Character generators for compact semisimple Lie groups

R. W. Gaskell
Department of Physics, Lafayette College, Easton, Pennsylvania 18042

(Received 27 January 1983; accepted for publication 8 April 1983)

A new character formula is presented which leads automatically to character generators in the positive form. The character generator for B_3 is constructed as an example.

PACS numbers: 02.20.Qs

1. INTRODUCTION

The problem of determining character generators for compact semisimple Lie groups has recently attracted a great deal of interest. Such generators are functions $X(A,\eta)$ of l variables A_i and l class labels η_i such that when $X(A,\eta)$ is expanded in a power series in A, the coefficients represent the characters of the irreducible representations labeled by the exponents of the A:

$$X(A,\eta)=\sum_{\lambda}\chi_{\lambda}(\eta)\prod_{i=1}^{l}A_i^{\lambda_i}. \tag{1.1}$$

The number l is the rank of the Lie group. The characters $\chi_\lambda(\eta)$ are polynomials in the η_i such that the exponents of the η in a given term designate the components of a weight in the irreducible representation (λ) while the coefficient of the term tells us the multiplicity of that weight.

In order to be useful as a character generator, $X(A,\eta)$ should be a sum of terms, each of which has no negative coefficient in its expansion. Otherwise, the evaluation of (1.1) would involve the cancellation of contributions from different terms. A general formula which does not satisfy this requirement has been known for some time.[1] The difficulty of rewriting the character generator in the required positive-definite form, however, increases dramatically with the rank of the group. Several authors have recently presented combinatorical methods which can be used to obtain positive definite character generators for the groups A_n, B_n, C_n, and D_n.[2-5] Here we present a general formula which can be used to obtain a positive-definite character generator for any compact semisimple Lie group.

In Sec. 2, we review the Weyl formula for the character generator, which is not positive definite, and discuss the Weyl reflection group. In Sec. 3, we derive a similar formula which does yield a positive-definite result. Section 4 gives an algorithm for the simple determination of the elements of the Weyl group and further simplifies the character formula. In Sec. 5 we discuss the construction of character generators and present some examples. Finally, in Sec. 6 we discuss other applications of our character formula.

2. THE WEYL FORMULA

The character of an irreducible representation (λ) of a compact semisimple Lie group can be written as[6]

$$\chi_\lambda(\eta)=\xi_\lambda(\eta)/\xi_0(\eta), \tag{2.1}$$

where the characteristic function ξ_λ is given by

$$\xi_\lambda(\eta)=\sum_S \det(S)\prod_{k=1}^{l}(\eta_k^s)^{(1+\lambda_k)} \tag{2.2}$$

with

$$\eta_k^s=\prod_{i=1}^{l}\eta_i^{S_{ik}}. \tag{2.3}$$

In Eqs. (2.1)–(2.3) we have used the following notations and conventions. The irreducible representations (IR's) are labeled by the set of l integers λ_k such that the highest weight M of an IR is given by $M=\lambda_k m_k$ with m_k the highest weight of the k th fundamental IR. l is the rank of the Lie group. The basis of weight space is such that the m_k form an orthonormal set, $m_k=e_k$. The character χ defined in Eq. (2.1) is a polynomial in the η with terms of the form $\Pi\eta_i^{w_i}$, where w_i is the ith component of a weight w of the IR in our basis. Finally, the sum in Eq. (2.2) is over all Weyl reflections. The matrices S_{ik} in Eq. (2.3) are the representation matrices of the Weyl group in the basis of weight space.

The Weyl reflection group is generated by a set of l elements S_i under the restriction[7]

$$(S_iS_j)^{P_{ij}}=I \quad (i,j=1,\ldots,l). \tag{2.4}$$

If $i=j$, then the exponent P_{ij} in (2.4) is equal to 1, that is, any generator is its own inverse. If $i\neq j$, then we can obtain the exponents from the Dynkin diagram. The generator S_n is associated with the nth circle of the diagram and the exponent P_{ij} is 2, 3, 4, or 6 corresponding to the number of lines joining the ith and jth circles being 0, 1, 2, or 3. To construct the group elements, we proceed iteratively. With each element we associate a "word," a product of w of the S_n where w, the word length, is the minimum number of generators needed to construct the element. Words of length $w+1$ are obtained by multiplying all words of length w by each of the l generators S_n. Words which can be reduced by the relations (2.4) to a word which has already been found, either of the same length or shorter, are then discarded. The starting point is the identity I, which is taken to have length zero. The number of distinct words is the order of the Weyl group. These are given in Table I for the classical Lie groups. These

TABLE I. Orders of the Weyl reflection groups for the classical Lie groups.

Lie group	Order of Weyl group
A_n	$(n+1)!$
B_n, C_n	$2^n\cdot n!$
D_n	$2^{n-1}\cdot n!$
G_2	12
F_4	1 152
E_6	51 840
E_7	2 903 040
E_8	232 243 200

 0022-2488/83/102379-08$02.50

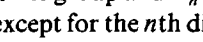
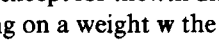
also represent the number of terms in the sum in Eq. (2.2).

In the basis of weight space, the matrices representing the generators $\mathbf{S}_n$ are

$$\mathbf{S}_n = \mathbf{I} - \mathbf{A}\cdot\mathbf{P}_n \quad (n = 1,...,l), \tag{2.5}$$

where $\mathbf{I}$ is the l-dimensional unit matrix, $\mathbf{A}$ is the Cartan matrix of the Lie group and $\mathbf{P}_n$ is the matrix whose elements are all zero except for the nth diagonal element which is one. When acting on a weight $\mathbf{w}$ the S_n have the effect

$$S_n\colon\ \mathbf{w} = \mathbf{w} - w_n\boldsymbol{\alpha}_n, \tag{2.6}$$

where w_n is the nth component of $\mathbf{w}$ and $\boldsymbol{\alpha}_n$ is the nth simple root, whose components in this basis are equal to the nth column of the Cartan matrix. From (2.3) we see that the k th column of each of the S_{ik} is one of the outside weights of the k th fundamental IR. An outside weight of an IR is one which is obtained from the highest weight by Weyl reflection. The number of outside weights in the k th fundamental IR is equal to the ratio of the order of the Weyl group to the order of the Weyl subgroup obtained by deleting the k th circle from the Dynkin diagram. For example, the Dynkin diagram for B_3 is  . Eliminating the first circle leaves  , the diagram for B_2. The number of outside weights in the (1,0,0) IR of B_3 is therefore equal to the order of the B_3 Weyl group divided by that of the B_2 Weyl group: $48/8 = 6$.

A generating function can be constructed for the Weyl characteristic function by multiplying (2.2) by $\Pi A_k^{\lambda_k}$ and summing over all $(\lambda) \geqslant (0)$. The result is

$$\Xi(A,\eta) = \sum_S \det(S) \prod_{k=1}^{l} \frac{\eta_k^s}{1 - A_k\eta_k^s}. \tag{2.7}$$

The coefficient of $A_k^{\lambda_k}$ in the expansion of (2.7) is the characteristic function $\xi_\lambda(\eta)$. In order to turn (2.7) into a generating function for characters, we simply divide by $\xi_0(\eta)$:

$$X(A,\eta) = \Xi(A,\eta)/\xi_0(\eta). \tag{2.8}$$

The coefficient of $A_k^{\lambda_k}$ in the expansion of (2.8) is the character $\chi_\lambda(\eta)$. The scalar characteristic can be written as

$$\xi_0(\eta) = \prod_p Q_p^{1/2}(1 - Q_p^{-1}), \tag{2.9}$$

where the index p labels positive roots and

$$Q_p = \prod_{j=1} \rho_j^{n_{jp}}, \quad \rho_j = \prod_{i=1} \eta_i^{A_{ij}}. \tag{2.10}$$

Here n_{jp} is the number of simple roots $\boldsymbol{\alpha}_j$ contained in the pth positive root and the exponent of η_i in ρ_j is the ith component of $\boldsymbol{\alpha}_j$. As usual, A_{ij} is the Cartan matrix.

While Eq. (2.8) provides us with a closed form expression for the character generator, it is not in a form we would like. First, the factor $\det(S)$ in (2.7) is equal to $(-1)^w$, where w is the word length of S. Therefore, terms appear with both positive and negative coefficients. Second, $\xi_0(\eta)$ introduces poles into each term of (2.8) which are spurious since they must disappear when all terms are added. Our problem is to rewrite (2.8) as a sum of rational expressions, each of which has a polynomial numerator with positive coefficients and a denominator which is a product of factors of the form $(1 - A_i\Pi\eta_j^{w_j})$, where A_i is the variable labeling the ith fundamental IR and w_j is the jth component of one of the outside weights of that IR. By counting the denominator factors in each term of (2.8) we can see that each of the rational expressions in the final form will have $\frac{1}{2}(r + l)$ denominator factors, where r is the order of the Lie algebra.

