rho\right)\right]\right\}|0,\text{in}\rangle. \tag{4.10}$$

Here we have employed the shorthand notation

$$\int (d\tau)\cdots \equiv \int d\tau \int d^4x \cdots. \tag{4.11}$$

Here $\mathbf{A}_{in}$ is the free field that reduces to the Heisenberg field in the remote past as regards evolution time. Since the factor $\exp(i\int(d\tau)\frac{1}{2}\rho\Box^{-2}\rho)$ has no operator structure, it can be brought outside the time ordering sign, and outside the matrix element. The factor

$$\langle 0,\text{in}|T\{\exp(-i\textstyle\int(d\tau)\mathrm{J}_\mu \mathrm{A}^\mu{}_{\text{in}})\}|0,\text{in}\rangle$$

is evaluated by expanding the exponential, using Wick's theorem, term by term, and summing the resultant series of c-numbers. The end product of these transformations is

$$\langle 0,\text{out}|0,\text{in}\rangle = \exp\left[-\tfrac{1}{2}i\int(d\tau_1)(d\tau_2)J^\mu(2)D_F(2,1)_{\mu\nu}J^\nu(1) + i\int(d\tau)\tfrac{1}{2}\rho\Box^{-2}\rho\right]. \tag{4.12}$$

The photon propagator $iD_F(2,1)_{\mu\nu}$ in Eq. (4.12) originated from the Wick pairing of two photon fields in accordance with Eq. (4.3). Next we substitute

$$iD_F(2,1)_{\mu\nu} = (-g_{\mu\nu} - \partial_{2\mu}\partial_{1\nu}\Box^{-2}{}_2) iD_F(2,1)$$

[a coordinate space form of Eq. (4.5)] in Eq. (4.12) and move the factors $\partial_{2\mu}$ and $\partial_{1\nu}$ to the currents by integration by parts. The 5-current conservation law (2.2) and two further integrations by parts then gives

$$\langle 0,\text{out}|0,\text{in}\rangle = \exp\left[-\tfrac{1}{2}i\int(d\tau_1)(d\tau_2)J^\mu(2)(-g_{\mu\nu})D_F(2,1)J^\nu(1) - \tfrac{1}{2}i\int(d\tau_1)(d\tau_2)\rho(2)\left[\Box^{-2}{}_2\left(\frac{\partial}{\partial\tau_2}\right)^2 D_F(2,1)\right]\rho(1) + i\int(d\tau)\tfrac{1}{2}\rho\Box^{-2}\rho\right]. \tag{4.13}$$

A final step in the reduction uses the differential equation (4.7) of the scalar propagator to eliminate the operator $(\partial/\partial\tau_2)^2$ in (4.13). The delta function term from the differential equation is found to exactly cancel the last factor $\exp(i\int(d\tau)\frac{1}{2}\rho\Box^{-2}\rho)$, and we obtain the rather simple result

$$\langle 0,\text{out}|0,\text{in}\rangle = \exp\left\{-\tfrac{1}{2}i\int(d\tau_1)(d\tau_2)\left[J_\mu(2)(-g^{\mu\nu})D_F(2,1)J_\nu(1) - \rho(2)\tfrac{1}{4}\Box^2{}_2 D_F(2,1)\rho(1)\right]\right\}. \tag{4.14}$$

This result justifies for this example the above suggestion that the $k_\mu k_\nu$ term is the propagator (4.5) can be dropped in Feynman integrals. Note, however, the attendant modification of the ρ dependent terms in the amplitude. The transition from (4.12) to (4.14) parallels exactly the transition in ordinary field theory from a Coulomb gauge Feynman integral to an equivalent Feynman integral in the Lorentz gauge. To exhibit this parallel more explicitly we cast Eq. (4.14) in 5-dimensional form as follows:

$$\langle 0,\text{out}|0,\text{in}\rangle = \exp\left\{-\tfrac{1}{2}i\int(d\tau_1)(d\tau_2)\Gamma_A(2)D_F(2,1)^{AB}\Gamma_B(1)\right\}. \tag{4.15}$$

Here Γ_A denotes the 5-current (2.18), and the components of the propagator $D_F(2,1)^{AB}$ are

$$D_F(2,1)^{AB} \equiv \begin{pmatrix} -g^{\mu\nu} & 0 \\ 0 & -4/\Box^2_2 \end{pmatrix} D_F(2,1). \tag{4.16}$$

It is the 5-dimensional propagator (4.16) that corresponds in the present formalism to the Lorentz gauge propagator of conventional quantum electrodynamics.

APPENDIX A: PHYSICAL MEANING OF THE 5-CURRENT CONSERVATION LAW

Observationally, we are bound to view phenomena as "evolving" in observer's time. We shall therefore attempt to interpret the 5-dimensional conservation law [Eq. (2.2) rewritten here for convenience]

$$\frac{\partial \rho}{\partial \tau} + \frac{\partial J^0}{\partial x^0} + \nabla\cdot\mathbf{J} = 0 \tag{A1}$$

from this point of view. Our interpretation will be in essentially classical terms. First we integrate (A1) over a volume V of 3-space and over an interval $\tau_1 < \tau < \tau_2$ of evolution time. The result is

$$I = II + III, \tag{A2}$$

where

$$I = \frac{\partial}{\partial x^0}\int_V d^3r \int_{\tau_1}^{\tau_2} d\tau J^0, \tag{A3}$$

$$\mathrm{II} = \iint_{\partial V} (-d\alpha)\cdot\int_{\tau_1}^{\tau_2} d\tau \mathbf{J}, \tag{A4}$$

and

$$\mathrm{III} = -\int_V d^3r\rho \Big|_{\tau_1}^{\tau_2}. \tag{A5}$$

Figure 1 suggests a photograph of the particles in V taken by an observer at one observer's time, x^0. Each particle is equipped with a clock reading its own evolution time, τ. Because of the indefiniteness of the relationship between observer's time and evolution time, the clocks show a statistical distribution of values of evolution time. We denote by $S(\tau_1, \tau_2)$ the subset of particles in V whose clocks read in the range $\tau_1 < \tau < \tau_2$. In Fig. 1 these clocks are represented by solid circles.

We know that the usual four-current density obeying $\partial_\mu j^\mu = 0$ is

$$j^\mu = \int_{-\infty}^{\infty} d\tau J^\mu.$$

We assume this usual four-current density to describe the total four-current of all particles in V, irrespective of the values of their evolution times. The representation of j^μ as an integral suggests that $d\tau J^\mu$ describes the partial four-current arising solely from particles in V, whose evolution times lie in the range $(\tau, \tau + d\tau)$. Accordingly, we interpret the integral

$$\int_{\tau_1}^{\tau_2} d\tau J^\mu,$$

which appears in expressions I and II, as the partial current due to only the particles present in V that belong to the subset $S(\tau_1, \tau_2)$.

Now consider the terms I, II, and III of Eq. (A2). Term I

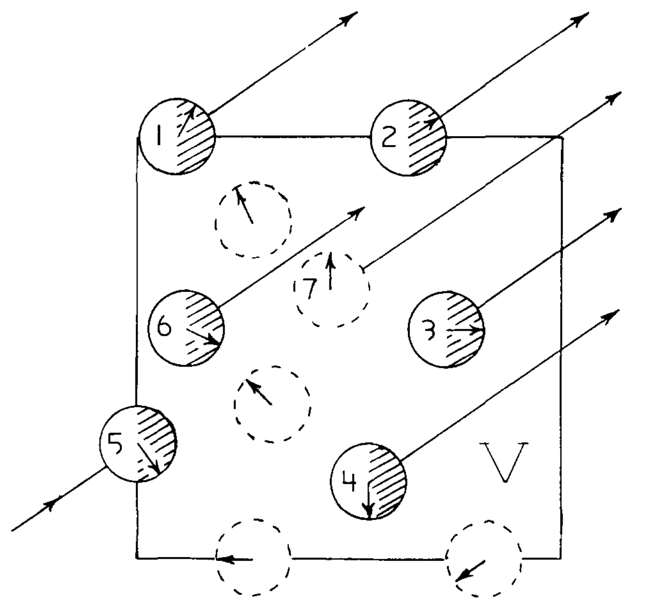


FIG. 1. Five-dimensional current conservation law. Each particle is equipped with a clock reading its own evolution time. The current density $\int_{\tau_1}^{\tau_2} d\tau J^\mu$ refers to only a subset of particles in V (see text discussion), depicted here as the subset S, whose evolution times lie in the range "12 o'clock" to "6 o'clock". In addition to simple transport across the boundary, the number $\int_V d^3r \int_{\tau_1}^{\tau_2} d\tau J^0$ of particles in S can change if the evolution times of interior particles change in such a way as to fall within the requisite range. This "selection" mechanism accounts for the term $-\partial\rho/\partial\tau$ in the current conservation law $\partial J^0/\partial x^0 = -\nabla\cdot\mathbf{J} - \partial\rho/\partial\tau$.

is the (net) rate of increase per unit of observer's time of the population of the subset $S(\tau_1, \tau_2)$. There are two mechanisms that contribute to this rate of increase. There is the simple transport into or out of V of particles having their evolution times in the requisite range (particles 1, 2, and 5 in Fig. 1) to belong to $S(\tau_1, \tau_2)$. The rate per unit of observer's time due to this transport mechanism is represented on the right hand side of Eq. (A2) by the term II.

Now consider particle 7 in Fig. 1. This particle is within V but at observer's time x^0 does not have its evolution time in the requisite range to belong to $S(\tau_1, \tau_2)$. Classically speaking, however, for any one particle a small increase in oberserver's time will imply a small increase of evolution time.

By observer's time $x^0 + dx^0$, therefore, particle 7 will have its evolution time in the requisite range and, being still within V, will join $S(\tau_1, \tau_2)$. Particle 4 is an example of the reverse process. We have here a second mechanism, a "selection" mechanism, whereby particles can join or leave the subset $S(\tau_1, \tau_2)$.

We can calculate the rate at which particles are added to $S(\tau_1, \tau_2)$ through the selection mechanism. We have an interpretation of ρ as a space–time probability density. From the standpoint of our observer, who is bound to see events as evolving in observer's time, the quantity

$$dx^0 \int_V d^3r\rho(\mathbf{x},\tau_1)$$

will measure the number of additional particles having evolution time τ_1 that are seen during the interval of observer's time dx^0. But these are precisely the particles that will join $S(\tau_1, \tau_2)$ during the interval dx^0 of observer's time. Dividing by dx^0 gives

$$\int_V d^3r\rho(\mathbf{x},\tau_1)$$

equals the rate, per unit of observer's time, of additions to $S(\tau_1, \tau_2)$ due to the selection mechanism. This accounts for the lower limit of the term III in the current conservation law (A2). Similarly, the rate at which particles leave $S(\tau_1, \tau_2)$ through the selection mechanism is

$$\int_V d^3r\rho(\mathbf{x},\tau_2).$$

This accounts for the part of III corresponding to the upper limit.

APPENDIX B: SCATTERING CROSS SECTIONS

Armed with the physical picture of the 5-dimensional current conservation law developed in Appendix A, we are now in a position to calculate scattering cross sections. The first order Fock equation of Ref. 7 will be used as an example. For simplicity we use first quantization techniques. The wave equation t · be solved [Eq. (3.16) of Ref. 7] is

$$\left(\tfrac{1}{2}\Box^2 + V + \frac{1}{i}\frac{\partial}{\partial\tau}\right)\phi = 0, \tag{B1}$$

$$V \equiv \tfrac{1}{2}ie(A^\mu\partial_\mu + \partial_\mu A^\mu) - \tfrac{1}{2}e^2A^\mu A_\mu.$$

In order to establish contact with the familiar results of conventional quantum theory of particles of definite mass, we neglect the new 5th component of the potential and assume that A_μ is independent of evolution time. Otherwise A_μ can be an arbitrary space–time function.

Equation (B1) can be integrated using the standard lore of scattering theory, but with τ as evolution parameter instead of x^0. Thus, the first step is to convert Eq. (B1) into an equivalent integral equation:

$$\phi^+{}_i(2) = \phi_i(2) + \int d^4x_1 d\tau_1 G_0(2,1)V(1)\phi^+{}_i(1). \tag{B2}$$

Here $\phi^+{}_i(2)$ denotes the exact solution of the wave Eq. (B1) that reduces in the remote past of evolution time to the plane wave $\phi_i(2)$. The function $G_0(2,1)$ is the free particle Green's function of Ref. 7 with the momentum space representation [Eq. (4.19) of Ref. 7, but with a new normalization]

$$G_0 \equiv 1/[\tfrac{1}{2}p^2 - \tfrac{1}{2}N(1 - i\epsilon)], \quad \epsilon > 0. \tag{B3}$$

The probability amplitude to see the state ϕ_f in the remote future of evolution time is given by the inner product

$$S_{fi} = \int d^4x_2\phi_f{}^*(2)\phi_i{}^+(2). \tag{B4}$$

It will be convenient to transform (B4) using standard reduction techniques[8] into the form

$$S_{fi} = -i\int d^4x_2 d\tau_2\phi_f{}^*(2)\left(-\tfrac{1}{2}\Box^2{}_2 - \frac{1}{i}\frac{\partial}{\partial\tau_2}\right)\phi_i{}^+(2), \quad f \neq i. \tag{B5}$$

In order to guarantee the interpretation of the dot product (B4) as a probability amplitude, it is necessary to require the plane wave states to be orthonormal with respect to the metric $(\phi_b;\phi_a) = \int d^4x\phi_b{}^*(\mathbf{x})\phi_a(\mathbf{x})$ appropriate to the wave equation (B1). Accordingly, the plane wave solutions of the free particle Fock equation are written

$$\phi_i(\mathbf{x}) = e^{-i\mathbf{P}\cdot\mathbf{x} + im^2\tau/2}/(VT)^{1/2}, \tag{B6}$$

$$P^0 \equiv +(m^2 + P_aP_a)^{\frac{1}{2}},$$

(electrons);

and

$$\phi_i(\mathbf{x}) = e^{i\mathbf{P}\cdot\mathbf{x} + i\frac{1}{2}m^2\tau}/(VT)^{1/2}, \tag{B7}$$

$$P^0 \equiv +(m^2 + P_aP_a)^{\frac{1}{2}},$$

(positrons); and similarly for ϕ_f. Periodic boundary conditions in a space–time box are assumed. Next, we need the current of the initial state (B6) or (B7). The discussion of Appendix A indicates that the experimentally observed current will be the total current of all particles seen at one observer's time, irrespective of the values of their evolution times: $\mathbf{J} = \int d\tau\tfrac{1}{2}\phi_i{}^*(-i)\overleftrightarrow{\nabla}\phi_i$. Substituting either (B6) or (B7), we find

$$J_a = \pm(\Lambda/VT)(P_a)_i, \tag{B8}$$

in which Λ represents the duration in evolution time of the scattering.

Note that we did not encounter an overall sign change in Eq. (B5) for negative frequency final states, in contrast to the corresponding situation in conventional quantum mechanics of particles of definite mass. This absence of a sign change reflects the fact that the boundary conditions in evolution time are strictly retarded boundary conditions for all timelike states. Accordingly, the initial state at $\tau = -\infty$ can for us be either an electron state (B6) or a positron state (B7). It is known[2,7] that these strictly retarded boundary conditions in evolution time for the timelike states imply the usual Feynman boundary conditions in observer's time. Evolution time $\tau = -\infty$ thus corresponds to observer's time $x^0 = -\infty$ for electrons and to observer's time $x^0 = +\infty$ for positrons.[12] In either case $\tau = -\infty$ is the initial point on the world line of the particle in a conventional Feynman diagram. Similarly, $\tau = +\infty$ is always the final point on the world line in a conventional Feynman diagram.

To proceed, we solve Eq. (B2) by iteration in the usual way, obtaining

$$\phi_i{}^+(2) = \int d^4x_1 d\tau_1 G(2,1)\left(-\tfrac{1}{2}\Box^2{}_1 - (1 - i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_1}\right) \times \phi_i(1), \tag{B9}$$

in which $G(2,1)$ is the fully interacting Green's function of our wave equation (B1):

$$G \equiv \left(\tfrac{1}{2}\Pi_\mu\Pi^\mu - (1 - i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau}\right)^{-1}. \tag{B10}$$

Substituting the result (B9) into Eq. (B5) for the transition amplitude gives

$$S_{fi} = -i\int d^4x_2 d\tau_2 d^4x_1 d\tau_1 \phi_f^*(2)$$
$$\times\left(-\tfrac{1}{2}\Box^2{}_2 - (1-i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_2}\right)$$
$$\times G(2,1)\left(-\tfrac{1}{2}\Box^2{}_1 - (1-i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_1}\right)\phi_i(1), \quad \text{(B11)}$$

$f\neq i$.

Under our assumption that the 4-potential shall not contain the evolution time, the operator sandwiched between the states ϕ_f^* and ϕ_i in Eq. (B11) is diagonal in the mass-squared. We can thus perform the τ_1 and τ_2 integrations in Eq. (B11) explicitly. Substituting plane wave states and integrating leads to

$$S_{fi} = -\frac{i}{VT}\int d^4x_2 d^4x_1 \int d\tau_2 d\tau_1$$
$$\times e^{i\epsilon_f \mathbf{P}_f\cdot\mathbf{x}_2 - im^2{}_f\tau_2/2}\left(-\tfrac{1}{2}\Box^2{}_2 - (1-i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_2}\right)$$
$$\times\int\frac{dN}{4\pi}e^{iN(\tau_2-\tau_1)/2}G(\mathbf{x}_2,\mathbf{x}_1,N)$$
$$\times\left(-\tfrac{1}{2}\Box^2{}_1 - (1-i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_1}\right)e^{-i\epsilon_i\mathbf{P}_i\cdot\mathbf{x}_1 + im^2{}_i\tau_1/2},$$

which reduces to

$$S_{fi} = -i\frac{4\pi\delta(m^2{}_f - m^2{}_i)}{VT}M_{fi}, \quad \text{(B12)}$$

$$M_{fi}\equiv\int d^4x_2 d^4x_1 e^{i\epsilon_f\mathbf{P}_f\cdot\mathbf{x}_2}T(2,1,m^2{}_i)e^{-i\epsilon_i\mathbf{P}_i\cdot\mathbf{x}_1}, \quad \text{(B13)}$$

$f\neq i$, $\epsilon_{i,f}\equiv 1$(electrons), $\epsilon_{i,f}\equiv -1$(positrons),

$$T(2,1,m^2{}_i)\equiv\left(-\tfrac{1}{2}\Box^2{}_2 - (1-i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_2}\right)G(\mathbf{x}_2,\mathbf{x}_1,m^2{}_i)$$
$$\times\left(-\tfrac{1}{2}\Box^2{}_1 - (1-i\epsilon)\frac{1}{i}\frac{\partial}{\partial\tau_1}\right). \quad \text{(B14)}$$

In the derivation leading to Eq. (B12) the τ_1 integration gives a factor $4\pi\delta(N - m^2{}_i)$. The N integration then collapses, with a replacement of N by $m_i{}^2$ throughout. The factor $4\pi\delta(m^2{}_f - m^2{}_i)$ results from the final τ_2 integration. The mass-squared representation $G(\mathbf{x}_2,\mathbf{x}_1,m^2{}_i)$ of the fully interacting propagator $G(2,1)$ is defined through the equation

$$G(2,1)\equiv\int\frac{dN}{4\pi}e^{iN(\tau_2-\tau_1)/2}G(\mathbf{x}_2,\mathbf{x}_1,N). \quad \text{(B15)}$$

The Green's function $G(\mathbf{x}_2,\mathbf{x}_1,N)$ is two times the familiar interacting Green's function of the Klein–Gordon equation with mass-squared N:

$$\{\tfrac{1}{2}\Pi^2{}_2 - \tfrac{1}{2}(1-i\epsilon)N\}G(\mathbf{x}_2,\mathbf{x}_1,N) = \delta^4(\mathbf{x}_2-\mathbf{x}_1). \quad \text{(B16)}$$

To go over from a probability amplitude to an actual probability we take the absolute square of Eq. (B12). Using the replacement $4\pi\delta(m^2=0) = \int d\tau = \Lambda$, we find

$$|S_{fi}|^2 = \frac{4\pi\delta(m^2{}_f - m^2{}_i)\Lambda}{V^2T^2}|M_{fi}|^2, \quad \text{(B17)}$$

f≠i.

Next, we divide by T to obtain the transition probability per unit of observer's time, and divide again by the magnitude of the incident current (B8), in order to convert the transition probability per unit of observer's time into an equivalent differential cross section

$$d\sigma' = \frac{4\pi\delta(m^2{}_f - m^2{}_i)}{VT^2P_i}|M_{fi}|^2, \quad f\neq i. \quad \text{(B18)}$$

Note the complete disappearance of the evolution time from the calculated differential cross section: the expression (B18) involves the familiar space–time degrees of freedom only. The mass-squared conserving delta function in Eq. (B18) will have the effect of limiting the scattering of a timelike initial state to timelike final states of the same mass-squared. To get the scattering cross section familiar from conventional quantum theory of particles of definite mass, we sum $\delta\sigma'$ over a small range of final mass-squared values centered on the value $m^2{}_i$. There are $d^4n_f = VTd^4p_f/(2\pi)^4$ states in the range d^4p_f of the 4-dimensional final energy-momentum space. Using the identity $d^4p = dm^2d^3p/2E$, we find

$$d\sigma\equiv d^4n_f d\sigma' = (VT/(2\pi)^4)(d^3p_f/2E_f)$$
$$\times dm^2{}_f 4\pi\delta(m^2{}_f - m^2{}_i)|M_{fi}|^2/(VT^2P_i).$$

This simplifies to

$$d\sigma = \frac{d^3P_f}{(2\pi)^3E_fP_i}\frac{|M_{fi}|^2}{T}, \quad f\neq i. \quad \text{(B19)}$$

Equation (B19) is our final answer for the scattering cross section for scattering of timelike states of mass m_i into a final state whose 3-momentum $\mathbf{P}_f$ lies in the range d^3P_f.

It is worthwhile writing out explicitly the matrix element (B13) for the various possible processes. From Eq. (B13) we find

$$M_{fi} = \int d^4x_2 d^4x_1 e^{i\mathbf{P}_f\cdot\mathbf{x}_2}T(2,1,m^2{}_i)e^{-i\mathbf{P}_i\cdot\mathbf{x}_1} \quad \text{(B20)}$$

(electron-electron scattering, $f\neq i$),

$$M_{fi} = \int d^4x_2 d^4x_1 e^{-i\mathbf{P}_f\cdot\mathbf{x}_2}T(2,1,m^2{}_i)e^{-i\mathbf{P}_i\cdot\mathbf{x}_1} \quad \text{(B21)}$$

(pair annihilation),

$$M_{fi} = \int d^4x_2 d^4x_1 e^{i\mathbf{P}_f\cdot\mathbf{x}_2}T(2,1,m^2{}_i)e^{i\mathbf{P}_i\cdot\mathbf{x}_1} \quad \text{(B22)}$$

(pair production),
and

$$M_{fi} = \int d^4x_2 d^4x_1 - e^{-i\mathbf{P}_f\cdot\mathbf{x}_2}T(2,1,m^2{}_i)e^{i\mathbf{P}_i\cdot\mathbf{x}_1} \quad \text{(B23)}$$

(positron-positron scattering, $f\neq i$).