3. AN ALTERNATIVE FORMULA

For a given irreducible representation (λ), the exponents of the η in the character $\chi_\lambda(\eta)$ provide us with a complete set of weights Δ_λ. This set can be written as a sum of subsets Δ_λ^k, where k is the number of simple roots which must be subtracted from the highest weight in order to obtain a weight in the subset. The subset Δ_λ^0 contains the highest weight, whose components in our basis are $(\lambda_1,...,\lambda_l)$. The usual procedure for determination of the complete set of weights proceeds iteratively in k.[8] All weights in the $k = 0$ level are known. If all weights in levels 0,...,k are known, then the weights in level $k + 1$ are determined by the following algorithm: Let $\mathbf{w}$ be an arbitrary weight in Δ_λ^k. If its nth component w_n is greater than zero, then by (2.6) it has a Weyl reflection $\mathbf{w} - w_n\boldsymbol{\alpha}_n$. Both of these weights are part of a chain which includes $\mathbf{w} - \boldsymbol{\alpha}_n$ so that this latter is a weight in level $k + 1$. If w_n is less than or equal to zero, then $\mathbf{w} - \boldsymbol{\alpha}_n$ is a weight if its reflection $\mathbf{w} + (1 - w_n)\boldsymbol{\alpha}_n$ is part of an α_n chain containing $\mathbf{w}$.

In order to construct complete sets of weights, we shall use an equivalent procedure which proceeds iteratively on the word length of the Weyl reflections. We shall show that if S is a Weyl reflection obtained by multiplying the minimum number of generators S_i, then the character $\chi_\lambda(\eta)$ is given by

$$\chi_\lambda(\eta) = \sum_S \tilde{S}\colon \prod_{k=1}^{l} \eta_k^{\lambda_k}, \tag{3.1}$$

where the sum is over all elements of the Weyl reflection group and where $\tilde{S}$ is an operator obtained by replacing all S_i in S by

$$\tilde{S}_i = (1 - \rho_i)^{-1}(S_i - I). \tag{3.2}$$

The quantity ρ_i defined in (2.10) has η exponents corresponding to the weight of the ith simple root.

Our method of construction of the complete set of weights is the following: We take the highest weight $\mathbf{M}$ to be the sole member of the set Γ^0. Acting on $\mathbf{M}$ with each of the generators S_i (words of length 1), we obtain a set of reflections and implied chain members which are placed in the set Γ^1. Notice that this set does not include $\mathbf{M}$. Continuing in this fashion, we act on all members of the set Γ^k with the generators and include all new weights in the set Γ^{k+1}. The procedure is continued until no new weights are introduced. In order to prove that Eq. (3.1) is valid, we must prove that the operators $\tilde{S}$ corresponding to elements of word length k of the Weyl group produce all terms $\eta^{\mathbf{w}}$, with $\mathbf{w} \in \Gamma^k$, and no others.

First, we prove that the only possible operators $\tilde{S}$ correspond to elements of the Weyl group. It is tedious but straight forward to prove by construction that $\tilde{S}_i\tilde{S}_i = -\tilde{S}_i$ and that $\tilde{S}_i\tilde{S}_j = \tilde{S}_j\tilde{S}_i$, $\tilde{S}_i\tilde{S}_j\tilde{S}_i = \tilde{S}_j\tilde{S}_i\tilde{S}_j$, $\tilde{S}_i\tilde{S}_j\tilde{S}_i\tilde{S}_j = \tilde{S}_j\tilde{S}_i\tilde{S}_j\tilde{S}_i$, or $\tilde{S}_i\tilde{S}_j\tilde{S}_i\tilde{S}_j\tilde{S}_i\tilde{S}_j = \tilde{S}_j\tilde{S}_i\tilde{S}_j\tilde{S}_i\tilde{S}_j\tilde{S}_i$ when the ith and jth circles of the Dynkin diagram are connected by 0, 1, 2, or 3 lines.

These relations can be used to eliminate any word which is not of minimum length or which is equal to one of the same length. Thus the sum over S in (3.1) is complete.

Suppose that a weight $\mathbf{w}$ has a positive ith component w_i. Then by (2.6) the reflection of this weight under $\tilde{S}_i$ is $\mathbf{w} - w_i\boldsymbol{\alpha}_i$, where α_i is the ith simple root. This implies the existence of a chain of weights $\mathbf{w}, \mathbf{w} - \boldsymbol{\alpha}_i, \ldots, \mathbf{w} - w_i\boldsymbol{\alpha}_i$. The effect of $\tilde{S}_i$ on $\eta^{\mathbf{w}} = \Pi\eta_k^{w_k}$ is

$$\tilde{S}_i\colon\ \eta^{\mathbf{w}} = \sum_{k=1}^{w_i} \eta^{\mathbf{w}}\rho_i^{-k} = \sum_{k=1}^{w_i} \eta^{\mathbf{w}-k\boldsymbol{\alpha}_i}. \tag{3.3}$$

Notice that $\tilde{S}_i$ introduces terms corresponding to the *additional* members of the chain. If $\mathbf{w}$ belongs to Γ^k and is assumed to have been produced by an operator $\tilde{S}$ of length k, then these new weights are produced by the operator $\tilde{S}_i\tilde{S}$ of length $k+1$ and appear in Γ^{k+1} as required. Notice as well that if Γ^k contains both $\mathbf{w}$ and its reflection $S_i\mathbf{w}$, then no additional weights are produced. The only potential problem occurs if $\mathbf{w}$ has a negative ith component and is unaccompanied by its reflection. In that case, the terms introduced by $\tilde{S}_i$ are negative. We shall show that this never happens.

There are only two ways in which a weight with a negative ith component can occur. First, it may occur as a part of a chain generated by $\tilde{S}_i$. Second, it may be produced when another operator, $\tilde{S}_j$, acts on a weight with a negative jth component and which is unaccompanied by its j-reflection. The latter possibility can be neglected if we concentrate on the first occurrence of this potential problem. Suppose that $\mathbf{w}$ is a weight obtained by the operation $\tilde{S}$. Then $\tilde{S}_i\tilde{S}$ gives us the weights $\mathbf{w} - p\boldsymbol{\alpha}_i$, where $p = 1,\ldots,w_i$. Notice that $\mathbf{w} - w_i\boldsymbol{\alpha}_i$ is unaccompanied by its i-reflection. This is no problem since $\tilde{S}_i\tilde{S}_i\tilde{S}$ is not a reduced element. If we now act with $\tilde{S}_j$, where $\tilde{S}_i\tilde{S}_j \neq \tilde{S}_j\tilde{S}_i$ (if $\tilde{S}_i$ and $\tilde{S}_j$ commute, we put $\tilde{S}_j$ into $\tilde{S}$), we obtain $\mathbf{w} - q\boldsymbol{\alpha}_j - p\boldsymbol{\alpha}_i$, where $q = 1,\ldots,$ $(w_j - pA_{ji}) \geqslant 1$ since the component A_{ji} of the Cartan matrix is the jth component of the ith simple root in this basis. Let us suppose that the ith component of one of the above weights, $w_i - qA_{ij} - 2p$, is less than zero. If the i-reflection of this weight, $\mathbf{w} - q\boldsymbol{\alpha}_j - (w_i - p - qA_{ij})\boldsymbol{\alpha}_i$, is one of the weights above, then we have no problem at this stage. Since A_{ij} is a negative integer and since $w_i - p \geqslant 0$, we see that the reflection has the form $\mathbf{w} - q\boldsymbol{\alpha}_j - p'\boldsymbol{\alpha}_i$ with $1 \leqslant p' < p \leqslant w_i$ so that this is in fact one of the weights above. This argument is easily extended if we act with further operators and so we conclude that if some operator $\tilde{S}$ produces a term corresponding to a weight $\mathbf{w}$ with $w_i < 0$, then either a term corresponding to $S_i\mathbf{w}$ is also produced or $\tilde{S}_i\tilde{S}$ is not a reduced word. Therefore, Eq. (3.1) does accomplish the construction described above.

As an example, let us construct the character of the (1,2) representation of SU(3). The elements of the Weyl group are $I, S_1, S_2, S_1S_2, S_2S_1$, and $S_1S_2S_1$, where $S_1\colon \eta_1 = \eta_1^{-1}\eta_2$ and $S_2\colon \eta_2 = \eta_1\eta_2^{-1}$. The simple roots have components $(2, -1)$ and $(-1,2)$ so that $\rho_1 = \eta_1^2\eta_2^{-1}$ and $\rho_2 = \eta_1^{-1}\eta_2^2$. I acting on $\eta_1\eta_2^2$ gives us this highest weight term back again. The other reflections give

$$\tilde{S}_1\colon\ \eta_1\eta_2^2 = (\eta_1^{-1}\eta_2^3 - \eta_1\eta_2^2)/(1 - \eta_1^2\eta_2^{-1}) = \eta_1^{-1}\eta_2^3,$$

$$\tilde{S}_2\colon\ \eta_1\eta_2^2 = (\eta_1^3\eta_2^{-2} - \eta_1\eta_2^2)/(1 - \eta_1^{-1}\eta_2^2) = \eta_1^3\eta_2^{-2} + \eta_1^2,$$

$$\tilde{S}_1\tilde{S}_2\colon\ \eta_1\eta_2^2 = (\eta_1^{-3}\eta_2 + \eta_1^{-2}\eta_2^2 - \eta_1^3\eta_2^{-2} - \eta_1^2)/(1 - \eta_1^2\eta_2^{-1})$$
$$= \eta_1^{-3}\eta_2 + \eta_1^{-1} + \eta_1\eta_2^{-1} + \eta_1^{-2}\eta_2^2 + \eta_2,$$

$$\tilde{S}_2\tilde{S}_1\colon\ \eta_1\eta_2^2 = (\eta_1^2\eta_2^{-3} - \eta_1^{-1}\eta_2^3)/(1 - \eta_1^{-1}\eta_2^2)$$
$$= \eta_1^2\eta_2^{-3} + \eta_1\eta_2^{-1} + \eta_2,$$

$$\tilde{S}_1\tilde{S}_2\tilde{S}_1\colon\ \eta_1\eta_2^2 = (\eta_1^{-2}\eta_2^{-1} + \eta_1^{-1} - \eta_1^2\eta_2^{-3} - \eta_1\eta_2^{-1})/(1 - \eta_1^2\eta_2^{-1})$$
$$= \eta_1^{-2}\eta_2^{-1} + \eta_2^{-2} + \eta_1^{-1}.$$

The sum of these terms provides the character $\chi_{(1,2)}(\eta)$. Notice that the last line can also be obtained from the equivalent word $\tilde{S}_2\tilde{S}_1\tilde{S}_2$.

Equation (3.1) can be turned into an equation for the character generator by replacing $\eta_k^{\lambda_k}$ by $(1 - A_k\eta_k)^{-1}$, so we have

$$X(A,\eta) = \sum_{\tilde{S}}\tilde{S}\colon \prod_{k=1}^{l}(1 - A_k\eta_k)^{-1}. \tag{3.4}$$

It is important to notice, as can be seen from the discussion earlier in this section, that each term in (3.4) is positive. Moreover, the only factors which occur in the denominators of these terms correspond to the outside weights of the fundamental irreducible representations. Therefore, this expression has neither the negative terms nor the spurious poles which occur in (2.8).

There are two relationships which are useful in evaluating the terms of (3.4). First,

$$\tilde{S}_i\colon (FG) = (\tilde{S}_i\colon F)G + (S_i\colon F)(\tilde{S}_i\colon G) \tag{3.5}$$

allows us to act separately with $\tilde{S}_i$ on each factor of an expression. Second, if $\mathbf{w}$ is an outside weight of some fundamental irreducible representation which is labeled by A, then the corresponding denominator factor satisfies

$$\tilde{S}_i\colon (1 - A\eta^{\mathbf{w}})^{-1} = (A\tilde{S}_i\colon\eta^{\mathbf{w}})/(1 - A\eta^{\mathbf{w}})(1 - AS_i\colon\eta^{\mathbf{w}}). \tag{3.6}$$

Thus each term in (3.4) generated by a word S of length h has $h + l$ denominator factors. Since the word of maximum length has $h = \frac{1}{2}(r - l)$, the terms in the final expression for the generating function will have terms with $\frac{1}{2}(r + l)$ denominator factors, as is the case for the Weyl formula.

4. DETERMINATION OF THE WEYL ELEMENTS

In practice, the construction of the elements of the Weyl group by using the relations (2.4) to eliminate equivalent words is a very tedious one. It is far more convenient to construct them by considering their action in weight space.

In order to simplify the procedure further, we shall consider the Lie subgroup, and corresponding Weyl subgroup, obtained by deleting one of the circles from the Dynkin diagram. We shall eliminate the l th circle in our discussion, but the arguments presented are completely general. The set $\{S\}$ of elements of the Weyl group can be factored into a set $\{H\}$ of subgroup elements and a set $\{T\}$ such that

$$\{S\} = \{T\}\{H\}. \tag{4.1}$$

The elements in the set $\{H\}$ are generated by the $l - 1$ gener-

ators $S_1,\ldots,S_{l-1}$. These elements do not affect η_l, the highest weight term in the character of the l th fundamental IR. We can therefore determine the elements of $\{T\}$ in terms of their action on η_l.

A simple graphical algorithm suffices to determine the elements of $\{T\}$. The highest point on the graph is labeled by η_l. Points on each level are determined from those of the level above by applying S_i to the label of each of the higher points iff that label has a positive exponent of η_i. The generated point, labeled by the action of S_i on the higher label, is connected to the higher point by a line labeled by S_i. The action of S_i on any label is obtained by replacing

$$\eta_i \to \eta_i \Pi \eta_j^{-A_{ji}} \tag{4.2}$$

in that label. If a point is connected to those of the level above by more than one line, then all but one of those lines are eliminated. Each point on the completed graph represents one of the outside weights of the corresponding fundamental IR and is connected to the highest point by a unique path. The product of the labels of the lines of this path, with higher labels to the right, represents one of the elements of $\{T\}$. The connection of the highest point with itself represents the element I.

The graph shown in Fig. 1 determines the outside weights and elements of $\{T\}$ for the (0,1,0) representation of

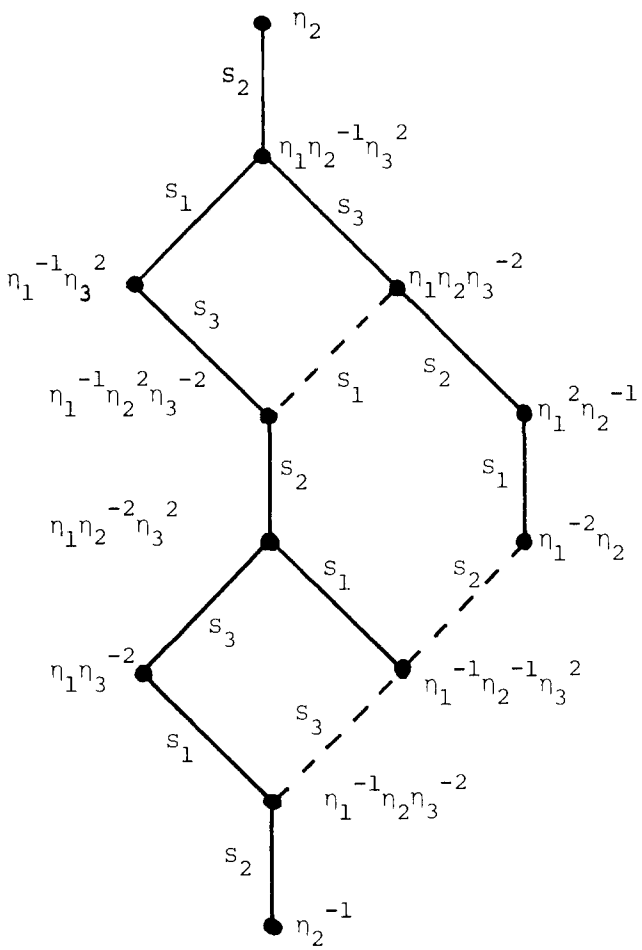


FIG. 1. Outside weights and Weyl generators for the (0,1,0) representation of B_3.

B_3. The dashed lines are to be eliminated from the graph. The resulting elements of $\{T\}$ are: I, S_2, S_1S_2, S_3S_2, $S_3S_1S_2$, $S_2S_3S_2$, $S_2S_3S_1S_2$, $S_1S_2S_3S_2$, $S_1S_2S_3S_1S_2$, $S_3S_2S_3S_1S_2$, $S_1S_3S_2S_3S_1S_2$, and $S_2S_1S_3S_2S_3S_1S_2$.

The elements of $\{H\}$ can be factored further by eliminating other circles of the Dynkin diagram until, ultimately, only a single circle representing A_1 remains. In the example above, the diagram ○—○⇒● for B_3 was reduced to ○ ○ of $A_1\times A_1$. The elements of the set $\{H\}$ are $(I+S_3)(I+S_1)$. The elements of $\{S\}$ for B_3 are the 48 terms in the product $\{T\}\{H\}$. The set of elements $\{\tilde{S}\}$ appearing in (3.1) and (3.4) are just those of $\{S\}$ with all S_i replaced by $\tilde{S}_i$.