These integrals are set up starting with a conventional Feynman diagram. The $\tau = -\infty$ state is taken to be the plane wave describing the particle or antiparticle found at the beginning of the world line, in accordance with our earlier discussion. Similarly, the $\tau = +\infty$ state corresponds to the particle or antiparticle found at the end of the world line. The expressions for $d\sigma$ computed by substituting Eqs. (B20)–(B23) into Eq. (B19) will be found to be identical to the expressions obtained for the same processes by use of the conventional quantum theory of particles of definite mass.[8]

[1]V. A. Fock, Phys. Z. Sowjetunion **12,** 404 (1937).
[2]Yoichiro Nambu, Prog. Theor. Phys. (Kyoto) **5,** 82, (1950).
[3]Yasuhisa Katayama, Katuro Sawada, and Shuji Takagi, Prog. Theor. Phys. (Kyoto) **5,** 14 (1950).
[4]John R. Fanchi, Phys. Rev. D **20,** 3108 (1979).
[5]L. P. Horwitz and Y. Lavie, Tel Aviv University Preprint TAUP 833-80.
[6]Hiroshi Enatsu, Prog. Theor. Phys. (Kyoto) **30,** 236 (1963); H. Enatsu and S. Kawaguchi, Nuovo Cimento **27,** 458 (1975); H. Enatsu, A. Takenaka, and M. Okazaki, *ibid.* **43,** 575 (1978).
[7]Levere Hostler, J. Math, Phys. **21,** 2461 (1980).
[8]See, for example, James D. Bjorken and Sidney D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965), Chap. 14.
[9]In going over from the "covariant" form (2.5) to the "contravariant" form (2.16) of the 5-gradient operator, there is an implied metric tensor whose momentum space form

$$g^{AB} = \begin{pmatrix} -\frac{1}{4}k^2 g^{\mu\nu} & 0 \\ 0 & 1 \end{pmatrix}$$

can be the metric tensor of an ordinary Riemannian manifold. This observation could perhaps point the way to a 5-dimensional invariance group of our equations, but this has not as yet been pursued further.
[10]For an exposition of the Gupta–Bleuler method a convenient reference is J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley, Reading, MA, 1955).
[11]Equation (4.39) of Ref. 7. Note, however, the new normalization: the present iD_F is twice that of Ref. 7.
[12]This point was recognized earlier in Ref. 5, where an alternative treatment of the scattering problem will be found.

Analytical solutions of geometric optics with an approximation of diffraction

P. Hillion
Institut Henri Poincaré, 75231 Paris, France

S. Quinnez
Faculté des Sciences Mirande, 21000 Dijon, France

(Received 21 August 1979; accepted for publication 19 June 1980)

In the first part of this paper, starting with geometric optics equations, we give some simple formulas for the intensity of collimated and focused beams when the refractive index has the special form $n(r) = 1 + \epsilon\mu(r)$, where ϵ is a very small number. In the second part, we deal with diffraction as a perturbation of the refractive index and give corresponding formulas for essentially transversally bounded beams, i.e., beams with smooth boundary and most of the intensity near the propagation axis.

PACS numbers: 42.10.Dy, 42.10.Hc

I. GEOMETRIC OPTICS

A. Introduction

In this paper, starting with the geometric optics approximation of the scalar Helmholtz equation, we prove that for a light beam propagating in a medium with refractive index:

$$n(r) = 1 + \epsilon\mu(r), \tag{1}$$

where ϵ is a very small parameter and $\mu(r)$ a real function, the intenstiy $I(r)$ can be written in the form

$$I(r) = I_0\exp\{F(r)\}, \quad F(r) = \sum_{n=0}^{\infty}\epsilon^n F_n(r). \tag{2}$$

Here we assume the convergence of the series $\Sigma_{n=0}^{\infty}\epsilon^n F_n(r)$ (this difficult question will be discussed more in detail elsewhere) and we compute $F_1(r)$ for collimated and focused beams.

Then, in a second part, we show that diffraction can be accounted for as a perturbation of the refractive index using instead of $n(r)$

$$\hat{n}(r) = 1 + \epsilon\mu(r) + (1/2K_0^2)(1/\sqrt{I(r)})\Delta(\sqrt{I(r)}). \tag{1'}$$

In Eq. (1'), Δ is the Laplacian operator and K_0 the wave number. The results obtained in this paper have been applied to the propagation of a laser beam; in Ref. 1 we discuss thermal blooming and in Refs. 2 and 3 we examine the propagation in a random medium.

B. Geometric optics approximation

If we look for a solution of the scalar Helmholtz equation $(\Delta + K_0 n^2(r))\psi(r) = 0$ in the form $\psi(r) = u(r)e^{ik_0S(r)}$, where $u(r)$ and $S(r)$ are real functions, we obtain two equations:

$$\partial^j S(r)\partial_j S(r) = n^2(r) + (1/K_0^2 u(r))\partial^j\partial_j u(r), \tag{3}$$

$$2\partial^j S(r)\partial_j u(r) + u(r)\partial^j\partial_j S(r) = 0. \tag{4}$$

In Eqs. (3) and (4) ∂_j is the derivative $\partial/\partial X_j$, the index j takes the values 1,2, and 3 and the usual summation convention is used, and r denotes an arbitrary point in R^3. Geometric optics is obtained in the lim $K_0\to\infty$, so that Eq. (3) reduces to

$$\partial^j S(r)\partial_j S(r) = n^2(r). \tag{3'}$$

Remark: We are well aware that this derivation of geometric optics originally due to Sommerfeld and Runge[4] is open to many criticisms but that is irrelevant for our purpose; for a more rigorous derivation see Ref. 5.

When one substitutes the intensity $I(r) = n(r)u^2(r)$ of the light beam[5,6] into Eq. (2), it becomes

$$\partial^j S(r)\partial_j[I(r)/n(r)] + (I(r)/n(r))\partial^j\partial_j S(r) = 0. \tag{5}$$

As is well known[7] Eq. (3') can be solved using the method of characteristics. If we denote by s the arc length along the rays, we have

$$dX_j(s)/ds = (1/n(s))\partial_j S(s) \quad j = 1,2,3 \tag{6 a}$$

$$dS(s)/ds = n(s), \tag{6 b}$$

and

$$\frac{d}{ds}(\partial_j S(s)) = (1/2n(s))\partial_j n^2(s) \quad j = 1,2,3. \tag{6 c}$$

$X_j, j = 1,2,3$ denote the components of the vector $\mathbf{r}$ and to simplify we write $X_j(s)$, $S(s)$,$n(s)$ instead of $X_j(r(s))$,$S(r(s))$,$n(r(s))$. Then Eq. (5) takes the form of a conservation law

$$\partial^j\Big(I(s)(dX_j(s)/ds)\Big) = 0. \tag{7}$$

Since $d/ds = (dX_j(s)/ds)\partial^j$, Eq. (7) becomes

$$dI(s)/ds + I(s)\partial^j(dX_j(s)/ds) = 0,$$

and this leads to

$$I(s) = I(s_0)\exp\left\{-\int_{s_0}^{s}\partial^j\frac{dX_j(s)}{ds}\,d\sigma\right\}. \tag{8}$$

Substituting (1) into (6 b) gives

$$S(s) = S(s_0) + (s - s_0) + \epsilon\int_{s_0}^{s}\mu(\sigma)\,d\sigma. \tag{9}$$

The wave fronts are the surfaces $S(s) = $ const and the light rays are their orthogonal trajectories with tangent vectors

$$\partial_j S(s) = \partial_j s + \epsilon\left\{\partial_j s\,\mu(s) + \int_{s_0}^{s}\partial_j\mu(\sigma)\,d\sigma\right\}, \tag{10}$$

so that substituting (10) into (6a) and using (1), one has

$$\frac{dX_j(s)}{ds} = \left\{\partial_j s + \epsilon\int_{s_0}^{s}\partial_j\mu(\sigma)\,d\sigma\right\}(1 + \epsilon\mu(s))^{-1}. \tag{11}$$

Then Eq. (2) follows from (8) and (11) when $(1+\epsilon\mu(s))^{-1}$ is approximated by a power series expansion.

In the following two sections, we compute the first two terms $F_0(r), F_1(r)$ of Eq. (2) for a collimated beam and for a focused beam. The case of isotropic beams is given in Appendix B.

C. Collimated beam

Let us remark that in Eq. (11), $\partial_j s$ does not depend on ϵ, which implies that ∂_j is the derivation operator along the unperturbed ray; from now on, to avoid confusion, we denote this operator by $\bar{\partial}_j$, keeping ∂_j for the derivation operator along the perturbed ray. Then Eq. (11) becomes to first order

$$\frac{dX_j(s)}{ds} = \left(\bar{\partial}_j s + \epsilon \int_{s_0}^{s} \partial_j \mu(\sigma)\, d\sigma\right) + O(\epsilon^2). \qquad (11')$$

For $\epsilon = 0$, the unperturbed wave front is planar and we have $\bar{\partial}_j s = a_j$; $dX_j(s)/ds = a_j$, where a_j is a constant vector such that $\partial^j s \partial_j s = 1$ and $(dX^j(s)/ds)(dX_j(s)/ds) = 1$, which implies $a^j a_j = 1$. For $\epsilon \neq 0$ the latter relations become, to first order approximation,

$$\partial^j s \partial_j s = 1 + O(\epsilon^2) \qquad (12\text{ a})$$

$$(dX^j(s)/ds)(dX_j(s)/ds) = 1 + O(\epsilon^2). \qquad (12\text{ b})$$

This suggests taking $\bar{\partial}_j s = a_j(1+\epsilon b(s)) + O(\epsilon^2)$ and determining the function $b(s)$ by (12 b). Using the relations

$$a^j a_j = 1 \qquad (13\text{ a})$$

$$a^j \partial_j \mu(s) = \frac{dX^j(s)}{ds} \partial_j \mu(s) + O(\epsilon) = d\mu(s)/ds + O(\epsilon), \qquad (13\text{ b})$$

one obtains easily

$$b(s) = \mu(s) - \mu(s_0) + O(\epsilon),$$

which leads to

$$\frac{dX_j(s)}{ds} = a_j + \epsilon\left(a_j \mu(s_0) - a_j \mu(s) + \int_{s_0}^{s} \partial_j \mu(\sigma)\, d\sigma\right) + O(\epsilon^2) \qquad (14)$$

and after integration,

$$X_j(s) = X_j(s_0) + a_j(s - s_0) n(s_0) - \epsilon \times \left(a_j \int_{s_0}^{s} \mu(\sigma)\, d\sigma - \int_{s_0}^{s} d\tau \int_{s_0}^{\tau} \partial_j \mu(\sigma)\, d\sigma\right) + O(\epsilon^2), \quad j = 1,2,3. \qquad (15)$$

In Appendix A, we show that the derivation operator ∂_j is

$$\partial_j = \left\{ [1 - \epsilon\mu(s_0) + \epsilon\mu(s)]\delta_{jk} - \epsilon a_j \int_{s_0}^{s} \partial^k \mu(\sigma)\, d\sigma \right\} \times \frac{\partial}{\partial(a_k s)} + O(\epsilon^2), \qquad (16)$$

where δ_{jk} is the Kronecker symbol. Then it is easy to check (12.a) since one has $\partial_j s = a_j + O(\epsilon)$. Using (16) and this last relation, one obtains easily

$$\partial^j \frac{dX_j(s)}{ds} = \epsilon\left[- a^j \partial_j \mu(s) + \partial^j s \partial_j \mu(s) + \int_{s_0}^{s} \partial^j \partial_j \mu(\sigma)\, d\sigma \right] + O(\epsilon^2)$$
$$= \epsilon \int_{s_0}^{s} \partial^j \partial_j \mu(\sigma)\, d\sigma + O(\epsilon^2).$$

Then Eq. (8) becomes

$$I(s) = I(s_0) \exp\left[-\epsilon \int_{s_0}^{s} d\tau \int_{s_0}^{\tau} \partial^j \partial_j \mu(\sigma)\, d\sigma \right] + O(\epsilon^2). \qquad (17)$$

In Eqs. (14), (15), and (17), $\mu(s)$ is defined by the relation

$$\mu(s) = \mu\{X_j(s_0) + a_j(s - s_0)\} + O(\epsilon), \qquad (17')$$

which follows from Eq. (15) to zero order.

These results depend on the parameters a_j which are determined by the boundary conditions; then, considering a collimated beam as focusing to infinity in the Oz direction, we shall see that one has $a_\alpha = 0$, $\alpha = 1,2$ and $a_3 = 1$. So Eq. (15) gives for $j = 3$, $s = z + O(\epsilon^2)$ and for $\alpha = 1,2$,

$$X_\alpha(z) = X_\alpha(z_0) + \epsilon \int_{z_0}^{z} d\xi \int_{z_0}^{\xi} \partial_\alpha \mu(\zeta)\, d\zeta + O(\epsilon^2), \quad \alpha = 1,2, \qquad (18)$$

with, according to (17'),

$$\mu(s) = \mu(x,y,z) + O(\epsilon), \qquad (18')$$

while Eq. (17) becomes

$$I(x,y,z) = I(x_0,y_0,z_0) \exp \times \left[-\epsilon \int_{z_0}^{z} d\xi \int_{z_0}^{\xi} \partial^j \partial_j \mu(x,y,\zeta)\, d\zeta + O(\epsilon^2) \right]. \qquad (19)$$

Now, provided that $I(x,y,z_0)$ has first order continuous partial derivatives, one deduces from (18):

$$I(x_0,y_0,z_0) = I(x,y,z_0) - \epsilon \partial^\alpha I(x,y,z_0) \times \int_{z_0}^{z} d\xi \int_{z_0}^{\xi} \partial_\alpha \mu(x,y,\zeta)\, d\zeta + O(\epsilon^2)$$

that we write since $I(x,y,z_0)$ and $I(x_0,y_0,z_0)$ must be some positive quantities:

$$I(x_0,y_0,z_0) = I(x,y,z_0) \exp\left[-\epsilon \int_{z_0}^{z} d\xi \int_{z_0}^{\xi} \times \frac{\partial_\alpha I(x,y,z_0)}{I(x,y,z_0)} \partial^\alpha \mu(x,y,\zeta)\, d\zeta + O(\epsilon^2) \right].$$

Substituting this last relation into (19) gives

$$I_p(x,y,z) = I(x,y,z_0) \exp\left[-\epsilon \int_{z_0}^{z} d\xi \int_{z_0}^{\xi} \left[\frac{\partial^\alpha I(x,y,z_0)}{I(x,y,z_0)} \times \partial_\alpha \mu(x,y,\zeta) + \partial^j \partial_j \mu(x,y,\zeta) \right] d\zeta + O(\epsilon^2) \right]. \qquad (20)$$

In these relations the index α takes the values 1,2 and the summation convention is used; the index p of I_p means that the unperturbed wave front is planar. Equation (20) is used in Refs. 1 and 3 to discuss thermal blooming and random propagation.

D. Focused beam

1. Using s as parameter

We proceed as for a collimated beam but the derivation is a bit more combersome. For $\epsilon = 0$, the unperturbed light rays are straight lines converging to the focal point (of parameter s_f); then one may introduce a unit vector

$$\frac{\tau_j(s)}{1-s/s_f}, \tau^j(s)\tau_j(s) = \left(1-\frac{s}{s_f}\right)^2$$

such that

$$\bar{\partial}_j s = \frac{\tau_j(s)}{1-s/s_f} = -s_f\frac{d\tau_j(s)}{ds} = \text{const} = a_j \quad j=1,2,3. \tag{21}$$

For $\epsilon \neq 0$, we take in Eq. (11′)

$$\bar{\partial}_j s = \frac{\tau_j(s)}{1-s/s_f}(1+\epsilon d(s)) + O(\epsilon^2),$$

where the function $d(s)$ is determined so that Eq. (12 b) is fulfilled, which gives

$$\frac{dX_j(s)}{ds} = \frac{\tau_j(s)}{1-s/s_f} + \epsilon\left[(\mu(s_0)-\mu(s))\frac{\tau_j(s)}{1-s/s_f} + \int_{s_0}^{s}\partial_j\mu(\sigma)\,d\sigma\right] + O(\epsilon^2) \quad j=1,2,3. \tag{22}$$

Using (21) a straightforward integration of (22) results in

$$X_j(s) = X_j(s_0) - s_f n(s_0)(\tau_j(s)-\tau_j(s_0)) + \epsilon(\tau_j(s)\nu_0(s)+\nu_j(s)) + O(\epsilon^2), \quad j=1,2,3, \tag{23}$$

which is nothing but Eq. (15) written in a different form. In Eq. (23), one has

$$\nu_0(s) = \frac{-1}{1-s/s_f}\int_{s_0}^{s}\mu(\sigma)\,d\sigma$$
$$\nu_j(s) = \int_{s_0}^{s}d\xi\int_{s_0}^{\xi}\partial_j\mu(\sigma)\,d\sigma, \quad j=1,2,3. \tag{24}$$

The function $\nu_0(s)$ satisfies the following relation:

$$\frac{\nu_0(s)}{s_f} - \left(1-\frac{s}{s_f}\right)\frac{d\nu_0(s)}{ds} = \mu(s). \tag{24′}$$

For a beam converging to the focus on the Oz axis, one has $X_\alpha(s_f)=0, \alpha=1,2, X_3(s_f)=f$. Then the parameters a_j and s_f are given by the relations which follow from (23) with $\epsilon=0$:

$$X_\alpha(s_0) + a_\alpha(s_f-s_0) = 0, \quad \alpha=1,2;$$
$$a^j a_j = 1; \quad f = X_3(s_0) + a_3(s_f-s_0). \tag{23′}$$

In Appendix A, we show that the derivation operator in terms of $\partial/\partial\tau_j(s)$ is

$$\partial_j = -\frac{1}{s_f}\left\{\left[1-\epsilon\mu(s_0)+\frac{\epsilon\nu_0(s)}{s_f}\right]\delta_{jk} - \frac{\epsilon\tau_j(s)}{1-s/s_f}\times\left[\tau_k(s)\frac{d\nu_0(s)}{ds}+\frac{d\nu_k(s)}{ds}\right]\right\}\frac{\partial}{\partial\tau_k(s)} + O(\epsilon^2)$$
$$j=1,2,3. \tag{25}$$

Then using the following straightforward relations:

$$\frac{d}{ds} = \frac{dX_j(s)}{ds}\partial_j = \frac{\tau^j(s)}{1-s/s_f}\partial_j + O(\epsilon), \tag{26 a}$$

$$\frac{\tau^j(s)}{1-s/s_f}\frac{d\nu_j(s)}{ds} = \mu(s) - \mu(s_0) + O(\epsilon), \tag{26 b}$$

it is also proved in Appendix A, that the ∂_j operator (25) applied to Eq. (22) gives

$$\partial^j\frac{dX_j(s)}{ds} = -\frac{2}{s_f}\left[\frac{1}{1-s/s_f}+\epsilon\frac{d\nu_0(s)}{ds}\right] + \epsilon\int_{s_0}^{s}\partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon^2). \tag{27}$$

Using definition (24) of $\nu_0(s)$ and substituting (27) into (8) we get

$$I_f(s) = \frac{I_f(s_0)(1-s_0/s_f)^2}{(1-s/s_f)^2}\exp\left[-\frac{2\epsilon}{s_f-s}\times\int_{s_0}^{s}\mu(\sigma)\,d\sigma - \epsilon\int_{s_0}^{s}d\xi\int_{s_0}^{\xi}\partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon^2)\right], \tag{28}$$

with

$$\mu(s) = \mu\{X_j(s_0) - s_f\tau_j(s) + s_f\tau_j(s_0)\} + O(\epsilon). \tag{28′}$$

The index f in I_f means that one considers a focused beam. As expected, one has $\lim_{s_f\to\infty} I_f(s) = I_p(s)$.

2. Using z as parameter

For a beam propagating along the Oz axis, it is easier to work with z as parameter. Now, as noticed in the previous section, Eqs. (15) and (23) are equivalent, so we can use Eq. (15) which gives for $j=3$ to zero order:

$$z - z_0 = a_3(s-s_0) + O(\epsilon), \tag{29 a}$$

and using this last result to first order

$$z - z_0 = a_3(s-s_0) + \epsilon\left[(z-z_0)\mu(s_0) - \int_{s_0}^{s}\mu(\xi)\,d\xi + \frac{1}{a_3^2}\times\int_{z_0}^{z}d\xi\int_{z_0}^{\xi}\frac{\partial\mu(\zeta)}{\partial\zeta}d\zeta\right] + O(\epsilon^2). \tag{29 b}$$

From now on, we take for simplicity $z_0 = s_0 = 0$ and we set

$$\gamma(z) = \frac{1}{z}\int_0^z\mu(\xi)\,d\xi - \frac{1}{a_3^2 z}\int_0^z d\xi\int_0^\xi\frac{\partial}{\partial\zeta}\mu(\zeta)\,d\zeta, \tag{30}$$

so that Eqs. (29 b) and (30) lead to

$$s = (z/a_3)[1-\epsilon\mu(0)+\epsilon\gamma(z)] + O(\epsilon^2). \tag{31}$$

Using (31), the two other components, $\alpha=1,2$, of Eq. (15), become

$$X_\alpha(z) = X_\alpha(0) + \frac{a_\alpha z}{a_3}\left[1 - \frac{\epsilon}{a_3^2 z}\int_0^z d\xi\int_0^\xi\frac{\partial}{\partial\zeta}\mu(\zeta)d\zeta + \frac{\epsilon}{a_3^2}\int_0^z d\xi\int_0^\xi\partial_\alpha\mu(\zeta)\,d\zeta\right] + O(\epsilon^2),$$

while the boundary conditions are

$$\frac{a_\alpha}{a_3} = -\frac{X_\alpha(0)}{f}; \quad a_3^{-2} = 1 + \frac{X^\alpha(0)X_\alpha(0)}{f^2} \tag{32}$$

which leads to

$$X_\alpha(z) = X_\alpha(0) - X_\alpha(0)\frac{z}{f}\left(1-\frac{\epsilon}{z}\nu_3(z)\right) + \epsilon\nu_\alpha(z) + O(\epsilon^2), \quad \alpha=1,2, \tag{33}$$

with according to (24) and (29 a):

$$v_j(z) = a_3^{-2}\int_0^z d\xi \int_0^\xi \partial_j\mu(\zeta)\, d\zeta \quad j = 1,2,3. \tag{33'}$$

Let us note that one has

$$\frac{\partial}{\partial z} X_\alpha(z) = -\frac{X_\alpha(0)}{f} + O(\epsilon), \quad \alpha = 1,2. \tag{33''}$$

Let us then substitute these results into (28); according to (31) and to $s_f = f/a_3$, one has

$$(1 - s/s_f)^2 = \left[1 - \frac{z}{f} + \epsilon\frac{z}{f}(\mu(0) - \gamma(z))\right]^2,$$

$$\exp\left\{-\frac{2\epsilon}{s_f - s}\int_0^s \mu(\sigma)\, d\sigma\right\}$$

$$= \exp\left\{-\frac{2\epsilon}{f}\frac{1}{1 - z/f}\int_0^z \mu(\zeta)\, d\zeta + O(\epsilon^2)\right\}.$$

Using these two relations and Eq. (29 a), Eq. (28) becomes

$$I(x(z),y(z),z)$$

$$= I(x(0),y(0),0)F(z/f)\exp\left[-\frac{\epsilon}{a_3^2}\int_0^z d\xi \times \int_0^\xi \partial^j\partial_j\mu(x(\zeta),y(\zeta),\zeta)\, d\zeta + O(\epsilon^2)\right], \tag{34}$$

with

$$F(z/f) = \exp\left[-(2\epsilon/f)(1/1 - z/f)\int_0^z \mu(\zeta)\, d\zeta + O(\epsilon^2)\right] \times \left[1 - \frac{z}{f} + \epsilon\frac{z}{f}(\mu(0) - \gamma(z))\right]^2,$$

where we write in (34'), $\mu(\zeta)$, $\mu(0)$, $\gamma(z)$ for $\mu(x(\zeta),y(\zeta),\zeta)$, $\mu(x(0),y(0),0)$, $\gamma(x(z),y(z),z)$.