The set $\{S\}$ contains a single element S_M of maximum length which is the product of the elements of maximum length in its factor sets. We shall now show that the sum of the $\tilde{S}$ can be replaced by $\bar{S}_M$, the operator obtained by replacing each of the S_i in S_M by $(I+\tilde{S}_i)$. Consider the product of the h rightmost factors in $\bar{S}_M$ and assume that all words generated are inequivalent. We now multiply this set of words by the next factor, $(I+\tilde{S}_i)$. Multiplication by I clearly reproduces the previous set of inequivalent words. Let us suppose that S is a word in this set and that $S_iS=S'$ is equivalent to a longer word in the set. Then since $\tilde{S}_i\tilde{S}_i=-\tilde{S}_i$, we have $\tilde{S}_i(\tilde{S}+\tilde{S}')=0$ so that no new words are introduced which are equivalent to words in the previous set. The process terminates when no new word can be introduced by multiplication by any $(I+\tilde{S}_i)$. This is the case for S_M since its factors can be rearranged to place any given S_i on the left. Multiplication of $(I+\tilde{S}_i)$ by itself gives $(I+\tilde{S}_i)$ back again so that $(I+\tilde{S}_i)\bar{S}_M=\bar{S}_M$. The set of words generated by $\bar{S}_M$ is complete and inequivalent.

Equation (3.4) for the character generator can therefore be written as

$$X(A,\eta)=\bar{S}_M\colon \prod_{k=1}^{l}(1-A_k\eta_k)^{-1}. \tag{4.3}$$

The character generator for B_3, for example, becomes

$$\begin{aligned} &X_{B_3}(A,\eta)\\ &=(I+\tilde{S}_2)(I+\tilde{S}_1)(I+\tilde{S}_3)(I+\tilde{S}_2)(I+\tilde{S}_3)(I+\tilde{S}_1)(I+\tilde{S}_2)\\ &\quad\times(I+\tilde{S}_3)(I+\tilde{S}_1)[(1-A_1\eta_1)(1-A_2\eta_2)(1-A_3\eta_3)]^{-1}. \end{aligned} \tag{4.4}$$

This form of the character generator, while equivalent to (3.4), is more practical. The reason is that the sum in (3.4) will yield terms with fewer than $\frac{1}{2}(r+l)$ denominator factors which must be rearranged to produce the final form. In Eq. (4.3) the result of multiplication by each $(I+\tilde{S}_i)$ can be arranged into terms with one more denominator factor than the previous one and with no negative numerator coefficients so that no rearrangement is necessary to produce the final form.

Since $S_M=T_MH_M$, where T_M and H_M, are the longest words in $\{T\}$ and $\{H\}$, respectively, we should be able to write the character generator $X(A,\eta)$ in terms of the character generator $X_h(A,\eta)$ for the subgroup obtained by deleting the l th circle of the Dynkin diagram. This latter depends upon A_k and η_k for $k=1,\ldots,l-1$ so we must introduce the implied η_l dependence. To do this, we introduce the $l-1$

quantities

$$\mu_i = \sum_{k=1}^{l-1} A^{\,g}_{lk} A^{\,h^{-1}}_{ki} \quad (i = 1,\ldots,l-1) \tag{4.5}$$

and write

$$X(A,\eta) = \bar{T}_M\colon\ (1 - A_l\eta_l)^{-1} X_h(A_{\bar{\eta}^l}{}^{\mu}, \eta\eta_l^{\mu}), \tag{4.6}$$

where in (4.5) $A^{\,g}$ and $A^{\,h}$ are the group and subgroup Cartan matrices, respectively. These relations are easily generalized to subgroups obtained by eliminating other circles. Equation (4.6) allows us to work with chains of subgroups, so that we do not need to work out each character generator from scratch.

5. CONSTRUCTION OF CHARACTER GENERATORS

Equations (3.4), (4.3), and (4.6) provide us with three expressions for the character generator. The first of these is particularly useful if the Weyl group is small. Each term gives a positive-definite result with the only drawback being that the terms do not have the same number of denominator factors. Some rearrangement is necessary to put the result in the simplest form. The second the third formulae, which are equivalent, give us terms with the same number of denominator factors at each step with the third exhibiting the possibility of construction through a subgroup chain. In practice, however, the second equation, (4.3), seems to offer the easiest construction since the algebra of the $\tilde{S}_i$ can be used to simplify the work at each stage.

The basic idea of this last method is to use (3.5) and the similar relation

$$(I + \tilde{S}_i)\colon\ (FG) = ((I + \tilde{S}_i){:}F)G + (S_i{:}F)(\tilde{S}_i{:}G) \tag{5.1}$$

to move the $\tilde{S}_i$ operators through each of the $(1 - A_k\eta_k)^{-1}$ in turn, collecting and simplifying after each denominator has been passed before moving to the next. For example, the character generator for B_2 [SO(5)] can be written

$$X(A,\eta) = (I + \tilde{S}_2)(I + \tilde{S}_1)(I + \tilde{S}_2) \times (I + \tilde{S}_1)(1 - A_2\eta_2)^{-1}(1 - A_1\eta_1)^{-1}. \tag{5.2}$$

We can immediately move $(I + \tilde{S}_1)$ through the factor $(1 - A_2\eta_2)^{-1}$. Using Eq. (5.1) to move $(I + \tilde{S}_2)$ through now gives

$$(I + \tilde{S}_2)(I + \tilde{S}_1)[(I + \tilde{S}_2)(1 - A_2\eta_2)^{-1}](I + \tilde{S}_1)(1 - A_1\eta_1)^{-1} + (I + \tilde{S}_2)(I + \tilde{S}_1)[S_2(1 - A_2\eta_2)^{-1}]\tilde{S}_2(I + \tilde{S}_1)(1 - A_1\eta_1)^{-1}. \tag{5.3}$$

Notice that $\tilde{S}_2(1 - A_1\eta_1)^{-1} = 0$ so that the factor $\tilde{S}_2(1 + \tilde{S}_1)$ in the second term can be replaced by $\tilde{S}_2\tilde{S}_1$. We now move $(I + \tilde{S}_1)$ through to give

$$\begin{aligned}(I + \tilde{S}_2)[(I + \tilde{S}_1)(I + \tilde{S}_2)(1 - A_2\eta_2)^{-1}](I + \tilde{S}_1)(1 - A_1\eta_1)^{-1} \\ + (I + \tilde{S}_2)[S_1(I + \tilde{S}_2)(1 - A_2\eta_2)^{-1}]\tilde{S}_1(I + \tilde{S}_1)(1 - A_1\eta_1)^{-1} \\ + (I + \tilde{S}_2)[(I + \tilde{S}_1)S_2(1 - A_2\eta_2)^{-1}]\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1} \\ + (I + \tilde{S}_2)[S_1S_2(1 - A_2\eta_2)^{-1}]\tilde{S}_1\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1}.\end{aligned} \tag{5.4}$$

The second term in (5.4) vanishes since $\tilde{S}_i(I + \tilde{S}_i) = 0$. Moving $(I + \tilde{S}_2)$ through gives

$$\begin{aligned}&[(I + \tilde{S}_2)(I + \tilde{S}_1)(I + \tilde{S}_2)(1 - A_2\eta_2)^{-1}](I + \tilde{S}_1)(1 - A_1\eta_1)^{-1} \\ &+ [S_2(I + \tilde{S}_1)(I + \tilde{S}_2)(1 - A_2\eta_2)^{-1}]\tilde{S}_2(I + \tilde{S}_1)(1 - A_1\eta_1)^{-1} \\ &\quad + [(I + \tilde{S}_2)(I + \tilde{S}_1)(1 - A_2\eta_2)^{-1}]\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1} \\ &\quad + [S_2(I + \tilde{S}_1)S_2(1 - A_2\eta_2)^{-1}]\tilde{S}_2\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1} \\ &\quad + [(I + \tilde{S}_2)S_1S_2(1 - A_2\eta_2)^{-1}]\tilde{S}_1\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1} \\ &\quad + [S_2S_1S_2(1 - A_2\eta_2)^{-1}]\tilde{S}_2\tilde{S}_1\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1}.\end{aligned} \tag{5.5}$$

The last term vanishes since $\tilde{S}_2\tilde{S}_1\tilde{S}_2\tilde{S}_1 = \tilde{S}_1\tilde{S}_2\tilde{S}_1\tilde{S}_2$ and $\tilde{S}_2(1 - A_1\eta_1)^{-1} = 0$. The second, third, and fourth terms have the same factors following the bracket since $\tilde{S}_2(I + \tilde{S}_1)$ can be replaced with $\tilde{S}_2\tilde{S}_1$ and since $\tilde{S}_2\tilde{S}_2 = -\tilde{S}_2$. Evaluating the quantities in brackets we have the three terms

$$\begin{aligned}X(A,\eta) = &[(1 - A_2\eta_2^{-1})(1 - A_2\eta_1^{-1}\eta_2)]^{-1} \\ &\times\tilde{S}_1\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1} \\ &+ [(1 - A_2\eta_2^{-1})(1 - A_2\eta_1^{-1}\eta_2) \\ &\times(1 - A_2\eta_1\eta_2^{-1})]^{-1} \\ &\times\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1} \\ &+ [(1 - A_2\eta_2^{-1})(1 - A_2\eta_1^{-1}\eta_2) \\ &\times(1 - A_2\eta_1\eta_2^{-1})(1 - A_2\eta_2)]^{-1} \\ &\times(I + \tilde{S}_1)(1 - A_1\eta_1)^{-1}.\end{aligned} \tag{5.6}$$

The operators $\tilde{S}_i$ now act only on the A_1 factor. Since they form reduced words of the Weyl group, the results will be positive definite.