Now, from (33), it follows that one has

$$X_\alpha(0) = \frac{X_\alpha(z) - \epsilon v_\alpha(z)}{1 - z/f + (\epsilon/f)v_3(z)} = \frac{X_\alpha(z) - \epsilon v_\alpha(z)}{D_\epsilon(z)},$$

$$\text{where } D_\epsilon(z) = 1 - \frac{z}{f} + \frac{\epsilon}{f}v_3(z). \tag{35}$$

Then $I(x(0),y(0),0)$ becomes

$$I(x(0),y(0),0)$$

$$= I(x(z)/D_\epsilon(z), y(z)/D_\epsilon(z),0) - \epsilon v^\alpha(z)\partial_\alpha \times I(x(z)/D_\epsilon(z), y(z)/D_\epsilon(z),0) + O(\epsilon^2),$$

that we write as for a collimated beam

$$I(x(0),y(0),0) = I(z)\exp\left\{-\epsilon\frac{v^\alpha(z)\partial_\alpha I(z)}{I(z)} + O(\epsilon^2)\right\},$$

with $I(z) = I(x(z)/D_\epsilon(z), y(z)/D_\epsilon(z),0)$. Substituting this last result into (34) and taking (33') into account, we get

$$I_f(x(z),y(z),z) = I(z)F(z/f)\exp\left\{-\frac{\epsilon}{a_3^2}\int_0^z d\xi \times \int_0^\xi\left[\frac{\partial^\alpha I(z)}{I(z)}\partial_\alpha\mu(\zeta) + \partial^j\partial_j\mu(\zeta)\, d\zeta\right] + O(\epsilon^2)\right\}, \tag{36}$$

where the index f in I_f means a focused beam. In Eq. (36) the term $\mu(0)$ in the denominator of $F(z/f)$ becomes:

$$\mu(x(0),y(0),0)$$

$$= \mu(x(z)/D_\epsilon(z), y(z)/D_\epsilon(z),0) + O(\epsilon). \tag{36'}$$

Equations (34) and (36) can be compared respectively with (19) and (20) for a collimated beam; except for $D_\epsilon(z)$ and $F(z/f)$ the results are similar.

Remark 1: Let us prove that for $f \to \infty$, the expressions (31) and (33) for $s(z)$ and $X_\alpha(z)$ become those of the collimated case. Indeed, according to (32), one has

$$\lim_{f\to\infty} a_3 = 1, \quad \lim_{f\to\infty} a_\alpha = 0, \quad \alpha = 1,2$$

and since one has

$$\lim_{f\to\infty}\int_0^z [\partial\mu(\zeta)/\partial\zeta]\, d\zeta = \mu(z) - \mu(0),$$

it follows from (30) that $\lim_{f\to\infty}\gamma(z) = \mu(0)$, which leads to $s = z + O(\epsilon^2)$ while from (33) and (33'), one has

$$\lim_{f\to\infty} X_\alpha(z) = X_\alpha(0) + \epsilon\int_0^z d\xi \int_0^\xi \partial_\alpha\mu(\zeta)\, d\zeta + O(\epsilon^2),$$

which is the result obtained for a collimated beam.

Remark 2: As mentioned at the end of Sec. B, the anaical solutions for isotropic beams are discussed in Appendix B.

II. EXTENSION OF GEOMETRIC OPTICS

A. Generalities

1. Geometric optics with an approximation of diffraction

Let us now come back to Eqs. (3) and (4); we proved that for $K_0 \to \infty$, this system can be written as [see Eqs. (3') and (5)]:

$$\partial^j S(r)\partial_j S(r) = n^2(r),$$

$$\partial^j S(r)\partial_j(I_\infty(r)/n(r)) + (I_\infty(r)/n(r))\partial^j\partial_j S(r) = 0,$$

where in this part II, we note by $I_\infty(r)$ the intensity supplied by geometric optics. It is easy to see that Eqs. (3) and (4) can also be written for $K_0^2 \neq 0$:

$$\partial_j S(r)\partial^j S(r) = \hat{n}^2(r), \tag{37 a}$$

$$\partial^j S(r)\partial_j\left(\frac{I(r)}{\hat{n}(r)}\right) + \frac{I(r)}{\hat{n}(r)}\partial^j\partial_j S(r) = 0, \tag{37 b}$$

where one has

$$\hat{n}^2(r) = n^2(r) + \frac{1}{K_0^2 u(r)}\partial^j\partial_j u(r) = 0$$

$$I(r) = \hat{n}(r)u^2(r). \tag{38}$$

Of course $\hat{n}^2(r)$ is unknown but it is easy to find an approximation

$$\hat{n}_1^2(r) = n^2(r) + \frac{1}{K_0^2}\left(\frac{n(r)}{I_\infty(r)}\right)^{1/2}\partial^j\partial_j\left(\frac{I_\infty(r)}{n(r)}\right)^{1/2}, \tag{38'}$$

and the solution of Eqs. (37) and (38') by the method of characteristics used in part I will supply an approximation $I_1(r)$ of the diffracted intensity $I(r)$, exact to first order $O((K_0^2a^2)^{-1})$, where a is the e^{-1} width of the light beam. The process can be iterated: let $I_l(r)$ be the approximation obtained after l iterations, I_l is exact to the order $O((K_0^2a^2)^{-l})$ and $I_{l+1}(r)$ is obtained by solving (37) and (39) with

$$n_{l+1}^2(r) = n^2(r) + \frac{1}{K_0^2}\left(\frac{n_l(r)}{I_l(r)}\right)^{1/2} \partial^j\partial_j \left(\frac{I_l(r)}{n_l(r)}\right)^{1/2}. \tag{39}$$

Of course, one must have

$$\left| \frac{1}{K_0^2}\left(\frac{n_l(r)}{I_l(r)}\right)^{1/2} \partial^j\partial_j \left(\frac{I_l(r)}{n_l(r)}\right)^{1/2} \right| \ll 1, \tag{40}$$

with a similar condition for $I_\infty(r)$. [We shall see that it is difficult to fulfill (40) for a focused beam.] We assume here the convergence of this iterative process (This question will be discussed elsewhere) and we therefore write

$$\lim_{l\to\infty} I_l(r) = I(r).$$

An important point here is the existence of $\partial^j\partial_j(I_l(r)/n(r))^{1/2}$, which requires that $I_l(r)$, $I_\infty(r)$ and $n_l(r)$, $n(r)$ be sufficiently smooth functions. In the next section, we discuss a particular kind of light beams such that this existence is guaranteed.

2. *Essentially transversally bounded beams*

In this section, we assume $n(r) = 1$ and we only consider cylindrical beams. (For rectangular beam see Ref. 2).

Since $n(r) = 1$, one has for a collimated beam according to (19):

$$I_{p\infty}(x,y,z) = I(x,y,0), \tag{41}$$

while from (8′) one has

$$\epsilon_d\mu_d(r) = \frac{1}{2K_0^2}\frac{1}{[I_{p\infty}(r)]^{1/2}}\partial^\alpha\partial_\alpha[I_{p\infty}(r)]^{1/2}, \tag{41′}$$

where the index d means that the perturbation of the refractive index is due to diffraction and where α takes the values 1,2, since $I_{p\infty}(r)$ does not depend on z. Using (18) and (19), this gives

$$\begin{aligned} I_{p1}(x,y,z) = {}& I(x-\epsilon_d\nu_{1d}(z), y-\epsilon_d\nu_{2d}(z),0) \\ & \times \exp\{-\epsilon_d(z^2/2)\partial^\alpha\partial_\alpha\mu_d(r) + O(\epsilon_d^2)\}, \end{aligned} \tag{42}$$

with

$$\nu_{\alpha d} = \int_0^z d\xi \int_0^\xi \partial_\alpha\mu_d(\zeta)\, d\zeta = (z^2/2)\partial_\alpha\mu_d(z). \tag{42′}$$

An essentially transversally bounded beam[2] is defined by the following conditions: Let ρ,z be be the usual cylindrical coordinates and let $u=\rho/a$ be a dimensionless parameter where a is the e^{-1} width of the light beam. We assume that $I(\rho,0)$ has the form

$$I(\rho,0) = I_0\exp\{2G(u)\} = I_0\exp\left\{-2\int g(u)\,du\right\}, \tag{41″}$$

such that $G(u)$ and $g(u)$ have the following properties:

(i) $g(u)$ is a C^4 function;

(ii) $G(u) < 0$ for $u\to\infty$ and there exist u_0 such that most of the intensity is found inside a cylinder with radius u_0 (this condition need not be made more precise). Then a simple computation gives

$$\epsilon_d\mu_d(r) = (1/2K_0^2a^2)[g^2(u) + g'(u) + (1/u)g(u)]. \tag{43}$$

Because of the properties $g(u)$, $\mu_d(u)$ is a bounded function for $u<u_0$ so that the condition (40) is fulfilled, since $(2K_0^2a^2)^{-1}$ is a very small number for a light beam.

Moreover we assume that the following conditions are fulfilled:

$$\epsilon_d > 0,\quad \partial^\alpha\partial_\alpha\mu_d(r) > 0,\quad \frac{\nu_{\alpha d}}{X_\alpha} > 0,\quad \alpha = 1,2$$

and these qualities remain bounded for $x_\alpha \to 0$. (44)

The conditions (i), (ii) and Eq. (44) define an essentially transversally bounded beam and we proved in Ref. 2 that the Kogelnik–Li solutions[8] for a laser beam have this property.

According to (44), one has

$$\begin{aligned} X_\alpha(0) &= X_\alpha - \epsilon_d\nu_{\alpha d}(z) + O(\epsilon_d^2) \\ &= X_\alpha\Big/\left[1 + \frac{2\epsilon_d}{X_\alpha}\nu_{\alpha d}(z)\right]^{1/2} + O(\epsilon_d^2) \end{aligned} \tag{42″}$$

and

$$\begin{aligned} &\exp\{-\epsilon_d(z^2/2)\partial^\alpha\partial_\alpha\mu_d(r) + O(\epsilon_d^2)\} \\ &\quad = 1\Big/\left[1 + \epsilon_d\frac{z^2}{2}\partial^\alpha\partial_\alpha\mu_d(r)\right] + O(\epsilon_d^2). \end{aligned} \tag{42‴}$$

Substituting (42″) and (42‴) into Eq. (42), one sees that $I_{p_1}(x,y,z)$ is a bounded positive quantity

$$\begin{aligned} I_{p_1}(x,y,z) = {}& \frac{1}{1+\epsilon_d(z^2/2)\partial^\alpha\partial_\alpha\mu_d(r)} \\ & \times I\left(\frac{x}{[1+(2/x)\epsilon_d\nu_{1d}(z)]^{1/2}}, \frac{y}{[1+(2/y)\epsilon_d\nu_{2d}(z)]^{1/2}}, 0\right) \\ & + O(\epsilon_d^2) \end{aligned} \tag{45}$$

but, according to (43), one has

$$\epsilon_d = 1/2K_0^2a^2,\quad \mu_d(u) = g^2(u) + g'(u) + (1/u)g(u), \tag{43′}$$

while Eq. (42′) gives

$$\nu_{\rho d}(z) = \frac{z^2}{2a}\mu_d'(u)$$

and

$$\partial^\alpha\partial_\alpha\mu_d(\rho,z) = \frac{1}{a^2}\left(\mu_d''(u) + \frac{1}{u}\mu_d'(u)\right). \tag{46}$$

In these relations, the derivative is taken with respect to u. Using (46), Eq. (45) becomes

$$\begin{aligned} I_{p_1}(\rho,z) = {}& \frac{I_0}{1+(z^2/4K_0^2a^4)(\mu_d''(u) + (1/u)\mu_d'(u))} \\ & \times \exp\left\{\left[2G\left(\frac{\mu}{\{[1+(z^2/2K_0^2a^4)][\mu_d'(u)/u]\}^{1/2}}\right)\right] \right. \\ & \left. + O\left(\frac{1}{K_0^4a^4}\right)\right\}, \end{aligned} \tag{45′}$$

while the conditions (44) are

$$u^{-1}\mu_d'(u) > 0 \quad \mu_d'' + u^{-1}\mu_d'(u) > 0. \tag{44′}$$

We shall now prove that for a Gaussian beam Eq. (45′) gives the same result as wave optics. Indeed for a Gaussian beam, one has

$$I(\rho,0) = I_0\exp(-u^2),\quad u = \rho/a; \tag{47}$$

that is with previous notations

$$\begin{aligned} & G(u) = -u^2/2,\quad \mu_d(u) = u^2 - 2 \\ & \mu_d''(u) = (1/u)\mu_d'(u) = 2. \end{aligned} \tag{47′}$$

One can easily see that conditions (44) are fufilled. Then, substituting (47′) into (45′) gives:

$$I_{p_1}(\rho,z) = \frac{I_0}{1+z^2/K_0^2a^4}\exp\left\{-\frac{\rho^2}{a^2(1+z^2/K_0^2a^4)}\right\}, \tag{48}$$

which is the exact result supplied by wave optics.

As previously stated, the problem is not so easy for a focused beam since the condition (40) cannot be fulfilled for any z. Let us for instance consider a Gaussian beam. According to (36), one has instead of (41)

$$I_{f\infty}(x(z),y(z),z) = \frac{I_0}{(1-z/f)^2}\exp\left(-\frac{X^\alpha(z)X_\alpha(z)}{a^2(1-z/f)^2}\right) \tag{49}$$

and, after some computations, one obtains

$$\epsilon_d\mu_d(x(z),y(z),z) = \frac{1}{2K_0^2a^2(1-z/f)^2} \times\left\{\frac{X^\alpha(z)X_\alpha(z)}{a^2(1-z/f)^2}-2\right\}. \tag{50}$$

It is clear that the factor $(1-z/f)^2$ makes it impossible to satisfy (40) except for $z/f \ll 1$. Then, we shall deduce the diffracted intensity $I_f(x(z),y(z),z)$ for a focused beam from the corresponding quantity for a collimated beam by the following *ansätz*:

$$I_0 \to \frac{I_0}{(1-z/f)^2},\quad K_0^2 \to K_0^2(1-z/f)^2,$$
$$X_\alpha \to \frac{X_\alpha(z)}{1-z/f},\quad \alpha = 1,2, \tag{51}$$

which can be justified by the following arguments:

(i) For z/f small enough so that Eq. (40) is fulfilled, one obtains the same result starting directly from (49), (50) or using (45), (51). We shall not prove this fact here to avoid uninteresting computations.

(ii) Using (51), $I_f(x(z),y(z),z)$ is bounded and positive for any z, moreover for a Gaussian beam, one also obtains the exact result of wave optics. Indeed, applying (51) to (48) gives

$$I_{f_1}(\rho,z) = \frac{I_0}{(1-z/f)^2+z^2/K_0^2a^4} \times\exp\left\{-\frac{\rho^2(z)}{a^2[(1-z/f)^2+z^2/K_0^2a^4]}\right\}. \tag{52}$$

The idea to consider diffraction as a perturbation of the refractive index was also mentioned in Ref. 9.

B. Approximation of diffracted intensity for a Gaussian beam

1. Collimated beam

In the previous section, ϵ was zero while we now assume $\epsilon \neq 0$. Then, according to (38′), (41′), and (47′), the perturbed refractive index becomes

$$n_1^2(r) = n^2(r) + (1/K_0^2a^2)(\rho^2/a^2-2),\quad \rho^2 = X^\alpha X_\alpha,$$

that is,

$$n_1(r) = 1 + \tilde\epsilon\tilde\mu(r) + O(\tilde\epsilon^2)$$
$$= 1 + \epsilon\mu(r) + (1/2K_0^2a^2)(\rho^2/a^2-2) + O(\tilde\epsilon^2), \tag{53}$$

while Eqs. (18) and (19) becomes

$$X_\alpha(z) = X_\alpha(z_0) + \tilde\epsilon\int_{z_0}^{z} d\xi\int_{z_0}^{\xi}\partial_\alpha\mu(\zeta)\,d\zeta + O(\tilde\epsilon^2), \tag{54}$$

$$\hat I_p(x,y,z) = I(z_0,y_0,z_0) \times\exp\left\{-\tilde\epsilon\int_{z_0}^{z}d\xi\int_{z_0}^{\xi}\partial^j\partial_j\tilde\mu(x,y,\zeta)\,d\zeta + O(\tilde\epsilon^2)\right\}, \tag{55}$$

where $\hat I$ means that diffraction has been taken into account.To simplify we now write $d^2(z) = 1 + z^2/K_0^2a^4$ and since a Gaussian beam is essentially transversally bounded, we may write similary to (42″) and (42‴)

$$\tilde X_\alpha(z) = \frac{1}{d(z)}\{X_\alpha(z) - \epsilon v_\alpha(z)\} + O(\epsilon^2),$$
$$v_\alpha(z) = \int_0^z d\xi\int_0^\xi \partial_\alpha\mu(\zeta)\,d\zeta, \tag{56}$$

$$\exp\left\{-\tilde\epsilon\int_0^z d\xi\int_0^\xi\partial^j\partial_j\tilde\mu(x,y,\zeta)\,d\zeta + O(\tilde\epsilon^2)\right\}$$
$$= \frac{1}{d^2(z)}\exp\left\{-\epsilon\int_0^z d\xi\int_0^\xi\partial^j\partial_j\mu(x,y,\zeta)\,d\zeta + O(\epsilon^2)\right\}. \tag{56′}$$

Substituting (56) and (56′) into (55) gives

$$\hat I_p(x,y,z) = \frac{I}{d^2(z)}I\left(\frac{X_\alpha(z)-\epsilon v_\alpha(z)}{d(z)},0\right) \times\exp\left\{-\epsilon\int_0^z d\xi\int_0^\xi\partial^j\partial_j\mu(x,y,\zeta)\,d\zeta + O(\epsilon^2)\right\}, \tag{57}$$

which we also may write in the same form as Eq. (20):

$$\hat I_p(x,y,z) = \frac{1}{d^2(z)}I\left(\frac{x}{d(z)},\frac{y}{d(z)},0\right) \times\exp\left\{-\epsilon\int_0^z d\xi\int_0^\xi\left[\frac{\partial^\alpha I(x/d(z),y/d(z),0)}{I(x/d(z),y/d(z),0)}\partial_\alpha\mu(x,y,\zeta) + \partial^j\partial_j\mu(x,y,\zeta)\right]d\zeta + O(\epsilon^2)\right\}. \tag{57′}$$

For a Gaussian beam, this last expression becomes

$$\hat I_p(x,y,z) = \frac{1}{d^2(z)}\exp\left(-\frac{X^\alpha X_\alpha}{a^2d^2(z)}\right) \times\exp\left\{\epsilon\int_0^z d\xi\int_0^\xi\left[\frac{2X^\alpha}{a^2d^2(z)}\partial_\alpha\mu(x,y,\zeta) - \partial^j\partial_j\mu(x,y,\zeta)\right]d\zeta + O(\epsilon^2)\right\}. \tag{58}$$

This formula can be used, for instance, to discuss thermal blooming in the propagation of a Gaussian laser beam.

2. Focused beam

As for $\epsilon = 0$ and for the same reason, we shall use some *ansätz* to obtain the diffracted intensity of a focused beam directly from (57). Using (34) and (35) one has instead of (51)

$$I(x_0,y_0,z_0) \to I(x_0,y_0,z_0) F\left(\frac{z}{f}\right); \quad K_0^2 \to K_0^2 D_\epsilon^2(z);$$

$$X_\alpha(z) \to \frac{X_\alpha(z)}{D_\epsilon(z)}, \quad \nu_\alpha(z) \to \frac{\nu_\alpha(z)}{D_\epsilon(z)}, \quad \alpha = 1,2, \tag{59}$$

with $F(z/f)$ and $D_\epsilon(z)$ respectively given by (34′) and (35). Of course, for $\epsilon \to 0$, Eqs. (59) reduce to (41). Applying (59) to (56) gives

$$X_\alpha(z_0) = \frac{1}{\hat{D}_\epsilon(z)} (X_\alpha(z) - \nu_\alpha(z)) + O(\epsilon^2), \tag{60}$$

with according to (35),

$$\hat{D}_\epsilon^2 = D_\epsilon^2(z) + \frac{z^2}{K_0^2 a^4}. \tag{60′}$$

Now let us remark that using (30) and (30′), the expression (34′) of $F(z/f)$ can be simplified if the exponential term is expanded in a power series and one obtains

$$F\left(\frac{z}{f}\right) = \Delta_\epsilon^{-2}(z) = \left[1 - \frac{z}{f} + \frac{\epsilon}{f}(z\mu(0) + \nu_3(z)) + O(\epsilon^2)\right]^{-2}. \tag{61}$$

So according to (59) and (61), the term $I(x_0,y_0,z_0)d^{-2}(z)$ of (57) becomes

$$\frac{I(x_0,y_0,z_0)}{1 + z^2/K_0^2 a^4 D_\epsilon^2(z)} \frac{1}{\Delta_\epsilon^2(z)} = \frac{I(x_0,y_0,z_0)}{\hat{\Delta}_\epsilon^2(z)}, \tag{62}$$

with

$$\hat{\Delta}_\epsilon^2(z) = \Delta_\epsilon^2(z) + \frac{z^2}{K_0^2 a^4} \frac{\Delta_\epsilon^2(z)}{D_\epsilon^2(z)}. \tag{62′}$$

Then, according to (57), (60), and (62), the diffracted intensity of a focused beam is

$$\hat{I}_f(x(z),y(z),z) = \frac{1}{\hat{\Delta}_\epsilon^2(z)} I\left(\frac{x_\alpha(z) - \epsilon\nu_\alpha(z)}{\hat{D}_\epsilon(z)}, 0\right) \exp\left\{-\frac{\epsilon}{a_3^2}\int_0^z d\xi \times \int_0^\xi \partial^j\partial_j\mu(x(\zeta),y(\zeta),\zeta)\, d\zeta + O(\epsilon^2)\right\}, \tag{63}$$

with according to (33) $a_3^{-2} = 1 + X^\alpha(0)X_\alpha(0)/f^2$.

For a Gaussian beam, Eq. (63) becomes

$$\hat{I}_f x(z),y(z),z) = \frac{1}{\hat{\Delta}_\epsilon^2(z)} \exp\left(-\frac{X^\alpha(z)X_\alpha(z)}{\hat{D}_\epsilon^2(z)}\right) \times \exp\left\{\epsilon\int_0^z d\xi \int_0^\xi \left[\frac{2X^\alpha(\zeta)}{a^2\hat{D}_\epsilon^2(z)} \partial_\alpha\mu(x(\zeta)y(\zeta),\zeta) - \frac{1}{a_3^2}\partial^j\partial_j\mu(x(\zeta),y(\zeta),\zeta)\right] d\zeta + O(\epsilon^2)\right\}. \tag{64}$$

Then we obtain a slightly more intricate formula than for a collimated beam.

III. CONCLUSION

As proved by Kline[10] the general solution of Eqs. (3′) and (4) requires the principal radii of curvature along the light rays but the previous sections show that this solution becomes rather simple to first order when the refractive index has the form (1) and that it leads to analytical expressions which are tractable in those cases we considered. Even when diffraction is approximated as a perturbation of the refractive index, these expressions remain manageable provided that one considers essentially bounded beams, a class not as restricted as might be thought (see Ref. 2). These analytical expressions are used in Refs. 1 and 3 to discuss thermal blooming of laser beams and the propagation of light in turbulent medium. As previously said we intend to discuss elsewhere the convergence of iteration process used here.

APPENDIX A

In this Appendix, we compute the operator ∂_j along the perturbed rays. According to Eq. (15), one has for a collimated beam

$$dX_j(s) = \left\{(1 + \epsilon\mu(s_0) - \epsilon\mu(s))\delta_{jk} + \epsilon a^k \int_{s_0}^s \partial_j\mu(\sigma)\, d\sigma\right\} \times d(a_k s) + O(\epsilon^2), \quad j = 1,2,3,$$

which gives at once

$$\partial_j = \left\{(1 - \epsilon\mu(s_0) + \epsilon\mu(s))\delta_{jk} - \epsilon a_j \int_{s_0}^s \partial^k\mu(\sigma)\, d\sigma\right\} \times \frac{\partial}{\partial(a_k s)} + O(\epsilon^2), \quad j = 1,2,3,$$

where δ_{jk} is the Kronecker symbol. Now using the relation $a^j a_j = 1$, one may write

$$\begin{aligned}\partial_j s &= \partial_j(a^l a_l s) = a^l \partial_j(a_l s), \\ &= a^l \delta_{kl}\left\{(1 - \epsilon\mu(s_0) + \epsilon\mu(s))\delta_{jk} - \epsilon a_j \times \int_{s_0}^s \partial^k\mu(\sigma)\, d\sigma\right\} + O(\epsilon^2) = a_j + O(\epsilon^2),\end{aligned}$$

where we used the relation

$$\partial^j\partial_j\mu(s) = d\mu(s)/ds + O(\epsilon), \quad j = 1,2,3.$$

For a focused beam, the computations are similar. According to (23), one has

$$dX_j(s) = \left\{\left(1 + \epsilon\mu(s_0) - \frac{\epsilon}{s_f}\nu_0(s)\right)\delta_{jk} + \frac{\epsilon\tau_k(s)}{1 - s/s_f} \times \left[\tau_j(s)\frac{d\nu_0(s)}{ds} + \frac{d\nu_j(s)}{ds}\right]\right\}(-s_f d\tau^k(s)) + O(\epsilon^2), \quad j = 1,2,3,$$

which leads to

$$d\tau_k(s) = -\frac{1}{s_f}\left\{\left[1 - \epsilon\mu(s_0) + \frac{\epsilon\nu_0(s)}{s_f}\right]\delta_{jk} - \frac{\epsilon\tau_j(s)}{1 - s/s_f} \times \left[\tau_k(s)\frac{d\nu_0(s)}{ds} + \frac{d\nu_k(s)}{ds}\right]\right\} dX^j(s) + O(\epsilon^2), \quad k = 1,2,3,$$

and this results in

$$\partial_j = -\frac{1}{s_f}\left\{\left[1 - \epsilon\mu(s_0) + \frac{\epsilon\nu_0(s)}{s_f}\right]\delta_{jk} - \frac{\epsilon\tau_j(s)}{1 - s/s_f} \times \left[\tau_k(s)\frac{d\nu_0(s)}{ds} + \frac{d\nu_k(s)}{ds}\right]\right\}\frac{\partial}{\partial\tau_k(s)} + O(\epsilon^2), \quad j = 1,2,3.$$

Then using (24') and (26b), it is easy to prove that one has

$$\partial_j s = \frac{\tau_j(s)}{1 - s/s_f} + O(\epsilon^2). \tag{A1}$$

Let us then consider the quantity $\partial^j dX_j(s)/ds$. For the last term on the right-hand side of Eq. (22), one has

$$\begin{aligned}\partial^j\left(\frac{\tau_j(s)}{1 - s/s_f}\right) &= \frac{\partial^j \tau_j(s)}{1 - s/s_f} + \frac{\tau_j(s)}{s_f}\frac{\partial^j s}{(1 - s/s_f)^2}\\ &= \frac{\partial^j \tau_j(s)}{1 - s/s_f} + \frac{\tau_j^2}{s_f(1 - s/s_f)^3} + O(\epsilon^2).\end{aligned} \tag{A2}$$

Now, using (24') and (26b) one has

$$\begin{aligned}\partial^j \tau_j(s) &= -\frac{1}{s_f}\left\{3 - 3\mu(s_0) + \frac{3\nu_0(s)}{s_f} - \epsilon\right.\\ &\quad\left.\times\left[\left(1 - \frac{s}{s_f}\right)\frac{d\nu_0(s)}{ds} + \frac{\tau^k(s)}{1 - s/s_f}\frac{d\nu_k(s)}{ds}\right]\right\} + O(\epsilon^2)\\ &= -\frac{1}{s_f}\left[3 - 2\epsilon\mu(s_0) + \frac{3\nu_0(s)}{s_f}\right] + O(\epsilon^2).\end{aligned}$$

Substituting this last result into (A2) gives

$$\partial^j\left(\frac{\tau_j(s)}{1 - s/s_f}\right) = \frac{-2}{s_f - s}\left[1 - \epsilon\mu(s_0) + \frac{\epsilon\nu_0(s)}{s_f}\right] + O(\epsilon^2). \tag{A3}$$

For the second term, on the right-hand side of Eq. (22), one has

$$\begin{aligned}&\partial^j\left\{(\mu(s_0) - \mu(s))\frac{\tau_j(s)}{1 - s/s_f} + \int_{s_0}^s \partial_j\mu(\sigma)\,d\sigma\right\}\\ &= -\frac{1}{s_f}\frac{\partial}{\partial\tau_j(s)}\left[(\mu(s_0) - \mu(s))\frac{\tau_j(s)}{1 - s/s_f}\right]\\ &\quad + \partial^j s\partial_j\mu(s) + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon)\\ &= -\frac{2}{s_f - s}[\mu(s_0) - \mu(s)] - \frac{d\mu(s)}{ds} + \frac{d\mu(s)}{ds}\\ &\quad + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon)\\ &= -\frac{2}{s_f - s}[\mu(s_0) - \mu(s)] + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon),\end{aligned} \tag{A4}$$

where we used the relation

$$-\frac{1}{s_f}\frac{ds}{d\tau_j(s)} = \partial_j s = \frac{\tau_j(s)}{1 - s/s_f}.$$

Combining (A3) and (A4), one obtains for $\partial^j dX_j(s)/ds$

$$\begin{aligned}\partial^j\frac{dX_j(s)}{ds} &= -\frac{2}{s_f - s}\left[1 - \epsilon\mu(s) + \epsilon\frac{\nu_0(s)}{s_f}\right] + \epsilon\\ &\quad\times\int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon^2)\\ &= -\frac{2}{s_f}\left[\frac{1}{1 - s/s_f} + \epsilon\frac{d\nu_0(s)}{ds}\right] + \epsilon\\ &\quad\times\int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon^2).\end{aligned} \tag{A5}$$

APPENDIX B: ISOTROPIC BEAMS

We proceed as for collimated and focused beams. For $\epsilon = 0$ the unperturbed wave fronts are spherical, so one has $\bar{\partial}_j s = \rho_j(s)/s = a_j$ with $\rho_j(s)\rho^j(s) = (s^2)$. Then for $\epsilon \neq 0$, we take in Eq. (11') $\bar{\partial}_j s = (\rho_j(s)/s)(1 + \epsilon c(s) + O(\epsilon^2)$ and we determine the function $c(s)$ so that the condition (12b) is fulfilled which gives,

$$\begin{aligned}\frac{dX_j(s)}{ds} &= \frac{\rho_j(s)}{s} + \epsilon\left[(\mu(s_0) - \mu(s))\frac{\rho_j(s)}{s}\right.\\ &\quad\left.+ \int_{s_0}^s \partial_j\mu(\sigma)\,d\sigma\right] + O(\epsilon^2), \quad j = 1,2,3.\end{aligned} \tag{B1}$$

Since one has

$$\frac{\rho_j(s)}{s} = \frac{d\rho_j(s)}{ds} = \text{const} = a_j, \tag{B2}$$

the integration of Eq. (A6) gives

$$\begin{aligned}X_j(s) &= X_j(s_0) + \{\rho_j(s) - \rho_j(s_0)\}n(s_0)\\ &\quad - \epsilon\{\rho_j(s)\nu(s) - \nu_j(s)\} + O(\epsilon^2), \quad j = 1,2,3.\end{aligned} \tag{B3}$$

The functions $\nu_j(s)$ are defined as in (24) and one has

$$\nu(s) = \frac{1}{s}\int_{s_0}^s \mu(\sigma)\,d\sigma \tag{B4}$$

with both relations:

$$\nu(s) + s\frac{d\nu(s)}{ds} = \mu(s) \tag{B5a}$$

$$\rho^j(s)\frac{d\nu_j(s)}{ds} = s[\mu(s) - \mu(s_0)] + O(\epsilon). \tag{B5b}$$

We shall prove later that the ∂_j operator is

$$\begin{aligned}\partial_j &= \left\{(1 - \epsilon\mu(s_0) + \epsilon\nu(s))\delta_{jk} + \frac{\epsilon\rho_j(s)}{s}\right.\\ &\quad\left.\times\left(\rho_k(s)\frac{d\nu(s)}{ds} - \frac{d\nu_k(s)}{ds}\right)\right\}\frac{\partial}{\partial\rho_k(s)} + O(\epsilon^2)\end{aligned} \tag{B6}$$

and that one has

$$\begin{aligned}\partial^j\frac{dX_j(s)}{ds} &= \frac{2}{s}[1 + \epsilon(\nu(s) - \mu(s))]\\ &\quad + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon^2)\\ &= \frac{2}{s}\left[1 - \epsilon s\frac{d\nu(s)}{ds}\right] + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\,d\sigma + O(\epsilon^2).\end{aligned} \tag{B7}$$

Since one has $\nu(s_0) = 0$, Eq. (8) becomes for $s \geqslant s_0 \geqslant 0$,

$$\begin{aligned}I_{sp}(s) &= I(s_0)\frac{s_0^2}{2}\exp\left\{-2\epsilon\nu(s) - \epsilon\int_{s_0}^s d\xi\right.\\ &\quad\left.\times\int_{s_0}^\xi \partial^j\partial_j\mu(\sigma)\,d\sigma\right\} + O(\epsilon^2),\end{aligned} \tag{B8}$$

which can be compared with Eqs. (19) and (28). In (B8) one has

$$\mu(s) = \mu(X_j(s) + \rho_j(s) - \rho_j(s_0)) + O(\epsilon).$$

Let us now prove (B6) and (B7). It follows from (B3) that one has

$$dX_j(s) = \left\{ [1 + \epsilon\mu(s_0) - \epsilon\nu(s)]\delta_{jk} - \frac{\epsilon\rho_k(s)}{s} \times \left[\rho_j(s)\frac{d\nu(s)}{ds} - \frac{d\nu_j(s)}{ds} \right] \right\} d\rho^k(s) + O(\epsilon^2),$$

which leads to

$$d\rho_k(s) = \left\{ [1 - \epsilon\mu(s_0) + \epsilon\nu(s)]\delta_{jk} + \frac{\epsilon\rho_j(s)}{s} \times \left[\rho_k(s)\frac{d\nu(s)}{ds} - \frac{d\nu_k(s)}{ds} \right] \right\} dX^j(s) + O(\epsilon^2).$$

Eq. (B6) follows from the relation $\partial_j = [d\rho^k(s)/dX_j(s)]\, \partial/\partial\rho_k(s)$ and it is trivial to show that one has

$$\partial_j s = \frac{\rho_j(s)}{s} + O(\epsilon^2)$$

so that Eq. (12a) is fulfilled. Let us now prove (B7). For the first term on the right-hand side of (B1), we get

$$\partial^j\left(\frac{\rho_j(s)}{s}\right) = \frac{1}{s}\partial^j\rho_j(s) - \frac{1}{s^2}\rho^j(s)\partial_j(s) = \frac{1}{s}\partial^j\rho_j(s) - \frac{1}{s^3}\rho^j(s)\rho_j(s) + O(\epsilon^2)$$

but, according to (B5a) and (B5b) one has

$$\partial^j\rho_j(s) = 3 + \epsilon\left[3(\nu(s) - \mu(s_0)) + s\frac{d\nu(s)}{ds} - \frac{\rho_j(s)}{s}\frac{d\nu_j(s)}{ds}\right] + O(\epsilon^2) = 3 + 2\epsilon[\nu(s) - \mu(s_0)] + O(\epsilon^2)$$

so that $\partial^j(\rho_j(s)/s)$ becomes

$$\partial^j\left(\frac{\rho_j(s)}{s}\right) = \frac{2}{s}[1 + \epsilon\nu(s) - \epsilon\mu(s_0)] + O(\epsilon^2). \tag{B9}$$

The second term of Eq. (B1) gives

$$\begin{aligned}
\partial^j &\left\{(\mu(s_0) - \mu(s))\frac{\rho_j(s)}{s} + \int_{s_0}^s \partial_j\mu(\sigma)\, d\sigma\right\} \\
&= \frac{\partial}{\partial\rho_j(s)}\left\{(\mu(s_0) - \mu(s))\frac{\rho_j(s)}{s}\right\} + \partial^j s\partial_j\mu(s) + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\, d\sigma + O(\epsilon), \\
&= \frac{2}{s}((\mu(s_0) - \mu(s)) - \frac{d\mu(s)}{ds} + \frac{d\mu(s)}{ds} + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\, d\sigma + O(\epsilon) \\
&= \frac{2}{s}(\mu(s_0) - \mu(s)) + \int_{s_0}^s \partial^j\partial_j\mu(\sigma)\, d\sigma + O(\epsilon),
\end{aligned} \tag{B10}$$

where we used the relation $\partial_j s = \partial s/\partial\rho_j(s) + O(\epsilon) = \rho_j(s)/s + O(\epsilon)$ which follows from (B6) and from $\rho^j(s)\rho_j(s) = s^2$.

Combining (B9) and (B10) gives (B7).

[1]P. Hillion and S. Quinnez, "Thermal blooming with analytical diffraction approximated expressions," J. Math. Phys. **22**, 897 (1981).
[2]P. Hillion and S. Quinnez, "Scalar diffraction theory as a perturbation of the refractive index," J. Opt. (Paris) **11**(2), 1980.
[3]P. Hillion and S. Quinnez, "Propagation of a collimated Gaussian laser beam in a turbulent medium," (unpublished).
[4]A. Sommerfeld and J. Runge, Ann. Phys. **35**, 289 (1911).
[5]M. Kline and I.W. Kay, *Electromagnetic Theory and Geometrical Optics* (Interscience, New York, 1965).
[6]M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, New York, 1965).
[7]A. H. Nayfey, *Perturbation Methods* (Wiley, New York, 1973).
[8]H. Kogelnik and T. Li, Proc. I.E.E.E. **54**, 1312 (1966).
[9]V. I. Talanov, JETP Lett. **2**, 138 (1965) [JETP Pis'ma v Redaktsiyu, **2**, 218 (1965)].
[10]M. Kline, Comm. Pure. Appl. Math. **14**, 473 (1961).

2326,2331

Shift operator techniques for the classification of multipole-phonon states. VIII. R(5) ↓ SU(2) ⊗ SU(2) reduction for obtaining quadrupole O_l^0 eigenvalues

G. Vanden Berghe and H. E. De Meyer[a)]
Seminarie voor Wiskundige Natuurkunde, R. U. G., Krijgslaan 271-S9, B-9000 Gent, Belgium

(Received 30 December 1980; accepted for publication 20 February 1981)

With the aid of previously derived expressions for the reduced matrix elements of the $R(5)$ generators, which are obtained by considering $R(5)\downarrow \mathrm{SU}(2)\otimes \mathrm{SU}(2)$ reduction, a method is set up to evaluate analytical expressions for the eigenvalues of the quadrupole scalar shift operator O_l^0. Explicit expressions are listed for states with angular momenta $l=2v-k$ ($k=0$ up to 13).

PACS numbers: 03.65.Ca, 63.20.Dj

I. INTRODUCTION

In a set of previous papers[1–5] part of the quadrupole eigenvalue spectrum of the $R(3)$ scalar shift operator O_l^0 has been derived by means of relations existing between product operators $O_{l+k}^j O_l^k$ ($-3\leqslant j, k\leqslant 3$, and $0\leqslant |j+k|\leqslant 5$), the shift operator O_l^k and, if $j+k=0$, the seniority operator V^*. In the first series of three papers[1–3] have we restricted ourselves to the use of $R(3)$ scalar product operators. A tree generating mechanism has been developed by which it was possible to determine in an easy way O_l^0 eigenvalues for l-nondegenerated states. By introducing the nonscalar $R(3)$ product operators[4,5] the tree generating mechanism could be completely abandoned and new formulas for the O_l^0 eigenvalues valid for all possible seniority v could be derived. As an example of an l-degenerated case the $l=2v-6$ state has been considered. Although the shift operator technique applied to the symmetric irreducible representations of the $R(5)$ group is self-consistent, as was proved in Ref. 5, it is a recursive method where a certain O_l^0 eigenvalue $\alpha_{v,l}$ can only be derived if all $\alpha_{v,l'}$ with $l'>l$ are already known. In this paper we wish to discuss another method by which the $\alpha_{v,l}$ values can be obtained in a direct way, without having knowledge of other eigenvalues.

It will be proven that the O_l^0 eigenvalues are proportional to the reduced matrix elements of the $R(5)$ generator q_μ [see Ref. 1, expression (1.7)] between physical states, i.e., states with a definite angular momentum. Analytical expressions for such reduced matrix elements are available in the literature.[6] These results were obtained by using a "natural basis" in which the representations of $R(5)$ are fully reduced with respect to the subgroup $R(4)=\mathrm{SU}(2)\otimes\mathrm{SU}(2)$. The $R(5)$–$R(3)$ basis function, used for the evaluation of these matrix elements, have been projected from a small subset of the natural basis functions by Hill–Wheeler type integrals. In Sec. II we shall review the formulas we need for further use and rectify some results erroneously given by Williams and Pursey.[6] The O_l^0 eigenvalues for l-nondegenerated states will be derived in Sec. III, while the case where the $R(3)$ representation l occurs at most twice in the $R(5)$ representation $(v,0)$ is treated in Secs. IV and V.

II. BASIC FORMULAS

Using the explicit expressions given in Ref. 1 for the shift operators O_l^k ($k=-3,-2,-1,0,1,2,3$), applying the Wigner–Eckart theorem and introducing the analytical expression for the occurring 3-j symbol, it is easy to verify that

$$\begin{aligned}\alpha_{v,l} &\equiv \langle v,l,m|\,O_l^0|v,l,m\rangle\\ &=\frac{8}{\sqrt{5}}\,\frac{l(l+1)(l-1)(l+2)(2l-1)(2l+3)}{(2l+1)^{1/2}}\\ &\quad\times\langle v,l\,||q||\,v,l\rangle A_{30}(l),\end{aligned}\tag{2.1}$$

where $A_{jk}(l)$ is defined as follows[7]:

$$A_{jk}(l)=\left[\frac{(2l+j+k+1)^{(2j+1)}}{2l+2k+1}\right]^{-1/2},\tag{2.2}$$

with $U^{(x)}=U(U-1)\cdots(U-x+1)$. Here $|v,l,m\rangle$ is a shorthand notation for the quadrupole phonon state of seniority v, angular momentum l, and projection m.

By considering the $R(5)\downarrow\mathrm{SU}(2)\otimes\mathrm{SU}(2)$ reduction Williams and Pursey[6] introduce functions $\psi(v,\nu,l,m)$, which span the entire representation space of the irreducible representation $(v,0)$ of $R(5)$:

$$\psi(v,\nu,l,m)=\int D^{l*}_{m,K}(\Omega)\chi_\Omega(v,\nu)d\Omega,\tag{2.3}$$

where the label ν is introduced to resolve the degeneracy of l. It takes the values

$$\nu=0,1,2,\ldots,[v/3],\tag{2.4}$$

where $[v/3]$ denotes the integral part of $v/3$. The $D^l_{m,K}(\Omega)$ is an ordinary rotation matrix, while the $\chi(v,\nu)$ denotes the "intrinsic states," which form a small subset of the natural basis functions. Williams and Pursey[6] clearly prove that K, the l-projection of the intrinsic states, can take the values

$$K=v-3\nu,\tag{2.5}$$

while the possible values of l are

$$l=2K,\,2K-2,\,2K-3,\ldots,K.\tag{2.6}$$

The functions $\psi(v,\nu,l,m)$ defined by Eq. (2.3) are not normalized, and if two of them differ only in the value of ν, they are not orthogonal. Therefore, one has defined the Hilbert-space integral

$$A_l^v(\nu',\nu)\equiv(\psi(v,\nu',l,m),\,\psi(v,\nu,l,m)).\tag{2.7}$$

[a)]Bevoegdverklaard Navorser N.F.W.O. (Belgium).