In evaluating the brackets in (5.6), we have used (3.5), (3.6), and the relations $S_1{:}\eta_1 = \eta_1^{-1}\eta_2^2$, $S_2{:}\eta_2 = \eta_1\eta_2^{-1}$, $\rho_1 = \eta_1^2\eta_2^{-2}$, and $\rho_2 = \eta_1^{-1}\eta_2^2$. In order to see how this procedure works, we shall evaluate $\tilde{S}_2\tilde{S}_1(1 - A_1\eta_1)^{-1}$ starting with the partial result

$$\tilde{S}_1(1 - A_1\eta_1)^{-1} = A_1\eta_1^{-1}\eta_2^2[(1 - A_1\eta_1^{-1}\eta_2^2)(1 - A_1\eta_1)]^{-1}. \tag{5.7}$$

Let us now operate with $\tilde{S}_2$. Notice first that any denominator factor not containing η_i, or which is accompanied by its i-reflection, can be treated as a constant as far as $\tilde{S}_i$ is concerned. We can therefore ignore $(1 - A_1\eta_1)^{-1}$ for the moment. With (3.5), the action of $\tilde{S}_2$ on the remaining factors becomes

$$\begin{aligned}&[\tilde{S}_2(1 - A_1\eta_1^{-1}\eta_2^2)^{-1}](A_1\eta_1^{-1}\eta_2^2) \\ &\quad + [S_2(1 - A_1\eta_1^{-1}\eta_2^2)^{-1}]\tilde{S}_2(A_1\eta_1^{-1}\eta_2^2).\end{aligned} \tag{5.8}$$

The final factor in the second term is, with (3.2) and the definitions of ρ_2 and of S_2: η_2 above,

$$\begin{aligned}\tilde{S}_2(A_1\eta_1^{-1}\eta_2^2) &= [(A_1\eta_1\eta_2^{-2}) - (A_1\eta_1^{-1}\eta_2^2)]/(1 - \eta_1^{-1}\eta_2^2) \\ &= A_1\eta_1\eta_2^{-2} + A_1.\end{aligned} \tag{5.9}$$

According to (3.6),

$$\begin{aligned}&\tilde{S}_2(1 - A_1\eta_1^{-1}\eta_2^2)^{-1} \\ &\quad = [\tilde{S}_2(A_1\eta_1^{-1}\eta_2^2)]/(1 - A_1\eta_1^{-1}\eta_2^2)(1 - S_2A_1\eta_1^{-1}\eta_2^2) \\ &\quad = (A_1\eta_1\eta_2^{-1} + A_1)/(1 - A_1\eta_1^{-1}\eta_2^2)(1 - A_1\eta_1\eta_2^{-2})\end{aligned} \tag{5.10}$$

so that (5.8) becomes

$$\begin{aligned}&A_1\eta_1^{-1}\eta_2^2(A_1 + A_1\eta_1\eta_2^{-2})/(1 - A_1\eta_1^{-1}\eta_2^2)(1 - A_1\eta_1\eta_2^{-2}) \\ &\quad + (A_1 + A_1\eta_1\eta_2^{-2})/(1 - A_1\eta_1\eta_2^2) \\ &\quad = (A_1 + A_1\eta_1\eta_2^{-2})/(1 - A_1\eta_1^{-1}\eta_2^2)(1 - A_1\eta_1\eta_2^{-2}).\end{aligned} \tag{5.11}$$

The result for $\tilde{S}_2\tilde{S}_1(1-A_1\eta_1)^{-1}$ is (5.11) with the denominator factor $(1-A_1\eta_1)$ reinstated. Evaluating the other factors in (5.6) similarly, we finally obtain the B_2 character generator

$$\begin{aligned}X(A,\eta) = {} & [1/(1-A_2\eta_2)(1-A_2\eta_1\eta_2^{-1}) \\ & + (A_1 + A_1\eta_1\eta_2^{-2})/(1-A_2\eta_1\eta_2^{-1}) \\ & \times(1-A_1\eta_1\eta_2^{-2}) \\ & + A_1\eta_1^{-1}(1+A_1)/(1-A_1\eta_1\eta_2^{-2})(1-A_1\eta_1^{-1})] \\ & \times[(1-A_2\eta_2^{-1})(1-A_2\eta_1^{-1}\eta_2)(1-A_1\eta_1) \\ & \times(1-A_1\eta_1^{-1}\eta_2^2)]^{-1} \end{aligned} \tag{5.12}$$

This result agrees, to within a change of basis, with the generator obtained by other methods.[1]

The principal advantage of this method is that it allows us to do the construction in stages. For example, the character generator for B_3 can be written

$$\begin{aligned}X(A,\eta) = {} & (I+\tilde{S}_3)(I+\tilde{S}_2)(I+\tilde{S}_1)(I+\tilde{S}_3)(I+\tilde{S}_2)(I+\tilde{S}_3) \\ & \times(1-A_3\eta_3)^{-1}(I+\tilde{S}_2)(I+\tilde{S}_1) \\ & \times(1-A_1\eta_1)^{-1}(I+\tilde{S}_2)(1-A_2\eta_2)^{-1}, \end{aligned} \tag{5.13}$$

where the $\tilde{S}_i$ satisfy $\tilde{S}_1\tilde{S}_3 = \tilde{S}_3\tilde{S}_1$, $\tilde{S}_1\tilde{S}_2\tilde{S}_1 = \tilde{S}_2\tilde{S}_1\tilde{S}_2$, and $\tilde{S}_2\tilde{S}_3\tilde{S}_2\tilde{S}_3 = \tilde{S}_3\tilde{S}_2\tilde{S}_3\tilde{S}_2$. After moving the operators $(I+\tilde{S}_i)$ through $(1-A_3\eta_3)^{-1}$ we are left with

$$\begin{aligned}X(A,\eta) = {} & \tilde{P}(A_3,\eta)(I+\tilde{S}_2)(I+\tilde{S}_1)(1-A_1\eta_1)^{-1} \\ & (I+\tilde{S}_2)(1-A_2\eta_2)^{-1}, \end{aligned} \tag{5.14}$$

where

$$\begin{aligned}\tilde{P}(A_3,\eta) = {} & P_1\tilde{S}_2\tilde{S}_1\tilde{S}_3\tilde{S}_2\tilde{S}_3 + P_2\tilde{S}_1\tilde{S}_3\tilde{S}_2\tilde{S}_3 + P_3\tilde{S}_3\tilde{S}_2\tilde{S}_3 \\ & + P_4\tilde{S}_1\tilde{S}_2\tilde{S}_3 + P_5\tilde{S}_2\tilde{S}_3 + P_6\tilde{S}_3 + P_7 \end{aligned} \tag{5.15}$$

and

$$\begin{aligned} & P_1 = 1/(1-A_3\eta_3^{-1})(1-A_3\eta_2^{-1}\eta_3), \\ & P_2 = P_1/(1-A_3\eta_1^{-1}\eta_2\eta_3^{-1}), \\ & P_3 = P_2/(1-A_3\eta_1\eta_3^{-1}), \quad P_4 = P_2/(1-A_3\eta_1^{-1}\eta_3), \\ & P_5 = (P_3 + A_3\eta_1^{-1}\eta_3P_4)/(1-A_3\eta_1\eta_2^{-1}\eta_3), \\ & P_6 = P_5/(1-A_3\eta_2\eta_3^{-1}), \quad P_7 = P_6/(1-A_3\eta_3). \end{aligned} \tag{5.16}$$