When $\nu' = \nu$, $A^{\,v}_{\,l}(\nu,\nu)$ is the square of the normalization constant, and we have adopted the convention of taking the positive square root. For $\nu' \neq \nu$, (2.7) is the overlap integral for states of common l but different ν belonging to the irreducible representation $(v,0)$. Williams and Pursey[6] derive several analytical expressions for the $A^{\,v}_{\,l}(\nu',\nu)$ of which we withhold the following one:

$$A^{\,v}_{\,l}(\nu',\nu) = \frac{2^{\nu'-\nu}}{(2l+1)(K-K')!}$$
$$\times \left[\frac{(v-\nu)!(v-\nu')!\,\nu!\,\nu'!\,(l-K')!(l+K)!}{(l+K')!(l-K)!}\right]^{1/2}$$
$$\times \sum_{\beta} \frac{(-4)^{\nu-\beta}}{(v-\nu-\nu'+\beta)!(\nu'-\beta)!(\nu-\beta)!\,\beta!}$$
$$\times \int_0^1 dz(1-z)^{2(v-\nu-\nu'+\beta)} z^{3(\nu'-\beta)}(1-4z)^{v-\nu-\nu'+\beta}$$
$$\times {}_2F_1(K-l, K+l+1; K-K'+1; z). \tag{2.8}$$

This quantity is evaluated under the condition $\nu' \geqslant \nu$. If ν' is smaller than ν, Eq. (2.8) still applies, but with $\nu' \longleftrightarrow \nu$. This is equivalent to the symmetry rule

$$A^{\,v}_{\,l}(\nu',\nu) = A^{\,v}_{\,l}(\nu,\nu'). \tag{2.9}$$

This last relation was erroneously formulated by Williams and Pursey.[6] The authors present an expression for the reduced matrix elements of the generator q_μ with respect to the basis functions (2.3). The second term in this expression, which is based upon Eq. (23) in Ref. 6, is not correct. This can be easily understood as follows. The ten $R(5)$ generators in the $SU(2) \otimes SU(2)$ reduction are rearranged into twice three generators satisfying the SU(2) Lie algebras and four generators forming a bispinor.[8] The incorrect term follows from the application of the bispinor component $T^{[1/2\,1/2]}_{1/2\,1/2}$ to the intrinsic state function $\chi,(v,\nu)$. By considering very carefully the explicit matrix elements of the $T^{[1/2\,1/2]}_{\pm 1/2 \pm 1/2}$ we find that

$$T^{[1/2\,1/2]}_{1/2\,1/2}\chi(v,\nu) = -[\nu/(v-\nu+1)]^{1/2}\chi(v,\nu-1), \tag{2.10}$$

an expression which replaces the erroneous Eq. (23) in Ref. 6. Taking into account (2.10) and the results of Williams and Pursey[6] the reduced matrix element of q_μ with respect to (2.3) reads, in the present notation,

$$\langle v,\nu',l' \| q \| v,\nu,l \rangle$$
$$= [(2l+1)/10]^{1/2}$$
$$\times (\{ -[5\nu(v-\nu+1)]^{1/2}\langle lK\,33|l'\,K+3\rangle$$
$$-[5\nu l'(l'+1)/3(v-\nu+1)]^{1/2}$$
$$\times \langle l'\,K+3\;1\;-1|l'\,K+2\rangle$$
$$\times \langle lK\,32|l'\,K+2\rangle\} A^{\,v}_{\,l'}(\nu',\nu-1)$$
$$+[5(v-\nu)(\nu+1)]^{1/2}$$
$$\times \langle lK\,3\;-3|l'\,K-3\rangle A^{\,v}_{\,l'}(\nu',\nu+1)$$
$$+\{[\tfrac{3}{2}l'(l'+1)]^{1/2}\langle l'K\,11|l'\,K+1\rangle\langle lK\,31|l'\,K+1\rangle$$
$$-[\tfrac{2}{3}l'(l'+1)]^{1/2}\langle l'K\,1\;-1|l'\,K-1\rangle$$
$$\times \langle lK\,3\;-1|l'\,K-1\rangle$$
$$-(2v-\nu)\langle lK\,30|l'K\rangle\} A^{\,v}_{\,l'}(\nu',\nu)). \tag{2.11}$$

From the conditions (2.4)–(2.6) one finds for large enough v-values that all states with $l = 2v - k$ ($k = 0,2,3,4,5$, and 7) exist and are nondegenerated, that all states with $l = 2v - k'$ ($k' = 6,8,9,10,11$ and 13) exist and are doubly degenerated,... . In the following chapters we wish to study the $O^{\,0}_{\,l}$ eigenvalues for the classes of nondegenerated and doubly degenerated states.

III. THE $\alpha_{v,l}$-VALUES FOR THE NONDEGENERATED STATES WITH $\nu = 0$

For the case $\nu = 0$, which implies, due to (2.5), $K = v$, it follows that the normalized wave function for the states with angular momentum $l = 2v - k$ ($k = 0, 2,3,4,5$, and 7) can be defined as follows:

$$|v,l,m\rangle = [A^{\,v}_{\,l}(0,0)]^{-1/2}\psi(v,0,l,m), \tag{3.1}$$

and that the reduced q matrix element reads

$$\langle v,l \| q \| v,l \rangle$$
$$= [1/A^{\,v}_{\,l}(0,0)]\langle v,0,l \| q \| v,0,l\rangle$$
$$= [(2l+1)/10]^{1/2}(\sqrt{5}\nu\langle lv3\;-3|l\,v-3\rangle A^{\,v}_{\,l}(0,1)/A^{\,v}_{\,l}(0,0)$$
$$+\{[\tfrac{3}{2}l(l+1)]^{1/2}\langle lv11|l\,v+1\rangle\langle lv31|l\,v+1\rangle$$
$$-[\tfrac{2}{3}l(l+1)]^{1/2}\langle lv1\;-1|\,lv-1\rangle\langle lv3\;-1|l\,v-1\rangle$$
$$-2v\langle lv30|lv\rangle\}). \tag{3.2}$$

The occurring Clebsch–Gordan coefficients can all be written down analytically.[7] The expressions for the $A^{\,v}_{\,l}(\nu',\nu)$ directly follow from (2.8). For $A^{\,v}_{\,l}(0,0)$ one finds

$$A^{\,v}_{\,l}(0,0)$$
$$= \frac{1}{2l+1}\int_0^1 dz\,(1-z)^{2v}(1-4z)^v\,{}_2F_1(v-l, v+l+1;1;z), \tag{3.3}$$

an expression which can be transformed with the help of partial integration into

$$A^{\,v}_{\,l}(0,0) = \frac{4^{l-v}}{2l+1}\frac{v!}{(l-v)!(2v-l)!}$$
$$\times \int_0^1 (1-4z)^{2v-l}z^{l-v}(1-z)^{l+v}dz. \tag{3.4}$$

On the right-hand side we recognize an integral representation of the hypergeometric function, so that $A^{\,v}_{\,l}(0,0)$ finally takes the form

$$A^{\,v}_{\,l}(0,0) = \frac{4^{l-v}}{(2l+1)^2}\frac{v!(l+v)!}{(2l)!(2v-l)!}$$
$$\times {}_2F_1(l-2v, l-v+1; 2l+2; 4). \tag{3.5}$$

One can also remark that $A^{\,v}_{\,l}(0,0)$ becomes identically zero, for $l < v$, $l > 2v$, and $l = 2v - 1$, which shows that the $|v,l,m\rangle$ states with these specific l-values do not exist.

The determination of $A^{\,v}_{\,l}(0,1) = A^{\,v}_{\,l}(1.0)$ progresses along the same lines as was the case for $A^{\,v}_{\,l}(0,0)$. From (2.8) with $K = v$ and $K' = v - 3$ one obtains

$$A^{\,v}_{\,l}(1,0) = \frac{2}{(2l+1)3!}\left[\frac{v(l-v+3)!(l+v)!}{(l+v-3)!(l-v)!}\right]^{1/2}$$
$$\times \int_0^1 dz(1-z)^{2(v-1)}z^3(1-4z)^{v-1}$$
$$\times {}_2F_1(v-l, v+l+1; 4; z), \tag{3.6}$$

which again can be transformed into the simpler expression

$$A_l^v(1,0) = \frac{2\cdot 4^{l-v-1}}{(2l+1)^2}\frac{(v-1)!(v+l-3)!}{(2l)!(2v-l)!} \times\left[\frac{v(l-v+3)!(l+v)!}{(l+v-3)!(l-v)!}\right]^{1/2} \times\{v\,{}_2F_1(l-2v, l-v+4; 2l+2;4) + 3(2v-1)\,{}_2F_1(l-2v+1, l-v+4; 2l+2;4)\}. \tag{3.7}$$

Furthermore, it can be proven that for the l-values under consideration the following identity is valid[9]:

$$v\,{}_2F_1(l-2v,l-v+4; 2l+2;4) + 3(2v-1) \times{}_2F_1(l-2v+1, l-v+4;2l+2;4) = \{5l+9v+16[1-(-1)^l]\} \times {}_2F_1(l-2v, l-v+1;2l+2;4) \quad (2v-l=0,2,3,4,5,7). \tag{3.8}$$

Substituting (3.5), (3.7), and (3.8) into (3.2) and replacing the occurring Clebsch–Gordan coefficients by their analytical expression, the reduced matrix element of q_μ with respect to the normalized states (3.1) for $l=2v-k$ ($k=0,2,3,4,5,7$) takes the following form:

$$\langle v,l\,||q||v,l\rangle = [(2l+1)/10]^{1/2}A_{30}(l) \times\{5(l-v+3)^{(3)}[5l-9v+16(1-(-1)^l)] - [l(l+1)]^2 + 2l(l+1)[15v^2+15v+1] - 5v(v+1)(9v^2+v+2)\}. \tag{3.9}$$

By this the O_l^0 eigenvalue expression (2.1) reduces to the simple form

$$\alpha_{v,l} = (1/10\sqrt{2})[5(l-v+3)^{(3)} \times 5l-9v+16(1-(-1)^l)] - [l(l+1)^2 + 2l(l+1)(15v^2+15v+1) - 5v(v+1)(9v^2+v+2)] \quad \text{for } l=2v-k \quad (k=0,2,3,4,5,7). \tag{3.10}$$

The reader can convince himself that this single formula assumes all eigenvalues previously derived by means of the shift operator technique [Ref. 2, Eqs. (4.1)–(4.5); Ref. 5, Eqs. (2.4), (2.8), (3.15), (3.19), (3.20), (3.31)].

IV. DERIVATION OF $\alpha_{v,l}$ FOR DOUBLY DEGENERATED STATES: GENERAL THEORY

As already mentioned in Sec. 2, one can derive from (2.4)–(2.6) that for large enough v all states with $l=2v-k'$ ($k'=6,8,9,10,11,13$) are doubly degenerated. It is only for small v that states with these angular momenta are nondegenerated. These cases can be treated separately. Since the formulas of O_l^0-eigenvalues for these cases where $v<9$ are already obtained by another method,[2] we shall restrict ourselves here to the general situation.

For the doubly degenerated states, which we wish to consider here, we have at our disposal two basis functions of the type (2.3), i.e., the ones with $\nu=0$ and $\nu=1$. These basis functions are not normalized and are not orthogonal to each other. By using a suitable Hilbert–Schmidt procedure one can construct out of the basis functions two normalized and mutually orthogonal wavefunctions:

$$\Phi(v,l,\mathrm{I},m) = \frac{1}{[A_l^v(0,0)]^{1/2}}\psi(v,0,l,m), \tag{4.1}$$

$$\Phi(v,l,\mathrm{II},m) = \frac{[A_l^v(0,0)]^{1/2}}{\{A_l^v(1,1)A_l^v(0,0) - [A_l^v(0,1)]^2\}^{1/2}} \times\{\psi(v,1,l,m) - [A_l^v(0,1)/A_l^v(0,0)]\psi(v,0,l,m)\}. \tag{4.2}$$

These functions are found by construction eigenvectors of the Casimir operators of the $R(5)$, $R(3)$, and $R(2)$ groups, but not of O_l^0. The matrix elements of O_l^0 with respect to (4.1)–(4.2) can be written down, due to (2.1), as follows:

$$\langle v,l,i,m|O_l^0|v,l,j,m\rangle = \frac{8}{\sqrt{5}}\frac{l(l+1)(l-1)(l+2)(2l-1)(2l+3)}{(2l+1)^{1/2}} \times A_{30}(l)\langle v,l,i||q||v,l,j\rangle \quad \text{for } i,j=\mathrm{I} \text{ and } \mathrm{II}. \tag{4.3}$$

The following is due to (4.1) and (4.2):

$$\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle = 1/A_l^v(0,0)\langle v,0,l\,||q||v,0,l\,\rangle, \tag{4.4}$$

$$\langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle = A_l^v(0,0)/[A_l^v(1,1)A_l^v(0,0) - (A_l^v(0,1))^2] \times\{\langle v,1,l\,||q||v,1,l\,\rangle + [A_l^v(0,1)/A_l^v(0,0)]^2\langle v,0,l\,||q||v,0,l\,\rangle - [A_l^v(0,1)/A_l^v(0,0)][\langle v,1,l\,||q||v,0,l\,\rangle + \langle v,0,l\,||q||v,1,l\,\rangle]\}, \tag{4.5}$$

$$\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle = 1/[A_l^v(1,1)A_l^v(0,0) - (A_l^v(0,1))^2]^{1/2} \times[\langle v,0,l\,||q||v,1,l\,\rangle - [A_l^v(0,1)/A_l^v(0,0)]\langle v,0,l\,||q||v,0,l\,\rangle], \tag{4.6}$$

and

$$\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle = 1/[A_l^v(1,1)A_l^v(0,0) - (A_l^v(0,1))^2]^{1/2} \times[\langle v,1,l\,||q||v,0,l\,\rangle - [A_l^v(0,1)/A_l^v(0,0)]\langle v,0,l\,||q||v,0,l\,\rangle. \tag{4.7}$$

By this it is possible to construct explicitly the two by two matrix whose eigenvalues correspond to the $\alpha_{v,l}$ values we are looking for.

Although the described method is completely general and also applicable to cases of higher degeneracy, it gives rise to very complex calculations due to the complicated structure of each of the occurring matrix elements. Therefore, we prefer to develop in case of double degeneracy an alternative method to derive the O_l^0 eigenvalues. For this purpose we shall reformulate the problem. We try to determine two mutually orthogonal eigenvectors of the angular momentum operator L^2, the $R(5)$ Casimir operator V^*, and O_l^0 with respective eigenvalues $l(l+1)$, $-\frac{1}{2}v(v+3)$, and $\alpha_{v,l}^{(i)}$. In other words, we look for states $|v,l,i\rangle$ such that

$$O_l^0|v,l,i\rangle = \alpha_{v,l}^{(i)}|v,l,i\rangle \quad (i=1 \text{ or } 2). \tag{4.8}$$

Since these $|v,l,i\rangle$ states as well as the previously introduced $\Phi(v,l,j,m)$ ($j=\mathrm{I}$ or II) states form a complete set of orthonormalized basis vectors, in terms of which the L^2 and V^* eigenstates can be developed, there exists a simple relation between both sets:

$$|v,l,i\rangle = \sum_{j=1}^{\mathrm{II}} s_{ij}\Phi(v,l,j,m) \quad (i = 1 \text{ or } 2). \tag{4.9}$$

Due to the fact that both sets consist of orthonormalized vectors, the s_{ij} describes a simple rotation in two dimensions, further on defined in terms of the angle θ.

Using (4.3), (4.8), and (4.9), one easily deduces that

$$\langle v,l,i|O_l^0|v,l,j\rangle = \kappa \sum_{n,m} s_{in}s_{jm}\langle v,l,n||q||v,l,m\rangle, \tag{4.10}$$

where

$$\kappa = \frac{8}{\sqrt{5}}\,\frac{l(l+1)(l-1)(l+2)(2l-1)(2l+3)}{(2l+1)^{1/2}}A_{30}(l), \tag{4.11}$$

and where i,j take the values 1 and 2, while n and m can become I or II. For $i=j$ one obtains the following relations:

$$\begin{aligned}\alpha_{v,l}^{(1)} = \kappa\{&\cos^2\theta\,\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle\\ &+\sin^2\theta\,\langle v,\,,l,\mathrm{II}||\,q||v,\mathrm{II}\rangle\\ &+\cos\theta\sin\theta\,[\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle\\ &+\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle]\},\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}\alpha_{v,l}^{(2)} = \kappa\{&\sin^2\theta\,\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle\\ &+\cos^2\theta\,\langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle\\ &-\cos\theta\sin\theta\,[\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle\\ &+\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle]\}.\end{aligned} \tag{4.13}$$

From (4.12) and (4.13) it follows that the sum of the two eigenvalues, which is dependent of θ, is given by

$$\alpha_{v,l}^{(1)} + \alpha_{v,l}^{(2)} = \kappa[\langle v,l,\,\mathrm{I}||q||v,l,\mathrm{I}\rangle + \langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle]. \tag{4.14}$$

Assuming in (4.10) $i\neq j$ and taking into account that both basis sets consist of orthornormalized states, one obtains two other relations:

$$\begin{aligned}-\sin\theta\cos\theta\,&\langle v,l,\mathrm{I}\,||q||v,l,\mathrm{I}\rangle\\ &-\sin^2\theta\,\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle\\ &+\cos^2\theta\,\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle\\ &+\sin\theta\cos\theta\,\langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle = 0\end{aligned} \tag{4.15}$$

and

$$\begin{aligned}-\sin\theta\cos\theta\,&\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle\\ &+\cos^2\theta\,\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle\\ &-\sin^2\theta\,\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle\\ &+\sin\theta\cos\theta\,\langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle = 0.\end{aligned} \tag{4.16}$$

Adding first and subtracting then (4.15) and (4.16) results in

$$\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle = \langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle, \tag{4.17}$$

and

$$\begin{aligned}&2\cos 2\theta\,\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle\\ &\quad=\sin 2\theta\,[\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle - \langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle],\end{aligned} \tag{4.18}$$

respectively.

Taking into account (4.12), (4.13), (4.17), and (4.18), one can deduce that

$$\begin{aligned}\alpha_{v,l}^{(1)} - \alpha_{v,l}^{(2)} = \kappa\{&[\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle - \langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle]^2\\ &+4\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle^2\}^{1/2}.\end{aligned} \tag{4.19}$$

The eigenvalues we are looking for follow immediately from (4.14) and (4.19). The last-mentioned relations can all be expressed, on account of (4.4)–(4.7) in terms of $\langle v,i,l\,||q||v,j,l\,\rangle$ ($i,j = 1$ and 2), which are generally defined in (2.11). By introducing the analytical expressions for the occurring Clebsch–Gordan coefficients,[7] these matrix elements can be written in the following compact form:

$$\begin{aligned}\langle v,0,l\,||q||v,0,l\,\rangle = {}&[(2l+1)/10]^{1/2}A_{30}(l\,)\\ &\times[Q^{(1)}(l,v)A_l^v(0,1) + Q^{(2)}(l,v)A_l^v(0,0)]\end{aligned} \tag{4.20}$$

$$\begin{aligned}&\langle v,1,l\,||q||v,1,l\,\rangle = [(2l+1)/10]^{1/2}A_{30}(l\,)\\ &\times[Q^{(3)}(l,v)\,A_l^v(0,1) + Q^{(4)}(l,v)\,A_l^v(1,2) + Q^{(5)}(l,v)\,A_l^v(1,1)],\end{aligned} \tag{4.21}$$

$$\begin{aligned}\langle v,0,l\,||q||v,1,l\,\rangle = {}&[(2l+1)/10]^{1/2}A_{30}(l\,)\\ &\times[Q^{(1)}(l,v)A_l^v(1,1)Q^{(2)}(l,v)A_l^v(0,1)]\end{aligned} \tag{4.22}$$

$$\begin{aligned}\langle v,1,l\,||q||v,0,l\,\rangle = {}&[(2l+1)/10]^{1/2}A_{30}(l\,)\\ &[Q^{(3)}(l,v)A_l^v(0,0) + Q^{(4)}(l,v)A_l^v(0,2) + Q^{(5)}(l,v)A_l^v(0,1)],\end{aligned} \tag{4.23}$$

where

$$Q^{(1)}(l,v) = 10[v(l-v+3)^{(3)}(l+v)^{(3)}]^{1/2}, \tag{4.24}$$

$$\begin{aligned}Q^{(2)}(l,v) = {}&-[l(l+1)]^2 + 2l(l+1)(15v^2 + 15v + 1)\\ &-5v(v+1)(9v^2+v+2),\end{aligned} \tag{4.25}$$

$$Q^{(3)}(l,v) = (2/v)\,Q^{(1)}(l,v),$$

$$Q^{(4)}(l,v) = 10[2(v-1)(l+6-v)^{(3)}(l+v-3)^{(3)}]^{1/2}, \tag{4.26}$$

$$\begin{aligned}Q^{(5)}(l,v) = {}&-[l(l+1)]^2 + 2l(l+1)(15v^2-45v+1)\\ &+15(v-3)(-3v^3+17v^2-22v-8).\end{aligned} \tag{4.27}$$

Note that from (4.17), (4.6), and (4.7) it directly follows that

$$\langle v,0,l\,||q||v,1,l\,\rangle = \langle v,1,l\,||q||v,0,1\rangle. \tag{4.28}$$