We therefore have the complete A_3 content. In the next stage we move the operators through $(1-A_1\eta_1)^{-1}$ to give

$$X(A,\eta) = \tilde{Q}(A_1,A_3,\eta)(I+\tilde{S}_2)(1-A_2\eta_2)^{-1}, \tag{5.17}$$

where

$$\begin{aligned}\tilde{Q}(A_1,A_3,\eta) = {} & Q_1\tilde{S}_2\tilde{S}_3\tilde{S}_1\tilde{S}_2\tilde{S}_3\tilde{S}_1 + Q_2\tilde{S}_3\tilde{S}_1\tilde{S}_2\tilde{S}_3\tilde{S}_1 + Q_3\tilde{S}_3\tilde{S}_2\tilde{S}_3\tilde{S}_1 + Q_4\tilde{S}_1\tilde{S}_2\tilde{S}_3\tilde{S}_1 + Q_5\tilde{S}_1\tilde{S}_2\tilde{S}_3 \\ & + Q_6\tilde{S}_2\tilde{S}_3\tilde{S}_1 + Q_7\tilde{S}_2\tilde{S}_3 + Q_8\tilde{S}_3\tilde{S}_1 + Q_9\tilde{S}_3 + Q_{10}\tilde{S}_1 + Q_{11} \end{aligned} \tag{5.18}$$

and

$$\begin{aligned} Q_1 = {} & P_1/(1-A_1\eta_1^{-1})(1-A_1\eta_1\eta_2^{-1}), \\ Q_2 = {} & [P_2 + A_1\eta_1\eta_2^{-1}P_1/(1-A_1\eta_1\eta_2^{-1})]/(1-A_1\eta_1^{-1})(1-A_1\eta_2\eta_3^{-2}), \\ Q_3 = {} & [P_3 + A_1\eta_1^{-1}P_2/(1-A_1\eta_1^{-1})]/(1-A_1\eta_1\eta_2^{-1})(1-A_1\eta_2\eta_3^{-2}), \\ Q_4 = {} & [P_4 + (A_1(1+\eta_2\eta_3^{-2})P_2 + A_1\eta_1\eta_2^{-1}(1+A_1)P_1/(1-A_1\eta_1\eta_2^{-1}))/(1-A_1\eta_2\eta_3^{-2})] \\ & \times[(1-A_1\eta_1^{-1})(1-A_1\eta_2^{-1}\eta_3^2)]^{-1}, \\ Q_5 = {} & Q_4/(1-A_1\eta_1^{-1}\eta_2), \\ Q_6 = {} & [P_5 + A_1\eta_1^{-1}P_4/(1-A_1\eta_1^{-1}) + A_1(1+\eta_2\eta_3^{-2})(P_3 + A_1\eta_1^{-1}P_2/(1-A_1\eta_1^{-1})) \\ & \times(1-A_1\eta_2\eta_3^{-2})^{-1}]/(1-A_1\eta_1\eta_2^{-1})(1-A_1\eta_2^{-1}\eta_3^2), \\ Q_7 = {} & (Q_6 + A_1\eta_1^{-1}\eta_2Q_5)/(1-A_1\eta_1), \\ Q_8 = {} & [P_6/(1-A_1\eta_2\eta_3^{-2}) + (P_5[A_1\eta_1\eta_2^{-1}/(1-A_1\eta_2\eta_3^{-2}) \\ & + A_1\eta_2^{-1}\eta_3^2/(1-A_1\eta_2^{-1}\eta_3^2)] + A_1\eta_1^{-1}P_4[1/(1-A_1\eta_2\eta_3^{-2}) \\ & + A_1\eta_2^{-1}\eta_3^2)/(1-A_1\eta_2^{-1}\eta_3^2)]/(1-A_1\eta_1^{-1})/(1-A_1\eta_1\eta_2^{-1}) \\ & + A_1(1+A_1)Q_3/(1-A_1\eta_2^{-1}\eta_3^2)]/(1-A_1\eta_1^{-1}\eta_2), \\ Q_9 = {} & Q_8/(1-A_1\eta_1), \\ Q_{10} = {} & [P_7 + A_1(1+\eta_2\eta_3^{-2})P_6/(1-A_1\eta_2\eta_3^{-2}) \\ & + A_1\eta_1\eta_2^{-1}(1+A_1)P_5/(1-A_1\eta_1\eta_2^{-1})(1-A_1\eta_2\eta_3^{-2}) \\ & + A_1\eta_1^{-1}(1+A_1)P_4/(1-A_1\eta_1^{-1})(1-A_1\eta_1\eta_2^{-1})(1-A_1\eta_2\eta_3^{-2})] \\ & \times[(1-A_1\eta_2^{-1}\eta_3^2)(1-A_1\eta_1^{-1}\eta_2)]^{-1}, \\ Q_{11} = {} & Q_{10}/(1-A_1\eta_1). \end{aligned} \tag{5.19}$$

We now act with our operators on the remaining factor $(1-A_2\eta_2)^{-1}$ to obtain the final result for the B_3 character generator:

$$X(A,\eta) = \sum_{k=1}^{11} \frac{Q_kR_k}{(1-A_2\eta_1\eta_2^{-1}\eta_3^2)(1-A_2\eta_2)} \tag{5.20}$$

where

$$R_{11}=1,\quad R_{10}=A_2\eta_1^{-1}\eta_3^2/(1-A_2\eta_1^{-1}\eta_3^2),\quad R_9=A_2\eta_1\eta_2\eta_3^{-2}(1+\eta_2^{-1}\eta_3^2)/(1-A_2\eta_1\eta_2\eta_3^{-2}),$$

$$R_8=A_2\eta_1^{-1}\eta_2^2\eta_3^{-2}(1+\eta_2^{-1}\eta_3^2)[1/(1-A_2\eta_1\eta_2\eta_3^{-2})+A_2\eta_1^{-1}\eta_3^2/(1-A_2\eta_1^{-1}\eta_3^2)]/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2}),$$

$$R_7=A_2\eta_1^2\eta_2^{-1}(1+A_2\eta_1)/(1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2}),$$

$$\begin{aligned}R_6=&[A_2(1+A_2)(1+A_2\eta_1)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})\\&+A_2\eta_1\eta_2^{-2}\eta_3^2(1+A_2\eta_1)/(1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})\\&+(A_2\eta_2^{-1}\eta_3^2(1+A_2)+A_2^2\eta_2^{-2}\eta_3^4(1+\eta_1^{-1}\eta_2^2\eta_3^{-2}))/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)]/(1-A_2\eta_1\eta_2^{-2}\eta_3^2),\end{aligned}$$

$$\begin{aligned}R_5=&[A_2(1+A_2)(1+A_2\eta_1)/(1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})+A_2\eta_1^{-2}\eta_2(1+A_2\eta_1)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})\\&+A_2^2\eta_1^{-3}\eta_2^2(1+\eta_2^{-1}\eta_3^2)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)]/(1-A_2\eta_1^{-2}\eta_2),\end{aligned}$$

$$\begin{aligned}R_4=&[A_2(1+A_2)(1+A_2\eta_1)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2)\\&+A_2\eta_1\eta_2^{-2}\eta_3^2(1+A_2\eta_1)/(1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2)\\&+A_2\eta_2^{-1}\eta_3^2(1+A_2+A_2\eta_2^{-1}\eta_3^3(1+\eta_1^{-1}\eta_2^2\eta_3^{-2}))/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1\eta_2^{-2}\eta_3^2)\\&+(1+A_2)(1+A_2\eta_1)/(1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-2}\eta_2)\\&+A_2\eta_1^{-1}\eta_2^2\eta_3^{-2}(1+A_2\eta_1)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-2}\eta_2)\\&+A_2\eta_1^{-1}\eta_2(1+\eta_2^{-1}\eta_3^2)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2)]\\&\times A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2/(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2),\end{aligned}$$

$$\begin{aligned}R_3=&[A_2\eta_1\eta_3^{-2}(1+A_2)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})\\&+A_2^2\eta_1^3\eta_2^{-1}\eta_3^{-2}/(1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})+A_2^2(1+A_2)/(1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)]\\&\times(1+\eta_1^{-1}\eta_2)(1+\eta_2^{-1}\eta_3^2)/(1-A_2\eta_1\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2),\end{aligned}$$

$$\begin{aligned}R_2=&[A_2\eta_1\eta_3^{-2}(1+A_2)(1+\eta_1^{-1}\eta_2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^3))\\&+A_2^2\eta_1^3\eta_2^{-1}\eta_3^{-2}(1+\eta_1^{-1}\eta_2)/((1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2^2(1+A_2)(1+\eta_1^{-1}\eta_2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+(1+A_2)(1+A_2\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2\eta_1^2\eta_2^{-1}(1+A_2\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2\eta_1^{-1}\eta_3^2(1+A_2)(1+A_2\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2(1+\eta_1)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2))\\&+A_2\eta_1^{-2}\eta_2(1+\eta_1)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2))\\&+A_2^2(1+\eta_1)(1+\eta_1^2\eta_2^{-1})/((1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2))]\\&\times A_2\eta_1^{-1}\eta_2\eta_3^{-2}(1+\eta_2^{-1}\eta_3^2)/(1-A_2\eta_1^{-1}\eta_2\eta_3^{-2}),\end{aligned}$$

$$\begin{aligned}R_1=&[A_2\eta_1\eta_3^{-2}(1+A_2)(1+\eta_1^{-1}\eta_2)(1+\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1\eta_3^{-2}(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2^2\eta_1^3\eta_2^{-1}\eta_3^{-2}(1+\eta_1^{-1}\eta_2)(1+\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2^2(1+A_2)(1+\eta_1^{-1}\eta_2)(1+\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1\eta_3^{-2})(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+(1+A_2)(1+A_2\eta_1^{-1})(1+A_2^2\eta_1)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2\eta_1^2\eta_2^{-1}(1+A_2\eta_1^{-1})(1+A_2\eta_2^{-1}\eta_3^2)/((1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1\eta_2^{-2}\eta_3^2))\\&+A_2\eta_2^{-1}\eta_3^2(1+A_2)(1+A_2\eta_1^{-1})(1+\eta_1^{-1}\eta_2)/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)\\&\times(1-A_2\eta_1\eta_2^{-2}\eta_3^2))+A_2^2\eta_1^{-1}\eta_2^2\eta_3^{-2}(1+\eta_1)(1+A_2\eta_1^{-1})/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1\eta_2\eta_3^{-2})\\&\times(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2))+A_2\eta_1^{-2}\eta_2(1+\eta_1)(1+A_2\eta_1^{-1})/((1-A_2\eta_1^{-1}\eta_2^2\eta_3^{-2})(1-A_2\eta_1^{-1}\eta_3^2)\\&\times(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2))+A_2(1+A_2)(1+A_2\eta_1^{-1})(1+\eta_1)/((1-A_2\eta_1^2\eta_2^{-1})(1-A_2\eta_1\eta_2\eta_3^{-2})\\&\times(1-A_2\eta_1^{-1}\eta_2^{-1}\eta_3^2)(1-A_2\eta_1^{-2}\eta_2))]A_2\eta_2^{-1}/(1-A_2\eta_2^{-1})(1-A_2\eta_1^{-1}\eta_2\eta_3^{-2}).\end{aligned}\tag{5.21}$$

This method not only simplifies the construction, it also expresses the character generator in a relatively compact form. If Eq. (5.20) were written out, the result would have 127 terms, each with 12 denominator factors and a polynomial numerator.

6. DISCUSSION

Equation (4.3) provides a powerful method for the construction of character generators. However, its usefulness may be questioned since for higher rank groups these generators may simply prove too long to write down. This is evident from a comparison of the B_2 and B_3 character generators which were constructed in the last section. Nevertheless, this new form may prove useful as a starting point for the construction of other generating functions. There is some precedent for this hope: The construction of the generating function for the G_2 Clebsch–Gordan series[9] used as its starting point the Weyl form of the character generator discussed in

Sec. 2 rather than the positive definite form. In this way, the explicit Weyl symmetry could be used to simplify the problem.

One obvious application of our character formula is the testing of other generating functions. Such generating functions enumerate the irreducible representations of a group which are contained in some infinite set of reducible representations. If $G(A,B)$ is such a generating function, with the exponents of A labeling the reducible representation and the exponents of B labeling the irreducible representations, and if $X(A,\eta)$ is a generating function for the characters of the reducible representations, then

$$X(A,\eta) = \bar{S}_M : G(A,\eta). \tag{6.1}$$

It is particularly straightforward to set up a recursive computer program to calculate the right-hand side of this expression numerically, so we can easily test the validity of the expression for $G(A,B)$. This procedure uses the relationship

$$(I + \tilde{S}_i)F(\eta_1,...,\eta_i,...,\eta_l) = (1-\rho_i)^{-1}[F(\eta_1,...,\eta_i/\rho_i,...,\eta_l) - \rho_i F(\eta_1,...,\eta_i,...,\eta_l)]. \tag{6.2}$$

It has been used to test the validity of Eqs. (5.21) and also to verify a generating function for $F_4 \supset B_4$ branching rules.[10]

[1] J. Patera and R. T. Sharp, *Lecture Notes in Physics* (Springer-Verlag, New York, 1979), Vol. 94.
[2] R. P. Stanley, "The character generator of SU(n)," J. Math. Phys. **21**, 2321 (1980).
[3] R. C. King, "The character generator of Sp($2k$)," C. R. Math. Rep. Acad. Sci. Can. **3**, 149 (1981).
[4] K. Baclawski, "Character-generators for unitary and symplectic groups," preprint, Haverford College, 1981.
[5] R. C. King and N. G. I. El-Sharkaway, "The character generators of SO($2k$) and SO($2k+1$)," C. R. Math. Rep. Acad. Sci. Can. **4**, 299 (1982).
[6] H. Weyl, Math. Z. **23** (1925), reprinted in H. Weyl, *Selecta* (Birkhauser-Verlag, Basel, 1956).
[7] C. T. Benson and L. C. Grove, *Finite Reflection Groups* (Bogden and Quigley, New York, 1971).
[8] B. G. Wybourne, *Classical Groups for Physicists* (Wiley, New York, 1974).
[9] R. Gaskell and R. T. Sharp, "G_2 Van Der Waerden invariant," J. Math. Phys. **23**, 2016 (1982).
[10] R. Gaskell and R. T. Sharp, in preparation.

Character generators for elements of finite order in simple Lie groups A_1, A_2, A_3, B_2, and G_2[a)]

R. V. Moody
Department of Mathematics, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

J. Patera[b)]
Centre de recherches de mathématiques appliquées, Université de Montréal, Montréal, Québec, Canada

R. T. Sharp
Department of Physics, McGill University, Montreal, Quebec, Canada

(Received 15 October 1982; accepted for publication 18 February 1983)

Elements of finite order in the compact simple Lie groups SU(2), SU(3), SU(4), Sp(4) or O(5), and G(2) are considered. We provide the characters of the elements on irreducible representations of the Lie groups by assigning appropriate numerical values to the variables on which the characters of representations of the Lie group depend. In this way we specialize generating functions for the characters of the representations of the Lie groups to the generating functions for characters of the elements of finite order. Particular attention is paid to rational elements, all of whose characters are integers; they are listed and the generating functions for their characters are obtained in a simplified form from which the characters can be read. Gaussian elements are also studied in detail. Their characters are complex valued with integer real and imaginary parts.

PACS numbers: 02.20.Qs

1. INTRODUCTION

Elements of finite order (EFO) in compact simple Lie groups could be a very practical tool in applications of representations of Lie groups provided the characters of these elements in irreducible representations of the Lie groups were readily available. With only a few exceptions[1-4] it has not been the case until now. In parallel with the development of a general computational procedure for the characters of EFO in Lie groups[5] we describe here an independent approach to the problem exploiting known generating functions for characters of representations of the Lie group. In this article we consider in detail EFO in the five simple Lie groups of types A_1, A_2, A_3, B_2, and G_2 or, respectively, SU(2), SU(3), SU(4), Sp(4) or O(5), and G(2).

The theory of EFO in Lie groups is in its early stage. There is no account of the "state of the art" in the literature. During our work on the general procedure for machine computation of the characters of EFO [5] we found it necessary to fill many gaps of the theory. The present article is self-contained in what concerns its main objective: providing characters for an arbitrary given EFO in one of the five Lie groups by means of the corresponding generating function. Since characters are invariants under the action of the Lie group, we are interested only in conjugacy classes of EFO rather than individual elements.

The generating functions for characters of representations of simple Lie groups, or simply character generators, were introduced in Ref. 6 where also the first examples were calculated (for the groups A_1, A_2, B_2). Subsequently more efficient combinatorial methods were invented[7-11] which give, in principle, the character generator for any compact Lie group. A character generator is a rational function of $2l$ variables, where l is the rank of the Lie group. The first l variables are those of the characters. The remaining ones are auxiliary variables whose powers, in the power expansion of the character generator, indicate irreducible representations of the Lie group. The coefficient of each power is precisely the character of the representation specified by the power. The main idea of our approach is to fix the character variables in such a way that the character of an irreducible representation of the Lie group is specialized to the character of a given EFO. The character generator for representations of the Lie group is thus transformed into the character generator of the chosen EFO.

Perhaps the most interesting EFO are the so-called rational elements whose character values are always rational (hence integers). Here we have determined all the specializations of all the rational EFO for the groups A_1, A_2, A_3, B_2, and G_2.

Although this is apparently simply a process of passing from the general to the particular, a glance at any one of the character generators, say A_3, (2.3), and the corresponding table of specializations (Table IV) shows that there is considerable simplification in each case. Only in the specialized form is it possible to grasp the entire content of the set of all character values of a given EFO (take for example [1212] in A_3). As a result we are able to tabulate the complete (finite) set of all character values taken by the regular rational elements of our five groups.