This property however cannot be observed at first sight by comparing (4.22) and (4.23). In what follows we shall take profit of (4.28). The occurring $A_l^v(v',v)$ terms can be directly deduced from (2.8) by using analogous techniques as the ones leading to the results (3.5) and (3.7) for $A_l^v(0,0)$ and $A_l^v(1,0)$ respectively. The remaining $A_l^v(v',v)$ factors have the following form:

$$\begin{aligned}A_l^v(1,1) = {}&\frac{4^{l-v-1}}{(2l+1)^2}\,\frac{(v-1)!(v+l-3)!}{(2l\,)!(2v-l\,)!}\\ &\times\{v(v+3)^{(4)}\,{}_2F_1(1-2v,l-v+4;2l+2;4)\\ &+3(2v-l\,)(v+3)^{(4)}\,{}_2F_1(l-2v+1,l-v+4;2l+2;4)\\ &-6(v-10)(2v-l\,)^{(2)}(v+1)^{(2)}\\ &\times{}_2F_1(l-2v+2,l-v+4;2l+2;4)\\ &-2(5v-59)(2v-l\,)^{(3)}v\\ &\times{}_2F_1(l-2v+3,l-v+4;2l+2;4)\\ &+3(7v+20)(2v-l\,)^{(4)}\\ &\times{}_2F_1(l-2v+4,l-v+4;2l+2;4)\\ &-9(2v-l\,)^{(5)}\,{}_2F_1(l-2v+5,l-v+4;2l+2;4)\},\end{aligned} \tag{4.29}$$

$$A_l^v(2,1)=\frac{2\cdot 4^{l-v-2}}{(2l+1)^2}\frac{(v-2)!(v+l-6)!}{(2l)!(2v-l)!}$$
$$\times\Big[\frac{(v-1)(l-v+6)!(l+v-3)!}{2(l+v-6)!(l-v+3)!}\Big]^{1/2}$$
$$\times\{v(v+3)^{(5)}\,{}_2F_1(l-2v,l-v+7;2l+2;4)$$
$$+6(2v-l)(v+3)^{(5)}\,{}_2F_1(l-2v+1,l-v+7;2l+2;4)$$
$$+3(v+58)(2v-l)^{(2)}(v+1)^{(3)}$$
$$\times{}_2F_1(l-2v+2,l-v+7;2l+2;4)$$
$$-4(7v-149)(2v-l)^{(3)}v^{(2)}$$
$$\times{}_2F_1(l-2v+3,l-v+7;2l+2;4)$$
$$-9(v-92)(2v-l)^{(4)}(v-1)$$
$$\times{}_2F_1(l-2v+4,l-v+7;2l+2;4)$$
$$+54(v+7)(2v-1)^{(5)}$$
$$\times{}_2F_1(l-2v+5,l-v+7;2l+2;4)$$
$$-27(2v-l)^{(6)}\,{}_2F_1(l-2v+6,l-v+7;2l+2;4)\},\tag{4.30}$$

and

$$A_l^v(2,0)=\frac{4^{l-v-1}}{(2l+1)^2}\frac{(v-2)!(l+v-6)!}{(2l)!(2v-l)!}$$
$$\Big[\frac{v(v-l)(l-v+6)!(l-v)!}{2(l+v-6)!(l-v)!}\Big]^{1/2}$$
$$\times\{v^{(2)}\,{}_2F_1(l-2v,l-v+7;2l+2;4)+6(v-1)(2v-l)$$
$$\times{}_2F_1(l-2v+1,l-v+7;2l+2;4)$$
$$+9(2v-l)^{(2)}\,{}_2F_1(l-2v+2,l-v+7;2l+2;4)\}.\tag{4.31}$$

Due to (4.20)–(4.23), Eqs. (4.4)–(4.7) transform into

$$\langle v,l,\mathrm{I}||q||v,l,\mathrm{I}\rangle$$
$$=[(2l+1)/10]^{1/2}A_{30}(l)$$
$$\times[Q^{(1)}(l,v)A_l^v(0,1)/A_l^v(0,0)+Q^{(2)}(l,v)],\tag{4.32}$$

$$\langle v,l,\mathrm{II}||q||v,l,\mathrm{II}\rangle$$
$$=[(2l+1)/10]^{1/2}A_{30}(l)(-Q^{(1)}(l,v)A_l^v(0,1)/A_l^v(0,0)$$
$$+Q^{(4)}(l,v)[A_l^v(1,2)A_l^v(0,0)-A_l^v(0,2)A_l^v(0,1)]/$$
$$\{A_l^v(1,1)A_l^v(0,0)-[A_l^v(0,1)]^2\}+Q^{(5)}(l,v)],\tag{4.33}$$

$$\langle v,l,\mathrm{I}||q||v,l,\mathrm{II}\rangle=\langle v,l,\mathrm{II}||q||v,l,\mathrm{I}\rangle$$
$$=[(2l+1)/10]^{1/2}A_{30}(l)$$
$$\times\{A_l^v(1,1)A_l^v(0,0)-[A_l^v(0,1)]^2\}^{1/2}/A_l^v(0,0)\times Q^{(1)}(l,v).\tag{4.34}$$

By this the expressions for the sum and difference of the eigenvalues reduce to simple forms:

$$\alpha_{v,l}^{(1)}+\alpha_{v,l}^{(2)}$$
$$=\kappa[(2l+1)/10]^{1/2}A_{30}(l)(Q^{(2)}(l,v)+Q^{(5)}(l,v)+Q^{(4)}(l,v)$$
$$\times[A_l^v(1,2)A_l^v(0,0)-A_l^v(0,2)A_l^v(0,1)]/$$
$$\{A_l^v(1,1)A_l^v(0,0)-[A_l^v(0,1)]^2\}),\tag{4.35}$$

and

$$\alpha_{v,l}^{(1)}-\alpha_{v,l}^{(2)}$$
$$=\kappa[(2l+1)/10]^{1/2}A_{30}(l)\{X(l,v)^2+[4Q^{(1)}(l,v)/A_l^v(0,0)]$$
$$\times[A_l^v(1,0)X(l,v)+A_l^v(1,1)Q^{(1)}(l,v)]\}^{1/2},\tag{4.36}$$

where $X(l,v)$ is defined as

$$X(l,v)$$
$$=Q^{(2)}(l,v)-Q^{(4)}(l,v)[A_l^v(1,2)A_l^v(0,0)-A_l^v(2,0)A_l^v(1,0)]/$$
$$\times\{A_l^v(1,1)A_l^v(0,0)-[A_l^v(0,1)]^2\}-Q^{(5)}(l,v).\tag{4.37}$$

V. DERIVATION OF $\alpha_{v,l}$ FOR DOUBLY DEGENERATED STATES: APPLICATION

Starting from the formulas (4.35)–(4.37) it is possible to derive all $\alpha_{v,l}^{(i)}$ values for states with angular momenta $l=2v-k$ ($k=6,8,9,10,11$, and 13). For these specific l-values the term composed of $A_l^v(v',v)$ factors and which occurs in (4.35) as well as in (4.37) takes a very simple form, i.e.,

$$Q^{(4)}(l,v)\frac{A_l^v(1,2)A_l^v(0,0)-A_l^v(0,2)A_l^v(0,1)}{A_l^v(1,1)A_l^v(0,0)-[A_l^v(0,1)]^2}$$
$$=10(l+6-v)^{(3)}[5l-9(v-3)+16(1-(-1)^l)]$$
$$(2v-l=6,8,9,10,11,13).\tag{5.1}$$

Note that the last factor in the right-hand side of (5.1) looks very much like the first term in the right-hand side of (3.8); only the "v" is changed into "$v-3$." Due to (5.1), (4.35) and (4.37) reduce for these specific l-values to

$$\alpha_{v,l}^{(1)}+\alpha_{v,l}^{(2)}$$
$$=\kappa[(2l+1)/10]^{1/2}A_{30}(l)(-2[l(l+1)]^2$$
$$+4l(l+1)(15v^2-15v+1)$$
$$-5v(v+1)(9v^2+v+2)$$
$$+15(v-3)(-3v^3+17v^2-22v-8)$$
$$+10(l+6-v)^{(3)}\{5l-9(v-3)+16[1-(-1)^l]\})\tag{5.2}$$

and

$$X(l,v)=120l(l+1)-5v(v+1)(9v^2+v+2)$$
$$-15(v-3)(-3v^3+17v^2-22v-8)$$
$$-10(l+6-v)^{(3)}\{5l-9(v-3)+16[1-(-1)^l]\}.\tag{5.3}$$

In both expressions we have introduced the explicit forms for the $Q^{(i)}(l,v)$ terms. In fact, for the specific values we consider, (5.2) as well as (5.3) reduces to polynomials of the fourth degree in v. The only term in the derivation of the eigenvalues which we cannot write down in a general way is the second term in the right-hand side of (4.36). However, we have observed that for the six l-values under consideration the numerator is a multiple of the denominator and this expression can be written more simply in the following form:

$$\frac{4Q^{(1)}(l,v)}{A_l^v(0,0)}[A_l^v(0,1)X(l,v)+A_l^v(1,1)Q^{(1)}(l,v)]$$
$$=100\{2l+4+[1+(-1)^l]\}(l-v+3)^{(3)}Y(l,v),\tag{5.4}$$

where $Y(l,v)$ is a polynomial of fourth degree in v, i.e.,

$$Y(2v-6;v)=-v^4+6v^3+745v^2-1614v+1344,\tag{5.5}$$
$$Y(2v-8;v)=-v^4+31v^3+370v^2-2968v+4416,\tag{5.6}$$

$$Y(2v-9;v) = -v^4 - 20v^3 + 391v^2 - 1498v + 1320, \quad (5.7)$$

$$Y(2v-10,v) = -v^4 + 56v^3 - 419v^2 + 1048v - 660, \quad (5.8)$$

$$Y(2v-11;v) = -v^4 + 5v^3 + 343v^2 - 2909v + 6162, \quad (5.9)$$

$$Y(2v-13;v) = -v^4 + 30v^3 - 119v^2 - 294v + 1344. \quad (5.10)$$

Introducing these results into (4.36) and taking into account (5.2) and (5.3) we finally obtain the following results for the eigenvalues which are generally valid for all v:

$$\alpha_{2v-6}^{(1\text{ or }2)} = \frac{\sqrt{2}}{5}[(16v^4 - 116v^3 + 86v^2 + 194v - 165) \pm 5(64v^6 - 960v^5 + 6832v^4 - 21312v^3 + 32668v^2 - 26292v + 10161)^{1/2}], \quad (5.11)$$

$$\alpha_{2v-8}^{(1\text{ or }2)} = \frac{\sqrt{2}}{5}[(16v^4 - 180v^3 + 350v^2 + 60v - 321) \pm 5(64v^6 - 1728v^5 + 19696v^4 - 94560v^3 + 230620v^2 - 304572v + 181449)^{1/2}], \quad (5.12)$$

$$\alpha_{2v-9}^{(1\text{ or }2)} = \frac{\sqrt{2}}{5}[(16v^4 - 172v^3 + 74v^2 + 1522v - 2265) \pm 5(64v^6 - 960v^5 + 10288v^4 - 45504v^3 + 98332v^2 - 140340v + 123561)^{1/2}], \quad (5.13)$$

$$\alpha_{2v-10}^{(1\text{ or }2)} = \frac{\sqrt{2}}{5}[(16v^4 - 244v^3 + 806v^2 - 1154v + 1515) \pm 5(64v^6 - 2496v^5 + 37936v^4 - 224256v^3 + 615292v^2 - 799140v + 405441)^{1/2}], \quad (5.14)$$

$$\alpha_{2v-11}^{(1\text{ or }2)} = \frac{\sqrt{2}}{5}[(16v^4 - 236v^3 + 386v^2 + 1844v - 3975) \pm 5(64v^6 - 1728v^5 + 26608v^4 - 163680v^3 + 524380v^2 - 1071804v + 1149561)^{1/2}], \quad (5.15)$$

$$\alpha_{2v-13}^{(1\text{ or }2)} = \frac{\sqrt{2}}{5}[(16v^4 - 300v^3 + 890v^2 + 750 - 2181) \pm 5(64v^6 - 2496v^5 + 48304v^4 - 345216v^3 + 1106044v^2 - 1697700v + 1139625)^{1/2}]. \quad (5.16)$$

Naturally the result (5.11) for the $l = 2v - 6$ state is identical to the one obtained by the shift operator technique [Ref. 5, Eq. (3.30)]. The expressions for the other eigenvalues are new. It is striking that there is a great analogy between these expressions. In all cases the rational part of the expression consists of a polynomial of fourth degree in v and the coefficients of the highest power is always 16, while the irrational part is the square root of a polynomial of sixth degree in v where the coefficient of the highest power is always 64.

VI. DISCUSSION

In the present paper we have developed a method for calculating rigorously the expressions for the quadrupole O_l^0 eigenvalues. Although the method has been applied only to cases where nondegenerated and doubly-degenerated states are involved, it is completely general. Since the method is based on the fact that their exists a connection between the O_l^0 eigenvalue and the reduced matrix element of the $R(5)$ generator q_μ, it is only useful if analytical expressions for these matrix elements can be derived. In the $R(5)$ case this problem was already solved by Williams and Pursey by considering the $SU(2) \otimes SU(2)$ reduction of $R(5)$. We believe that the same method can also yield eigenvalue expressions for the scalar shift operators in the G_2 and $R(7)$ groups, which are involved in the description of the octupole-phonon case. However, the subgroup structure of these groups has not been studied in so much detail and expressions for the reduced matrix elements of the occurring generators are not available in the literature. The study of these problems will be considered in the near future.

[1]G. Vanden Berghe and H. E. De Meyer, J. Math. Phys. **21**, 1902 (1980).
[2]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980).
[3]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1913 (1980).
[4]G. Vanden Berghe and H. E. De Meyer, J. Math. Phys. **21**, 2632 (1980).
[5]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 2635 (1980).
[6]S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).
[7]A. P. Yutsis, I. B. Levinson, and V. V. Vanagas, *Mathematical Apparatus of the Theory of Angular Momentum* (Israel Program for Scientific Translations, Jerusalem, 1962).
[8]N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968).
[9]C. C. Grosjean (private communication).

Shift operator techniques for the classification of multipole-phonon states. IX. Properties of nonscalar R(3) product operators in the G_2 group

G. Vanden Berghe, H.E. De Meyer,[a] and J. Van der Jeugt[b]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent Krijgslaan 271-S9, B-9000 Gent, Belgium

(Received 9 April 1981; accepted for publication 18 May 1981)

Expressions connecting nonscalar R(3) products of operators shifting the eigenvalues of L^2 are constructed within the group G_2.

PACS numbers: 63.20.Hp, 02.20.Nq

I. INTRODUCTION

In a set of previous papers[1–8] (to be referred to as I–VIII) it was evident that operators shifting the eigenvalues l of the R(3) Casimir operator L^2 could play an important role in the classification of multipole-phonon states. By means of this specific technique the quadrupole-phonon state labeling problem could be completely solved.[7,8]

The symmetry group of the octupole Hamiltonian is the U(7) group. The labels specifying its symmetric representations, which are connected to the considered phonon states, are usually related to the Casimir operators of groups appearing in the chain

$$\mathrm{U}(7)\supset \mathrm{SU}(7)\supset \mathrm{R}(7)\supset \mathrm{G}_2\supset \mathrm{R}(3)\supset \mathrm{R}(2).$$

Only four independent ones can be deduced in this way. In Papers IV and V, G_2 and R(7) shift operators, i.e.,P_l^k $(-5\leqslant k\leqslant 5)$ and O_l^k $(-3\leqslant k\leqslant 3)$, respectively, have been introduced. It has been pointed out that either the P_l^0 or the O_l^0 operator could be used as fifth label generating operator for these U(7) irreps. For the R(7) case expressions connecting scalar as well as nonscalar R(3) quadratic products of these shift operators have been constructed (V and VI). For the G_2 group up to now only expressions between quadratic operator products of the scalar R(3) type have been reported (IV). It is rather important to notice that, in order to achieve proper relations between these quantities, one always needed six of the available 11 product operators. It then also followed that among the various relations which could be constructed only six independent ones exist. It was a striking property that only three of the six mentioned relations had to be deduced explicitly. Since by using the fact that every P_l^{+k} goes over in a P_l^{-k} on replacing l by $-(l+1)$, one easily derived from the first three constructed relations three other independent equations.

For the quadrupole case either a tree generating mechanism,[2] where only relations between product operators of the R(3) scalar type were involved, or a self-consistent single step algorithm,[7] where both types of relations were necessary, could be applied in order to obtain eigenvalues for the scalar R(3) shift operator. For the octupole-phonon situation we established that the first-mentioned mechanism does not generally lead to the P_l^0 and O_l^0 eigenvalues. Therefore, a study of quadratic product operators of the type $P_{l+k}^{+j}P_l^{+k}$ $(-5\leqslant j, k\leqslant 5$ and $0<|j+k|\leqslant 10)$ which have no R(3) scalar character is of great importance.

II. THE NONSCALAR R(3) PRODUCT OPERATORS AND THEIR MUTUAL RELATIONS

The quadratic operator $P_{l+k}^{+j}P_l^{+k}$ $(-5\leqslant j, k\leqslant 5$ and $0\leqslant|j+k|\leqslant 10)$ shifts the l values of the state upon which it acts by $(k+j)$. With the available shift operators (IV.2.4)–(IV.2.9) and the property (IV.2.3), which contain the components of the 11-dimensional tensor representation p_μ of R(3) [for definition, see (IV.1.5)] to first order only, one can construct 11 product operators with $s=j+k=0$, ten with $s=\pm 1$, nine with $s=\pm 2$, eight with $s=\pm 3$, seven with $s=\pm 4$, six with $s=\pm 5$, five with $s=\pm 6$, four with $s=\pm 7$, three with $s=\pm 8$, two with $s=\pm 9$, and one with $s=\pm 10$. For $s=0$ (the scalar case) we can refer to IV for the existing relations between the 11 product operators. It has to be noted that these expressions are only valid when they act to the right upon states with angular momentum projection $m=0$. It has been remarked that this seemingly drastic condition does not seriously detract from the generality of the presented calculations. Therefore, we shall work here also within this same convention.

The considered quadratic product operators consist of terms composed of two p_μ and ten or less R(3) generators l_i $(i=0,\pm)$. In order to obtain relations between them, it is clear that once again all product operators should be brought into a standard form. The procedure to reach that form has been fully discussed in I. It is evident, on account of the commuting properties of the p generators (see IV), that operators where p generators appear linearly can emerge in relations between quadratic products of the shift operators P_l^k. It is easy to understand that between expressions of the form $P_{l+k}^{+j}P_l^{+k}$, these operators linear in p_μ will be P_l^{j+k} themselves, if they are defined. By straightforward calculation we have arrived at the following final results for the case where $s<0$;

[a]Research Associate N.F.W.O. (Belgium).
[b]Research Assistant.

$$
\begin{aligned}
&(l+3)^2(l+4)^2(2l+5)(2l+7)^2(2l+9)P^{0}_{l-1}P^{-1}_{l} - (l+3)(l+4)^2(2l+1)(2l+7)^2(2l+9)(l+15)P^{-1}_{l}P^{0}_{l}\\
&\quad + 20l(l+4)^2(l+8)(2l+3)(2l+7)(2l+9)P^{-2}_{l+1}P^{+1}_{l}/(l+1)^2\\
&\quad + 60l(l+4)(l+6)(2l+1)(2l+5)(2l+9)P^{-3}_{l+2}P^{+2}_{l}/(l+1)^2(l+2)^2\\
&\quad + 30l(l+1)(2l+1)(2l+7)(4l+21)P^{-4}_{l+3}P^{+3}_{l}/(l+1)^2(l+2)^2(l+3)^2\\
&\quad + 30l(l+1)(2l+1)(2l+3)P^{-5}_{l+4}P^{+4}_{l}/(l+1)^2(l+2)^2(l+3)^2(l+4)^2\\
&\quad + (8/21\sqrt{3})l(l+1)(l+2)(l+3)^2(l+4)^2(2l+1)\\
&\quad \times(2l+3)(2l+5)(2l+7)^2(2l+9)P^{-1}_{l} = 0;
\end{aligned}
\tag{2.1}
$$

$$
\begin{aligned}
&(2l-1)(l+2)(l+3)^2(2l+5)(2l+7)(4l^2-8l-171)P^{0}_{l-1}P^{-1}_{l}\\
&\quad - (l-1)(l+3)(2l+1)(2l+5)(2l+7)(2l+9)(2l^2+17l-114)P^{-1}_{l}P^{0}_{l}\\
&\quad + 15(l-1)(2l-1)(2l+3)(l+3)(4l^3+32l^2-21l-350)P^{-2}_{l+1}P^{+1}_{l}/(l+1)^2\\
&\quad + 30l(l-1)(2l-1)(2l+5)(4l^2+12l-45)P^{-3}_{l+2}P^{+2}_{l}/(l+1)^2(l+2)^2\\
&\quad + 30l(l-1)(2l-1)(2l+1)(2l-3)P^{-4}_{l+3}P^{+3}_{l}/(l+1)^2(l+2)^2(l+3)^2\\
&\quad + 5(l+2)(l+3)^2(2l+3)(2l+5)^2(2l+7)P^{+1}_{l-2}P^{-2}_{l}/(l-1)^2\\
&\quad + (8/21\sqrt{3})(l-1)(2l-1)l(2l+1)(l+1)(2l+3)(l+2)\\
&\quad \times(2l+5)(l+3)^2(2l+7)(4l^2+4l+27)P^{-1}_{l} = 0;
\end{aligned}
\tag{2.2}
$$

$$
\begin{aligned}
&-(l+2)(l-2)(2l-1)(2l+3)(4l^4-24l^3-257l^2+444l+2773)P^{0}_{l-1}P^{-1}_{l}\\
&\quad + (l+2)(l-2)(2l+1)(2l-3)(4l^4+24l^3-257l^2-444l+2773)P^{-1}_{l}P^{0}_{l}\\
&\quad - 10(l-1)(l-2)(2l-3)(2l-5)(2l+3)(l^2+2l-20)P^{-2}_{l+1}P^{+1}_{l}/(l+1)^2\\
&\quad - 10(l-1)(l-2)^2(2l-1)(2l-3)(2l-5)P^{-3}_{l+2}P^{+2}_{l}/(l+1)^2(l+2)^2\\
&\quad + 10(l+1)(l+2)^2(2l+1)(2l+3)(2l+5)P^{+2}_{l-3}P^{-3}_{l}/(l-1)^2(l-2)^2\\
&\quad - 10(l+1)(l+2)(2l+3)(2l+5)(2l-3)(l^2-2l-20)P^{+1}_{l-2}P^{-2}_{l}/(l-1)^2\\
&\quad - (8/21\sqrt{3})l(l-1)(l-2)(l+1)(l+2)(2l-1)(2l-3)(2l+1)(2l+3)\\
&\quad \times(4l^4+15l^2+191)P^{-1}_{l} = 0;
\end{aligned}
\tag{2.3}
$$

$$
\begin{aligned}
&-(l+1)(l-3)(2l-1)(2l-5)(2l-7)(2l-9)(2l^2-17l-114)P^{0}_{l-1}P^{-1}_{l}\\
&\quad + (l-2)(l-3)^2(2l+1)(2l-5)(2l-7)(4l^2+8l-171)P^{-1}_{l}P^{0}_{l}\\
&\quad - 5(l-2)(l-3)^2(2l-3)(2l-5)^2(2l-7)P^{-2}_{l+1}P^{+1}_{l}/(l+1)^2\\
&\quad - 30l(2l-1)(l+1)(2l+1)(2l+3)P^{+3}_{l-4}P^{-4}_{l}/(l-1)^2(l-2)^2(l-3)^2\\
&\quad + 30l(l+1)(2l+1)(2l-5)(4l^2-12l-45)P^{+2}_{l-3}P^{-3}_{l}/(l-1)^2(l-2)^2\\
&\quad - 15(l+1)(l-3)(2l+1)(2l-3)(4l^3-32l^2-21l+350)P^{+1}_{l-2}P^{-2}_{l}/(l-1)^2\\
&\quad - (8/21\sqrt{3})l(l+1)(l-1)(l-2)(l-3)^2(2l+1)(2l-1)(2l-3)\\
&\quad \times(2l-5)(2l-7)(4l^2-4l+27)P^{-1}_{l} = 0;
\end{aligned}
\tag{2.4}
$$

$$
\begin{aligned}
&-(l-3)(l-4)^2(2l-1)(2l-7)^2(2l-9)(l-15)P^{0}_{l-1}P^{-1}_{l}\\
&\quad + (l-3)^2(l-4)^2(2l-5)(2l-7)^2(2l-9)P^{-1}_{l}P^{0}_{l}\\
&\quad + 30l(l-1)(2l-1)(2l-3)P^{+4}_{l-5}P^{-5}_{l}/(l-1)^2(l-2)^2(l-3)^2(l-4)^2\\
&\quad - 30l(l-1)(2l-1)(2l-7)(4l-21)P^{+3}_{l-4}P^{-4}_{l}/(l-1)^2(l-2)^2(l-3)^2\\
&\quad + 60l(l-4)(l-6)(2l-1)(2l-5)(2l-9)P^{+2}_{l-3}P^{-3}_{l}/(l-1)^2(l-2)^2\\
&\quad - 20l(l-4)^2(l-8)(2l-3)(2l-7)(2l-9)P^{+1}_{l-2}P^{-2}_{l}/(l-1)^2\\
&\quad - (8/21\sqrt{3})l(l-1)(l-2)(l-3)^2(l-4)^2(2l-1)(2l-3)\\
&\quad \times(2l-5)(2l-7)^2(2l-9)P^{-1}_{l} = 0;
\end{aligned}
\tag{2.5}
$$

$$
\begin{aligned}
&(l+2)(l+3)^2(l+4)(2l+3)(2l+5)^2(2l+7)P^{0}_{l-2}P^{-2}_{l}\\
&\quad + 35(l+2)(l+3)^2(l+4)(2l-1)(2l+5)(2l+7)P^{-1}_{l-1}P^{-1}_{l}\\
&\quad + (l-1)(l+3)(l+4)(2l+1)(2l+5)(2l+7)(l+10)(-2l+21)P^{-2}_{l}P^{0}_{l}\\
&\quad + 15(l-1)(l+3)(l+4)(2l-1)(2l+3)(2l^2+l-70)P^{-3}_{l+1}P^{+1}_{l}/(l+1)^2
\end{aligned}
$$

$$
\begin{aligned}
&+30(l-1)l(2l-1)(2l+5)(2l^2+3l-34)P_{l+2}^{-4}P_l^{+2}/(l+1)^2(l+2)^2\\
&+30(l-2)(l-1)l(2l-1)(2l+1)P_{l+3}^{-5}P_l^{+3}/(l+1)^2(l+2)^2(l+3)^2\\
&+(8/21\sqrt{3})(l-1)l(l+1)(l+2)(l+3)^2(l+4)(2l-1)\\
&\times(2l+1)(2l+3)(2l+5)^2(2l+7)P_l^{-2}=0;
\end{aligned}
\tag{2.6}
$$

$$
\begin{aligned}
&(l+1)(l+2)(l+3)(2l+3)(2l+5)(2l-3)(2l^2-l-106)P_{l-2}^{0}P_l^{-2}\\
&+35(l-2)(l+2)(l+3)(2l-1)(2l+3)(2l^2-l-43)P_{l-1}^{-1}P_l^{-1}\\
&+(l-2)(l+2)(2l-3)(2l+1)(-4l^4+4l^3+379l^2-64l-4515)P_l^{-2}P_l^{0}\\
&+10(l-1)(l-2)(l-3)(2l-3)(2l+3)(2l^2-53)P_{l+1}^{-3}P_l^{+1}/(l+1)^2\\
&+10(l-1)(l-2)(l-3)(2l-1)(2l-3)(2l-5)P_{l+2}^{-4}P_l^{+2}/(l+1)^2(l+2)^2\\
&+5(l+1)(l+2)^2(l+3)(2l+1)(2l+3)(2l+5)P_{l-3}^{+1}P_l^{-3}/(l-2)^2\\
&+(8/21\sqrt{3})l(l-1)(l-2)(l+1)(l+2)(l+3)(2l-1)(2l-3)\\
&\times(2l+1)(2l+3)(2l+5)(2l^2-5l+18)P_l^{-2}=0;
\end{aligned}
\tag{2.7}
$$

$$
\begin{aligned}
&(l+1)(l-3)(2l+1)(2l-3)(-4l^4+12l^3+367l^2-690l-4200)P_{l-2}^{0}P_l^{-2}\\
&-35(l+1)(l-3)(l-4)(2l-1)(2l-5)(2l^2-3l-42)P_{l-1}^{-1}P_l^{-1}\\
&+(l-2)(l-3)(l-4)(2l-5)(2l-7)(2l+1)(2l^2-3l-105)P_l^{-2}P_l^{0}\\
&-5(l-2)(l-3)^2(l-4)(2l-3)(2l-5)(2l-7)P_{l+1}^{-3}P_l^{+1}/(l+1)^2\\
&-10l(l+1)(l+2)(2l+1)(2l-5)(2l^2-4l-51)P_{l-3}^{+1}P_l^{-3}/(l-2)^2\\
&+10l(l+1)(l+2)(2l-1)(2l+1)(2l+3)P_{l-4}^{+2}P_l^{-4}/(l-2)^2(l-3)^2\\
&-(8/21\sqrt{3})(l-1)l(l+1)(l-2)(l-3)(l-4)(2l-1)(2l+1)\\
&\times(2l-3)(2l-5)(2l-7)(2l^2+l+15)P_l^{-2}=0;
\end{aligned}
\tag{2.8}
$$

$$
\begin{aligned}
&-l(l-4)(l-5)(2l-3)(2l-7)(2l-9)(l-11)(2l+19)P_{l-2}^{0}P_l^{-2}\\
&-35(l-3)(l-4)^2(l-5)(2l-1)(2l-7)(2l-9)P_{l-1}^{-1}P_l^{-1}\\
&+(l-3)(l-4)^2(l-5)(2l-5)(2l-7)^2(2l-9)P_l^{-2}P_l^{0}\\
&-15l(l-4)(l-5)(2l-1)(2l-5)(2l^2-5l-67)P_{l-3}^{+1}P_l^{-3}/(l-2)^2\\
&+30l(l-1)(2l-1)(2l-7)(2l^2-7l-29)P_{l-4}^{+2}P_l^{-4}/(l-2)^2(l-3)^2\\
&-30l(l+1)(l-1)(2l-1)(2l-3)P_{l-5}^{+3}P_l^{-5}/(l-2)^2(l-3)^2(l-4)^2\\
&-(8/21\sqrt{3})l(l-1)(l-2)(l-3)(l-4)^2(l-5)(2l-1)\\
&\times(2l-3)(2l-5)(2l-7)^2(2l-9)P_l^{-2}=0;
\end{aligned}
\tag{2.9}
$$

$$
\begin{aligned}
&(l+1)(l+2)^2(l+3)(2l+1)(2l+3)(2l+5)(2l+7)(l+13)P_{l-3}^{0}P_l^{-3}\\
&+20(l+1)(l+2)(l+3)(l-1)(2l+3)(2l+5)(2l+7)(2l+11)P_{l-2}^{-1}P_l^{-2}\\
&-l(l+2)(l-2)(2l+1)(2l+7)(4l^4+40l^3-145l^2-3220l-8439)P_l^{-3}P_l^{0}\\
&+5(l-1)(l-2)(2l+1)(2l+3)(2l+11)(2l^3+11l^2-111l-420)P_{l+1}^{-4}P_l^{+1}/(l+1)^2\\
&+10(l+1)(l-1)(l-2)(2l-1)(2l+21)(2l^2-5l-39)P_{l+2}^{-5}P_l^{+2}/(l+1)^2(l+2)^2\\
&-(2/21\sqrt{3})l(l+1)(l+2)^2(l+3)(2l+1)(2l+3)(2l+5)(2l+7)\\
&\times(l-1)(l-2)(2l-1)(2l^2+23l-879)P_l^{-3}=0;
\end{aligned}
\tag{2.10}
$$

$$
\begin{aligned}
&(l+1)(l+2)^2(l+3)(2l+1)(2l+3)(2l+5)(2l+7)P_{l-3}^{0}P_l^{-3}\\
&+20(l-1)(l+2)(l+3)(2l-1)(2l+3)(2l+7)(2l+11)P_{l-1}^{-2}P_l^{-1}\\
&+(l+2)(2l+1)(2l-1)(2l-3)(2l+7)(-l^3+3l^2+178l+660)P_l^{-3}P_l^{0}\\
&+5l(l-1)(2l-3)(2l+3)(2l+11)(2l^2-15l-98)P_{l+1}^{-4}P_l^{+1}/(l+1)^2\\
&+10(l-1)(2l-1)(2l+1)(2l-3)(l^2-5l-42)P_{l+2}^{-5}P_l^{+2}/(l+1)^2(l+2)^2\\
&-(2/21\sqrt{3})l(l+1)(l+2)^2(l+3)(l-1)(2l+1)(2l+3)(2l+5)\\
&\times(2l+7)(2l-1)(2l-3)(l-62)P_l^{-3}=0;
\end{aligned}
\tag{2.11}
$$

$$
\begin{aligned}
&(l-4)(2l-5)(2l-3)(2l-1)(2l-11)(-l^3+3l^2+178l-1020)P^0_{l-3}P_l^{-3}\\
&\quad -20(l-1)(l-4)(l-5)(2l-3)(2l-7)(2l-11)(2l-15)P_{l-2}^{-1}P_l^{-2}\\
&\quad +(l-3)(l-4)^2(l-5)(2l-5)(2l-7)(2l-9)(2l-11)P_l^{-3}P_l^0\\
&\quad -5(l-2)(l-1)(2l-1)(2l-7)(2l-15)(2l^2+7l-120)P_{l-4}^{+1}P_l^{-4}/(l-3)^2\\
&\quad +10(l-1)(2l-3)(2l-5)(2l-1)(l^2+l-48)P_{l-5}^{+2}P_l^{-5}/(l-3)^2(l-4)^2\\
&\quad +(2/21\sqrt{3})(l-1)(l-2)(l-3)(l-4)^2(l-5)(2l-1)(2l-3)\\
&\quad \times(2l-5)(2l-7)(2l-9)(2l-11)(l+60)P_l^{-3}=0;
\end{aligned}
\tag{2.12}
$$

$$
\begin{aligned}
&l(l-2)(l-4)(2l-5)(2l-11)(4l^4-72l^3+191l^2+3192l-15\,075)P^0_{l-3}P_l^{-3}\\
&\quad +20(l-1)(l-3)(l-4)(l-5)(2l-7)(2l-9)(2l-11)(2l-15)P_{l-1}^{-2}P_l^{-1}\\
&\quad -(l-3)(l-4)^2(l-5)(2l-5)(2l-7)(2l-9)(2l-11)(l-15)P_l^{-3}P_l^0\\
&\quad +5l(l-1)(2l-5)(2l-7)(2l-15)(2l^3-23l^2-43l+582)P_{l-4}^{+1}P_l^{-4}/(l-3)^2\\
&\quad -10l(l-1)(l-3)(2l-3)(2l-25)(2l^2-3l-41)P_{l-5}^{+2}P_l^{-5}/(l-3)^2(l-4)^2\\
&\quad -(2/21\sqrt{3})l(l-1)(l-2)(l-3)(l-4)^2(l-5)(2l-3)\\
&\quad \times(2l-5)(2l-7)(2l-9)(2l-11)(2l^2-31l-825)P_l^{-3}=0;
\end{aligned}
\tag{2.13}
$$

$$
\begin{aligned}
&l(l+1)(l+2)(l+3)(2l-1)(2l+1)(2l+3)(2l+5)P^0_{l-4}P_l^{-4}\\
&\quad +360(l+1)(l+3)(l-2)(2l-3)(2l+1)(2l+5)P_{l-2}^{-2}P_l^{-2}\\
&\quad +45(l+1)(l+3)(l+6)(l-9)(2l-1)(2l-3)(2l-5)P_{l-1}^{-3}P_l^{-1}\\
&\quad +(l-1)(l-2)(2l+1)(2l-5)(-4l^4+20l^3+513l^2-1071l-10\,935)P_l^{-4}P_l^0\\
&\quad +5(l-2)(l-4)(l+6)(l-9)(2l-1)(2l-3)(2l-5)P_{l+1}^{-5}P_l^{+1}/(l+1)^2\\
&\quad -(2/7\sqrt{3})(l-2)(l-1)l(l+1)(l+2)(l+3)(2l-5)(2l-3)(2l-1)\\
&\quad \times(2l+1)(2l+3)(2l+5)(2l-21)P_l^{-4}=0;
\end{aligned}
\tag{2.14}
$$

$$
\begin{aligned}
&l(l+1)(l+2)(l+3)(2l-1)(2l+1)(2l+3)(2l+5)P^0_{l-4}P_l^{-4}\\
&\quad +45l(l+1)(l+2)(l+3)(l-2)(2l+1)(2l+5)P_{l-3}^{-1}P_l^{-3}\\
&\quad +45(l+1)(l+3)(l-1)(l-3)(2l-1)(2l^2-7l-50)P_{l-1}^{-3}P_l^{-1}\\
&\quad +(l-2)(l-3)(2l-1)(2l+1)(-4l^4+20l^3+239l^2-660l-3375)P_l^{-4}P_l^0\\
&\quad +20l(l-2)(l-3)(l-4)(2l-3)(l^2-3l-19)P_{l+1}^{-5}P_l^{+1}/(l+1)^2\\
&\quad -(2/7\sqrt{3})(l-3)(l-2)(l-1)l(l+1)(l+2)(l+3)(2l-3)(2l-1)\\
&\quad \times(2l+1)(2l+3)(2l+5)(4l-25)P_l^{-4}=0;
\end{aligned}
\tag{2.15}
$$

$$
\begin{aligned}
&l(l-1)(2l-5)(2l-7)(-4l^4+28l^3+203l^2-882l-2988)P^0_{l-4}P_l^{-4}\\
&\quad -45l(l-2)(l-4)(l-6)(2l-5)(2l^2-5l-53)P_{l-3}^{-1}P_l^{-3}\\
&\quad -45(l-1)(l-3)(l-4)(l-5)(l-6)(2l-7)(2l-11)P_{l-1}^{-3}P_l^{-1}\\
&\quad +(l-3)(l-4)(l-5)(l-6)(2l-5)(2l-7)(2l-9)(2l-11)P_l^{-4}P_l^0\\
&\quad -20l(l+1)(l-1)(l-3)(2l-3)(l^2-3l-19)P_{l-5}^{+1}P_l^{-5}/(l-4)^2\\
&\quad +(2/7\sqrt{3})l(l-1)(l-2)(l-3)(l-4)(l-5)(l-6)(2l-3)\\
&\quad \times(2l-5)(2l-7)(2l-9)(2l-11)(4l+13)P_l^{-4}=0;
\end{aligned}
\tag{2.16}
$$

$$
\begin{aligned}
&l(l-1)(2l-3)(2l+3)(2l+5)(l+10)P^0_{l-5}P_l^{-5}+450l(l-2)(2l-5)(2l+5)P_{l-3}^{-2}P_l^{-3}\\
&+25(l-1)(l-2)(2l-7)(2l-15)(2l+15)P_{l-1}^{-4}P_l^{-1}-(l-10)(2l-3)(2l-5)(2l-7)(l^2+4l-75)P_l^{-5}P_l^0\\
&\quad +(2/7\sqrt{3})l(l-1)(l-2)(2l-3)(2l-5)(2l-7)(2l+3)(2l+5)(l^3-12l^2-163l+1410)P_l^{-5}=0;
\end{aligned}
\tag{2.17}
$$

$$
\begin{aligned}
&l(2l-3)(2l+5)(2l+7)P^0_{l-5}P_l^{-5}+25l(2l-5)(2l+5)P_{l-4}^{-1}P_l^{-4}\\
&\quad +25(l-4)(2l-3)(2l-13)P_{l-1}^{-4}P_l^{-1}-(l-4)(2l-5)(2l-13)(2l-15)P_l^{-5}P_l^0\\
&\quad +(4/7\sqrt{3})l(l-2)(l-4)(2l-3)(2l-5)(2l+5)(2l-13)(l^2-4l-57)P_l^{-5}=0;
\end{aligned}
\tag{2.18}
$$

$$l(l-1)(l+2)(2l-1)(2l+3)(2l+5)P^{0}_{l-5}P_l^{-5}+450l(l-2)(2l-5)(2l+5)P_{l-2}^{-3}P_l^{-2}$$
$$+25(l-2)(l-3)(2l-1)(2l+7)(2l-23)P_{l-1}^{-4}P_l^{-1}+(l-1)(l-3)(2l-5)(-4l^3+56l^2-47l-1610)P_l^{-5}P_l^0$$
$$+(2/7\sqrt{3})l(l-1)(l-2)(l-3)(l+2)(2l-1)(2l-5)(2l+3)(2l+5)(2l^2-45l+343)P_l^{-5}=0; \tag{2.19}$$
$$(l+1)(l-1)(l+2)(l-2)P_{l-5}^{-1}P_l^{-5}-10(l-1)(l+2)(l-3)P_{l-4}^{-2}P_l^{-4}$$
$$-10(l-2)(l-4)(l-7)P_{l-2}^{-4}P_l^{-2}-(l-3)(l-4)(l-6)(l-7)P_{l-1}^{-5}P_l^{-1}=0; \tag{2.20}$$
$$(l+1)(l-2)(l-1)(l+2)(2l+3)(2l-3)P_{l-5}^{-1}P_l^{-5}-270(l-1)(l+2)(l-3)(2l-5)P_{l-3}^{-3}P_l^{-3}$$
$$-10(2l-3)(2l-5)(2l-7)(l^2-5l-30)P_{l-2}^{-4}P_l^{-2}-(l-2)(l-3)(l-6)(2l-7)(2l^2-11l-93)P_{l-1}^{-5}P_l^{-1}=0; \tag{2.21}$$
$$(l-2)(2l+3)P_{l-5}^{-2}P_l^{-5}-30(l-3)P_{l-3}^{-4}P_l^{-3}-(l-12)(2l-7)P_{l-2}^{-5}P_l^{-2}=0; \tag{2.22}$$
$$(l+6)(2l-5)P_{l-5}^{-2}P_l^{-5}-30(l-3)P_{l-4}^{-3}P_l^{-4}-(l-4)(2l-15)P_{l-2}^{-5}P_l^{-2}=0; \tag{2.23}$$
$$5(2l-7)P_{l-4}^{-4}P_l^{-4}-(l+1)(l-3)P_{l-5}^{-3}P_l^{-5}+(l-8)(l-4)P_{l-3}^{-5}P_l^{-3}=0; \tag{2.24}$$
$$P_{l-5}^{-4}P_l^{-5}-P_{l-4}^{-5}P_l^{-4}=0. \tag{2.25}$$

For a fixed s, the mentioned relations are all completely independent. All other existing relations between the considered nonscalar operators for a particular s value can be derived from the cited ones. Using the fact that P_l^{+k} and P_l^{-k} go over into each other on replacing l by $-(l+1)$, 25 independent equations can be easily derived from (2.1)–(2.25). For these equations the present s values are essentially positive. The reader can easily perform this transformation himself if such relations are of importance.

III. DISCUSSION

As a first remark we want to draw attention to the fact that the total number of independent relations among nonscalar product operators in G_2 is not the one expected on grounds of analogy with the SU(3),[9] the R(5),[6] and the R(7)[6] cases, but is much larger indeed. To be more precise, let us recall that for the unitary and orthogonal groups mentioned we observed a common pattern in the variation of the number of independent relations among product operators shifting l to $l+s$, when we started at the scalar situation $s=0$ and systematically lowered (raised) s by one unit. Denoting for a fixed s the number of independent relations by N(s) we found for SU(3), R(5), and R(7) that $\mathrm{N}(s)=\mathrm{N}(0)-|s|$ if $1\leqslant s\leqslant \mathrm{N}(0)$ and $\mathrm{N}(s)=1$ for larger permissible $|s|$ values. In the present G_2 case, anomalies to this simple rule are encountered when $|s|$ equals 3,4,5,6, and 7. In fact, the reader can easily verify that for G_2 the foregoing formula should be replaced by $\mathrm{N}(s)=\mathrm{N}(0)-[|s|/2]-1$ for $1\leqslant|s|\leqslant 9$, where square brackets denote the integral part of the number inside. It remains an open question whether the first formula given above applies for all unitary and orthogonal Lie groups and hence also whether the exceptional behavior of G_2 should be attributed to the fact only that it is an exceptional Lie group.

It has been originally pointed out by Hughes and Yadegar[10] that it is always possible to turn an l-lowering shift operator into and l-raising one and vice versa by a formal change of the parameter l in the definition of that shift operator. For G_2 this property can be expressed by the operator equality $P_l^{-k}=P_{-l-1}^{+k}$. It has still been noted that on account of this equality the relations established in the present paper can be simply transformed into relations among product operators that raise l by 1,2,⋯, or 9 units. By setting up the relations (2.1)–(2.25), however, we noticed another kind of symmetry, which, for practical purposes, is a useful complement to the already known shift operator properties. Indeed, we want to conjecture that any of the relations among product operators that shift l to $l+s$ can be turned immediately into another relation (or exceptionally the same) of similar kind by carrying out the following operations:

(i) the parameter l in the coefficients and in these only should be formally replaced by $-l-s-1$;

(ii) each operator product of the form $P_{l+j}^{+k}P_l^{+j}$ $(j+k=s)$ should be replaced by the product $P_{l+k}^{+j}P_l^{+k}$;

(iii) shift operators occurring linearly in the relation should be kept unchanged.

Notice that for $s=0$ these operations become equivalent to the property $P_l^{-k}=P_{-l-1}^{+k}$ and hence allow us to derive one half of the number of scalar relations directly from the other half. For $s\neq 0$, however, we have now also the opportunity to set up only $[(\mathrm{N}(s)+1)/2]$ relations, after which the symmetry operations immediately lead to the complete set of N(s) independent relations. Furthermore, the reader can convince himself that the type of symmetry discussed here equally occurs in the groups SU(3), R(5), and R(7), and hence is independent of the particular group under consideration and merely a consequence of the fact that one deals with a linear Lie algebra of group generators.

Finally, it should be noticed that the nonscalar relations presented here occur with l-dependent coefficients, which allow much more factorizations than was the case for the scalar relations. In a forthcoming paper, we shall demonstrate how the appropriate use of scalar and nonscalar relations leads to very general formulae expressing the P_l^0 eigenvalues and eigenstates.

ACKNOWLEDGMENTS

We like to express our gratitute to Professor Dr. C. C. Grosjean for his continuing interest.

[1]G. Vanden Berghe and H. E. De Meyer, J. Math Phys. **21**, 1902 (1980).
[2]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1906 (1980).
[3]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1913 (1980).
[4]G. Vanden Berghe and H. E. De Meyer, J. Math. Phys. **21**, 1967 (1980).
[5]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 1973 (1980).
[6]G. Vanden Berghe and H. E. De Meyer, J. Math. Phys. **21**, 2632 (1980).
[7]H. E. De Meyer and G. Vanden Berghe, J. Math. Phys. **21**, 2635 (1980).
[8]G. Vanden Berghe and H. E. De Meyer, J. Math. Phys. **22**, 2326 (1981).
[9]H. E. De Meyer, G. Vanden Berghe, and J. W. B. Hughes, J. Math. Phys. **22**, 2360 (1981).
[10]J. W. B. Hughes and J. Yadegar, J. Math. Phys. **19**, 2068 (1978).

The inverse problem of the shear modulus and density profiles of a layered earth—torsional vibration data

Shimon Coen
Engineering Geoscience, Department of Materials Science and Mineral Engineering, University of California, Berkeley, California 94720

(Received 16 December 1980; accepted for publication 1 May 1981)

This paper shows that the shear modulus and density profiles of a layered earth are uniquely determined by the torsional stress and displacement on its surface at two frequencies. An analytical example is given in which these profiles are deduced from analytic data.

PACS numbers: 91.60.Ba, 03.40.Dz

I. INTRODUCTION

A torsional vibrator in the form of a rigid circular plate of finite radius is mounted on the free surface of a layered earth and is performing rotational oscillations about its center. Required are the shear modulus and density profiles of the layered earth from the measurements of the torsional stress under the vibrator and the torsional displacement of the surface of the earth. These data are assumed to be known precisely at two arbitrary frequencies. In a realistic geophysical experiment, the data is, of course, imprecise, and there may be in addition some difficulties in coupling torsional vibrations into the earth. Similar problems are encountered with the horizontally polarized shear wave vibrators, currently being used by the oil industry.

The static case corresponding to $\omega = 0$ has been already treated by Coen,[1] in which case the shear modulus profile of a layered earth is uniquely determined by the static torsional stress and displacement of the surface of the earth.

The equation describing the dynamic torsional displacement of a layered earth is transformed to the Schrödinger equation whose potential is related to the shear modulus and density profiles of the earth and the frequency of oscillation of the vibrator. Weidelt[2] solved the inverse problem for the Schrödinger equation by modifying the theory of Gel'fand and Levitan.[3] Weidelt's[2] theory is used in this paper to recover uniquely the potential of the Schrödinger equation from the torsional stress and displacement of the surface of the earth. If the surface data are available at two arbitrary frequencies, two different potentials will be reconstructed from which the shear modulus and density profiles of the earth can be separately obtained.

II. SOLUTION OF THE INVERSE PROBLEM

The dynamic torsional displacement $u(r,z,\omega) \equiv u_\theta(r,z,\omega)$ in the cylindrical coordinate system (r,θ,z) satisfies the equation

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} - \frac{u}{r^2} + \frac{\partial^2 u}{\partial z^2} + \frac{1}{\mu}\frac{\partial \mu}{\partial z}\frac{\partial u}{\partial z} + \frac{\omega^2}{c^2}u = 0, \quad r,z \geqslant 0, \tag{1}$$

where $\mu = \mu(z)$ is the shear modulus profile which is a function of the depth z only, $c = c(z)$ is the shear wave velocity profile which may be expressed in terms of the density profile $\rho(z)$, and the shear modulus profile by $c^2(z) = \mu(z)/\rho(z)$, ω is the frequency, and the displacement u is outgoing as $(r^2+z^2)^{1/2}$ tends to infinity.

A rigid torsional vibrator is place on the free surface $z = 0$, and it occupies the region $0 \leqslant r \leqslant a$, $0 < \theta \leqslant 2\pi$, where a is its radius in meters. The objective is to determine the shear modulus profile $\mu = \mu(z)$ and the shear wave velocity profile $c = c(z)$ from the torsional displacement $u(r,0,\omega_l)$, $0 \leqslant r < \infty$, and the torsional stress under the vibrator

$$\tau(r,\omega_l) = \mu \left.\frac{\partial u(r,z,\omega_l)}{\partial z}\right|_{z=0} \quad 0 \leqslant r \leqslant a, \tag{2}$$

at two frequencies $\omega_l, l = 1,2$. This is a nonlinear inverse problem. The transformation

$$v(r,z,\omega) = u(r,z,\omega)[\mu(z)/\mu(0)]^{1/2} \tag{3}$$

transform (1) into the equation

$$\frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} - \frac{v}{r^2} + \frac{\partial^2 v}{\partial z^2} + q(z,\omega)\, v = 0, \quad z \geqslant 0, \tag{4}$$

where

$$q(z,\omega) = \frac{\omega^2}{c^2(z)} - \frac{[(\mu(z))^{1/2}]''}{(\mu(z))^{1/2}}, \quad z \geqslant 0, \tag{5}$$

$$\tau(r,\omega) = \begin{cases} \{-\frac{1}{2}\mu'(z)v(r,z,\omega) + \mu(z)v'(r,z,\omega)\}_{z=0}, & 0 < r \leqslant a, \\ 0, \quad r > a. \end{cases} \tag{6}$$

In (5) and (6) the prime and double prime denote the first and second derivative with respect to z, respectively.

Let $\tilde{v}(\xi,z,\omega)$ denote the Hankel transform of order 1 of $v(r,z,\omega)$:

$$\tilde{v}(\xi,z,\omega) = \int_0^\infty v(r,z,\omega) r\, J_1(r\xi)\, dr. \tag{7}$$

Equation (4) now shows that

$$\frac{\partial^2 \tilde{v}(\xi,z,\omega)}{\partial z^2} - \xi^2 \tilde{v}(\xi,z,\omega) = -q(z,\omega)\, \tilde{v}(\xi,z,\omega), \quad z \geqslant 0, \tag{8}$$

and (6) transforms to

$$\tilde{\tau}(\xi,\omega) = \{-\tfrac{1}{2}\mu'(z)\tilde{v}(\xi,z,\omega) + \mu(z)\,\tilde{v}'(\xi,z,\omega)\}_{z=0}, \quad 0 \leqslant \xi \leqslant \infty, \tag{9}$$

In terms of the dimensionless function $f(\xi,z,\omega) = \tilde{v}(\xi,z,\omega)/\tilde{v}'(\xi,0,\omega)$, Eq. (8) is equivalent to

$$f''(\xi,z,\omega) - \xi^2 f(\xi,z,\omega) + q(z,\omega) f(\xi,z,\omega) = 0, \quad z \geqslant 0, \tag{10}$$

where f satisfies the initial conditions

$$f(\xi,0,\omega) = 1, \quad f'(\xi,0,\omega) = \frac{\eta'(0)}{\eta(0)} - \frac{1}{d(\xi,\omega)}, \tag{11}$$

in which $\eta(z) = [\mu(z)]^{1/2}$ and

$$d(\xi,\omega) = -\frac{\mu(0)\tilde{u}(\xi,0,\omega)}{\tilde{\tau}(\xi,\omega)} = -\frac{\tilde{u}(\xi,0,\omega)}{\tilde{u}'(\xi,0,\omega)}. \tag{12}$$

Note that since $\tau(r,\omega)$ and $u(r,0,\omega)$ were assumed to be known, $d(\xi,\omega)$ is a known function of ξ and ω.

If ω is regarded as a fixed parameter, then Weidelt[2] shows that the potential $q(z,\omega)$ can be uniquely recovered from the impedance function $d(\xi,\omega)$ provided that $d(\xi,\omega)$ is real for real ξ and it has no poles in the right-hand side of the complex ξ plane. It can be shown that $d(\xi,\omega)$ as given by Eq. (12) is not a real-valued function for all real ξ. The simplest example is a uniform half-space in which case

$$d(\xi,\omega) = \begin{cases} (\xi^2 - k^2)^{-1/2} & \text{for } \xi > k, \\ -i(k^2 - \xi^2)^{-1/2} & \text{for } k > \xi, \end{cases} \tag{13}$$

where $i^2 = -1$, $k = \omega/c$, and the time harmonic convention $e^{i\omega t}$ has been used.

However, if the independent variable ξ is changed to γ by

$$\xi = \left(\gamma^2 + \frac{\omega^2}{c_0^2}\right)^{1/2}, \tag{14}$$

where c_0 is a constant which is smaller than the lowest shear wave velocity in the earth

$$0 < c_0 < \min_{0 \leqslant z < \infty} c(z), \tag{15}$$

then it is shown in the Appendix that $d(\gamma,\omega)$ is a real-valued function for real γ, and it has no poles in the right-hand side of the complex γ plane. In the particular example given by (13), the change of variables (14) with the constraint (15) show that

$$d(\gamma,\omega) = \left[\gamma^2 + \omega^2\left(\frac{1}{c_0^2} - \frac{1}{c^2}\right)\right]^{-1/2}, \tag{16}$$

which is real for real γ because $c_0 < c$.

In the general case, Eq. (10) is modified to

$$f''(\gamma,z,\omega) - \gamma^2 f(\gamma,z,\omega) = Q(z,\omega) f(\gamma,z,\omega), \tag{17}$$

where

$$Q(z,\omega) = \frac{\eta''(z)}{\eta(z)} - \frac{\omega^2}{c_0^2}\left[\frac{c_0^2}{c^2(z)} - 1\right]. \tag{18}$$

Similarly, Eq. (11) is modified to

$$f(\gamma,0,\omega) = 1, \quad f'(\gamma,0,\omega) = \frac{\eta'(0)}{\eta(0)} - \frac{1}{d(\gamma,\omega)}. \tag{19}$$

It now follows from Weidelt[2] that for a fixed ω, the potential $Q(z,\omega)$ can be uniquely determined by the following procedure:

Step 1:

$$b(\gamma,\omega) = \tfrac{1}{2}[1 - \gamma d(\gamma,\omega)].$$

Step 2:

$$b(\gamma,\omega) = \int_0^\infty B(z,\omega) e^{-\gamma z}\, dz.$$

Step 3:

$$A(z,y;\omega) = B(z+y;\omega) + \int_{-z}^{z} A(z,x;\omega) \times [B(y+x;\omega) + B(y-x;\omega)]\, dx, \quad |y| \leqslant z, \quad z \geqslant 0.$$

Step 4:

$$Q(z,\omega) = 2\frac{d}{dz}A(z,z;\omega), \quad 0 \leqslant z < \infty.$$

If this procedure is applied to $d(\gamma,\omega_1)$ and $d(\gamma,\omega_2)$, $Q(z,\omega_1)$ and $Q(z,\omega_2)$ will be obtained and Eq. (18) shows that the shear wave velocity profile can be recovered by

$$\frac{c^2(z)}{c_0^2} = \frac{\omega_1^2 - \omega_2^2}{\omega_1^2 - \omega_2^2 + c_0^2[Q(z,\omega_2) - Q(z,\omega_1)]}, \quad 0 \leqslant z < \infty, \tag{20}$$

from which the shear modulus profile or, more precisely, the square root of the shear modulus profile, $\eta(z) = [\mu(z)]^{1/2}$ is obtained as the solution of the ordinary differential equation

$$\eta''(z) - \epsilon(z)\eta(z) = 0, \quad 0 \leqslant z < \infty, \tag{21}$$

where

$$\epsilon(z) = Q(z,\omega_1) + \frac{\omega_1^2}{c_0^2}\left[\frac{c_0^2}{c^2(z)} - 1\right], \quad 0 \leqslant z \leqslant \infty. \tag{22}$$

The initial conditions for (21) are

$$\eta(0) = [\mu(0)]^{1/2}, \quad \eta'(0) = \tfrac{1}{2}\mu^{-1/2}(0)\mu'(0), \tag{23a}$$

which are assumed to be known values. The referee suggests that $\eta(0)$ and $\eta'(0)$ can be deduced from the data $\tilde{\tau}(\gamma,\omega)/\tilde{u}(\gamma,0,\omega)$ because a WBK approximation for $\gamma \to \infty$ yields

$$\frac{\tilde{\tau}(\gamma,\omega)}{\tilde{u}(\gamma,0,\omega)} = -\eta^2(0)\left[\gamma + \frac{\eta'(0)}{\eta(0)} + \frac{Q(0,\omega)}{2\gamma} + O(\gamma^{-2})\right]. \tag{23b}$$

This completes the solution of the inverse problem; the shear wave velocity and shear modulus profiles have been obtained from the torsional stress under the vibrator and the torsional displacement of the surface of the earth at two frequencies. The density profile can be deduced from these by $\rho(z) = \mu(z)/c^2(z)$, $0 \leqslant z < \infty$.

III. AN ILLUSTRATIVE EXAMPLE

An example is now given in which the shear wave velocity and shear modulus profiles are reconstructed from $d(\gamma,\omega)$ as given by Eq. (16). The inversion procedure now gives:

Step 1:

$$b(\gamma) = \tfrac{1}{2}\frac{(\gamma^2 + h^2)^{1/2} - \gamma}{(\gamma^2 + h^2)^{1/2}}, \quad h^2 = \omega^2\left(\frac{1}{c_0^2} - \frac{1}{c^2}\right).$$

Step 2:

$$B(z) = \begin{cases} \tfrac{1}{2}h\, J_1(hz), & z > 0, \\ 0, & z < 0, \end{cases}$$

where $J_1(\cdot)$ is the Bessel function of order 1.

Step 3:

$$A(z,y) = \frac{h}{2}\frac{z+y}{(z^2 - y^2)^{1/2}} I_1\left(h(z^2 - y^2)^{1/2}\right) \quad |y| \leqslant z, \ z \geqslant 0,$$

where $I_1(\cdot)$ is the modified Bessel function of the first kind of order 1.

Step 4:

$$Q(z,\omega) = h^2.$$

This is obtained by noting that $A(z,z) = \frac{1}{2}h^2 z$ and then by differentiation. h^2 is by Step 1 equal to $\omega^2(1/c_0^2 - 1/c^2)$, in which $c_0 < c$.

Equation (20) now shows that

$$\frac{c^2(z)}{c_0^2} = \frac{\omega_1^2 - \omega_2^2}{\omega_1^2 - \omega_2^2 + c_0^2[\omega_2^2(1/c_0^2 - 1/c^2) - \omega_1^2(1/c_0^2 - 1/c^2)]} = \frac{c^2}{c_0^2}, \quad 0 \leqslant z < \infty, \tag{24}$$

and Eq. (21) shows that

$$\eta''(z) = 0, \quad 0 \leqslant z < \infty, \tag{25}$$

because $\epsilon(z) = 0$, $0 \leqslant z < \infty$, by Eq. (22). The general solution of Eq. (25) is given by

$$\eta(z) = \alpha_1 z + \alpha_2, \quad 0 \leqslant z < \infty, \tag{26}$$

where α_1 and α_2 are to be determined by the initial conditions (23), which for this example are $\eta(0) = [\mu(0)]^{1/2}$ and $\eta'(0) = 0$ because the data $d(\gamma,\omega)$ corresponds to a uniform half-space. Applying the initial conditions results in

$$\eta(z) = \eta(0), \quad 0 \leqslant z < \infty, \tag{27}$$

which states the medium has a uniform shear modulus profile $\mu(z) = \mu(0)$, $0 < z < \infty$, and by Eq. (24) the medium has a uniform shear wave velocity profile c, $0 \leqslant z < \infty$, and, consequently, the density profile is uniform too.

IV. DISCUSSION

It has been demonstrated that the shear modulus and density profiles of a layered earth are uniquely determined by the torsional stress and displacement of the surface of the earth at two frequencies. This requires, however, an estimate of the lowest shear wave velocity within the earth and that the shear modulus and its derivative be known at the surface of the earth.

For real data the most difficult part in the inversion procedure is Step 2. This requires the numerical inversion of the Laplace transform $b(\gamma,\omega)$. This is discussed by Coen.[1]

It will be concluded by noting that the inversion procedure of this paper reduces to the inversion procedure for the static case[1] by setting $\omega = 0$.

ACKNOWLEDGMENTS

I am grateful to Peter Weidelt for stimulating discussions on the title problem. This research was supported by the Department of Energy Contract No. GF26790.

APPENDIX

Here it is shown that the impedance function $d(\xi,\omega)$ as given by Eq. (12) is real for real γ, and it has no poles in the right-hand side of the complex γ plane, where γ is related to ξ by Eq. (14) and c_0 satisfies the constraints (15).

If $\tilde{u}(\xi,z,\omega)$ denotes the Hankel transform of $u(r,z,\omega)$, then Eq. (1) shows that

$$\left[\frac{\partial^2}{\partial z^2} - \xi^2 + \frac{\omega^2}{c^2(z)} + \frac{1}{\mu(z)}\frac{d\mu(z)}{dz}\frac{\partial}{\partial z}\right]\tilde{u}(\xi,z,\omega) = 0, \quad 0 \leqslant z < \infty, \tag{A1}$$

where

$$\tilde{u}(\xi,z,\omega) = \int_0^\infty u(r,z,\omega)\, r\, J_1(r\xi)\, dr. \tag{A2}$$

The depth z is now transformed to an apparent depth s by

$$\frac{ds}{dz} = \frac{\mu(0)}{\mu(z)}, \tag{A3}$$

and Eq. (A1) is transformed to

$$\left[\frac{\partial^2}{\partial s^2} - \frac{\mu^2(s)}{\mu^2(0)}\xi^2 + \omega^2\frac{\mu^2(s)}{\mu^2(0)c^2(s)}\right]\tilde{u}(\xi,s,\omega) = 0, \quad 0 \leqslant s < \infty, \tag{A4}$$

where

$$s = \int_0^z \frac{\mu(0)}{\mu(z')}dz'. \tag{A5}$$

Note that because $\mu(z')$ is a positive valued function, $s = s(z)$ is a monotonic function of z.

The independent variable ξ is next changed to γ by Eq. (14); this changes Eq. (A4) into

$$\left\{\frac{\partial^2}{\partial s^2} - \gamma^2\frac{\mu^2(s)}{\mu^2(0)} + \frac{\omega^2}{c_0^2}\left[\frac{c_0^2}{c^2(s)} - 1\right]\frac{\mu^2(s)}{\mu^2(0)}\right\}\tilde{u}(\gamma,s,\omega) = 0, \quad 0 \leqslant s < \infty. \tag{A6}$$

Now multiply both sides of this equation by $\tilde{u}^*(\gamma,s,\omega)$, where the asterisk denotes complex conjugate, and integrate with respect to s over $(0,\infty)$; upon integration by parts, this results in

$$-\tilde{u}^*(\gamma,0,\omega)\tilde{u}'(\gamma,0,\omega) - \int_0^\infty |\tilde{u}'(\gamma,s,\omega)|^2 ds - \gamma^2\int_0^\infty \frac{\mu^2(s)}{\mu^2(0)}|\tilde{u}(\gamma,s,\omega)|^2 ds + \frac{\omega^2}{c_0^2}\int_0^\infty\left[\frac{c_0^2}{c^2(s)} - 1\right]\frac{\mu^2(s)}{\mu^2(0)}|\tilde{u}(\gamma,s,\omega)|^2 ds = 0. \tag{A7}$$

The poles and zeros of $d(\gamma,\omega)$ in the complex γ plane are the zeros of $\tilde{u}'(\gamma,0,\omega)$ and $\tilde{u}(\gamma,0,\omega)$, respectively, on the complex γ plane because $d(\gamma,\omega) = -\tilde{u}(\gamma,0,\omega)/\tilde{u}'(\gamma,0,\omega)$. Therefore, the poles and zeros of $d(\gamma,\omega)$ in the complex γ plane are from Eq. (A7) given by

$$\gamma^2 = \left\{ -\int_0^\infty |\tilde{u}'(\gamma,s,\omega)|^2 ds - \frac{\omega^2}{c_0^2}\int_0^\infty \left[1 - \frac{c_0^2}{c^2(s)}\right] \frac{\mu^2(s)}{\mu^2(0)} |\tilde{u}(\gamma,s,\omega)|^2 ds \right\} \left[\int_0^\infty \frac{\mu^2(s)}{\mu^2(0)} |\tilde{u}(\gamma,s,\omega)|^2\, ds \right]^{-1}, \tag{A8}$$

which, in view of the constraints (15), namely $0 < c_0 < \min_{0 \leqslant s < \infty} c(s)$, $\gamma^2 < 0$, and this states that the poles and zeros of $d(\gamma,\omega)$ are located on the imaginary axis of the complex γ plane. So, $d(\gamma,\omega)$ has no poles in the right-hand side of the complex γ plane.

It will be next shown that $d(\gamma,\omega)$ is real for real γ. If the constraints (15) is satisfied Eq. (A7) shows that

$$\tilde{u}(\gamma,0,\omega)\tilde{u}'(\gamma,0,\omega) < 0 \tag{A9}$$

for real γ. From the definition of $d(\gamma,\omega)$ it follows that $u'(\gamma,0,\omega) = -u(\gamma,0,\omega)/d(\gamma,\omega)$ and Eq. (A9) shows that

$$|\tilde{u}^*(\gamma,0,\omega)|^2/d(\gamma,\omega) > 0, \tag{A10}$$

from which it is concluded that $d(\gamma,\omega)$ is real and positive for real γ.

[1]S. Coen, "The inverse problem of the shear modulus profile of a layered earth," J. Geophys. Res. **85** (310), 5364–6 (1980).

[2]P. Weidelt, "The inverse problem of geomagnetic induction," Z. Geophys. **38**, 257–89 (1972).

[3]I. M. Gel'fand and B. M. Levitan, "On the determination of a differential equation from its spectral function," Trans. Amer. Math. Soc. Series **2.1**, 253–304 (1955).

Erratum: The Gel'fand–Levitan equation can give simple examples of non-self-adjoint operators with complete eigenfunctions and spectral representations. I. Ghosts and resonances [J. Math. Phys. 21, 1716 (1980)].

B. DeFacio
Ames Laboratory-USDOE and Department of Mathematics, Iowa State University, Ames, Iowa 50011

H. E. Moses
Center for Atmospheric Research, University of Lowell, Lowell, Massachusetts 01854

(Received 30 December 1980; accepted for publication 9 January 1981)

PACS numbers: 03.65.Ge, 02.30.Jr, 99.10. + g

Equation (41) should read

$$V(r) = \frac{6r(r^3 - 6C)}{(r^3 + 3C)^2}.$$

Erratum: The Gel'fand–Levitan equation can give simple examples of non-self-adjoint operators with complete eigenfunctions and spectral representations. I. Ghosts and resonances [J. Math. Phys. 21, 1716 (1980)].

B. DeFacio
Ames Laboratory-USDOE and Department of Mathematics, Iowa State University, Ames, Iowa 50011

H. E. Moses
Center for Atmospheric Research, University of Lowell, Lowell, Massachusetts 01854

(Received 30 December 1980; accepted for publication 9 January 1981)

PACS numbers: 03.65.Ge, 02.30.Jr, 99.10. + g

Equation (41) should read

$$V(r)=\frac{6r(r^3-6C)}{(r^3+3C)^2}.$$