We found that prior to specialization it is very convenient to rewrite the character generating functions in such a way that all the coefficients appearing are characters (rather than simply Weyl invariant expressions). The resulting forms are cleaner than the standard ones appearing in the literature.

In Sec. 2 we recall the character generators for the five simple Lie groups in a suitable form. In Sec. 3 we specialize

[a)] Work supported in part by the Natural Science and Engineering Research Council of Canada and by the Ministère de l'Education du Québec.
[b)] Present address: Caltech, Pasadena, CA 91125.

 0022-2488/83/102387-10$02.50

them to a general EFO in each group. In Sec. 4 we describe in detail the character generators for the elements whose characters take only integer values on any representation of the Lie group. Gaussian EFO are considered in Sec. 5. The final section contains some comments, remarks, and examples.

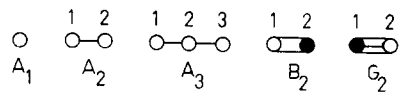


FIG. 1. Coxeter–Dynkin diagrams and numbering of simple roots.

2. CHARACTER GENERATORS

Here we rewrite the five character generators in a suitable form for our purposes.

An irreducible representation of G is specified by its highest weight, using the standard conventions and numbering of simple roots of G, as shown in Fig. 1. The character of the representation $(\lambda_1,\ldots,\lambda_l)$ is denoted by $\langle\lambda_1,\ldots,\lambda_l\rangle$. It is a function of l variables which we do not need to specify here. For examples, see Refs. 6–12.

The A_1-character generator, Eq. (1) of Ref. 6, can be rewritten as the rational function

$$\frac{1}{1-\langle 1\rangle A+A^2}=1+\langle 1\rangle A+(\langle 1\rangle^2-1)A^2+\cdots, \tag{2.1}$$

where $\langle 1\rangle$ is the character of the two-dimensional representation (1) of A_1. Since the character $\langle 0\rangle$ of the trivial representation equals 1, we do not use $\langle 0\rangle$. The coefficient of the power A^k in the power series is precisely the character of the A_1-representation (k) of dimension $k+1$.

The A_2-character generator, Eq. (2) of Ref. 6, is

$$\frac{1-PQ}{(1-\langle 10\rangle P+\langle 01\rangle P^2-P^3)(1-\langle 01\rangle Q+\langle 10\rangle Q^2-Q^3)}. \tag{2.2}$$

Again in the corresponding power series, the coefficient of the power P^pQ^q is the character of the A_2-representation (p,q).

The A_3-character generator can be inferred from the content of Ref. 13, or computed directly using Ref. 7. It can be brought to the form N/D, where

$$\begin{aligned}N=&(1-B^2)(1-AC+\langle 010\rangle ABC+A^2B+BC^2-AB^2C+A^2B^2C^2)\\&-(\langle 001\rangle-\langle 100\rangle B)AB(1+BC^2)-(\langle 100\rangle-\langle 001\rangle B)BC(1+A^2B),\\D=&(1-\langle 100\rangle A+\langle 010\rangle A^2-\langle 001\rangle A^3+A^4)\\&\times[1-\langle 010\rangle B+\langle 101\rangle B^2-(\langle 200\rangle+\langle 002\rangle)B^3+\langle 101\rangle B^4-\langle 010\rangle B^5+B^6]\\&\times(1-\langle 001\rangle C+\langle 010\rangle C^2-\langle 100\rangle C^3+C^4).\end{aligned} \tag{2.3}$$

Here the coefficients $\langle\lambda_1\lambda_2\lambda_3\rangle$ are the characters of representations $(\lambda_1\lambda_2\lambda_3)$ of A_3.

The B_2-character generator, Eq. (3) of Ref. 6, can be written as

$$\frac{(1+B)(1+A^2B)-\langle 01\rangle AB}{[1+A^4-\langle 01\rangle(A+A^3)+(1+\langle 10\rangle)A^2][1+B^4+(1-\langle 10\rangle)(B+B^3)+(1-\langle 10\rangle+\langle 02\rangle)B^2]}. \tag{2.4}$$

The coefficient of the term A^aB^b in the power expansion of (2.4) is the character of the representation (b,a) of B_2.

The last character generator we need is that of G_2 of Eq. (4) of Ref. 12. It can be brought to the form N/D, where

$$\begin{aligned}N=&1+A^4B^4+A+A^3B^4+(1+\langle 10\rangle)(B+A^4B^3)+(1-\langle 20\rangle)(AB+A^3B^3)\\&+(1+\langle 10\rangle)(B^2+A^4B^2)+(1+\langle 01\rangle)(A^2B+A^2B^3)-(\langle 20\rangle+\langle 01\rangle)(AB^2+A^3B^2)\\&+B^3+A^4B+(1-\langle 10\rangle)(A^3B+AB^3)+(\langle 11\rangle+2\langle 10\rangle)A^2B^2,\\D=&[1+A^6+(1-\langle 10\rangle)(A+A^5)+(1+\langle 01\rangle)(A^2+A^4)+(\langle 01\rangle-\langle 10\rangle-\langle 20\rangle)A^3]\\&\times[1+B^6+(1+\langle 10\rangle-\langle 01\rangle)(B+B^5)+(2-\langle 11\rangle+\langle 30\rangle)(B^2+B^4)\\&+(1+\langle 10\rangle-2\langle 01\rangle+\langle 30\rangle-\langle 02\rangle)B^3].\end{aligned} \tag{2.5}$$

3. CHARACTERS OF ELEMENTS OF FINITE ORDER

The character generators of Sec. 2 generate characters of all elements of the simple Lie group G on all its irreducible representations. Each character generator contains as coefficients the characters of a few lowest representations of the group. These character-coefficients are functions of l continuous variables. Choosing a particular element $\mathbf{x}$ of the group is equivalent to fixing the character variables. Consequently, substituting for the character-coefficients in the character generator these fixed values of the characters, one gets the generator of characters of the element $\mathbf{x}$ on irreducible representations of G.

A convenient way to specify an EFO was introduced by Kac[3]: the sets $\mathbf{s}=[\![s_0,s_1,\ldots,s_l]\!]$ of relatively prime nonnegative integers are in one-one correspondence with the conjugacy classes of EFO in G. These numbers are thought of as being attached to the extended Dynkin diagram, where the extension is the 0th node of the diagram.

Below we particularize the general properties[5] of ele-

ments of finite order and their characters to the cases under consideration.

The order of the EFO designated by $\mathbf{s}$ is $N = MC$ according to

	M	C
A_l	$\sum_{k=0}^{l} s_k$	$(l+1)/\gcd\left(\sum_{i=1}^{l} i s_i, l+1\right)$
B_2	$s_0 + s_1 + 2s_2$	$2/\gcd(2, s_1)$
G_2	$s_0 + 3s_1 + 2s_2$	1.

(3.1)

The EFO has within its conjugacy class a unique diagonal representative which acts on any weight space V^λ of the weight $\lambda = \Sigma_{i=1}^{l} c_i \alpha_i$ of any representation of G as

$$v \to \exp(2\pi i(\lambda, \mathbf{s})/M)v, \quad v \in V^\lambda, \tag{3.2}$$

where $(\lambda, \mathbf{s})$ is evaluated through

$$(\alpha_i, \mathbf{s}) = s_i, \quad i = 1,...,l. \tag{3.3}$$

Here α_i, $i = 1,2,...,l$ are the simple roots of G.

The character of the element $\mathbf{s} = [s_0 s_1 \cdots s_l]$ in the representation $(\lambda_1,...,\lambda_l)$ is denoted by $\langle \lambda_1,...,\lambda_l \rangle_{s_0 s_1 \cdots s_l}$. Whenever no ambiguity can arise, we omit the subscripts $s_0 s_1 \cdots s_l$. The character of $\mathbf{s}$ can readily be calculated[5] for representations of small dimension using (3.2) and (3.3). This allows one to calculate, for instance, the few characters of a general EFO which occur as coefficients in the character generators. In particular, one finds the following:

$$\langle 1 \rangle = e^{\pi i s_1/M} + e^{-\pi i s_1/M} = 2\cos \pi s_1/M \tag{3.4a}$$

of A_1,

$$\langle 10 \rangle = \langle 01 \rangle^* = e^{2\pi i(2s_1 + s_2)/3M} + e^{2\pi i(-s_1 + s_2)/3M} + e^{-2\pi i(s_1 + 2s_2)/3M} \tag{3.4b}$$

of A_2,

$$\langle 100 \rangle = \langle 001 \rangle^* = e^{\pi i(3s_1 + 2s_2 + s_3)/2M} + e^{\pi i(-s_1 + 2s_2 + s_3)/2M} + e^{\pi i(-s_1 - 2s_2 + s_3)/2M} + e^{-\pi i(s_1 + 2s_2 + 3s_3)/2M}, \tag{3.4c}$$

$$\langle 010 \rangle = e^{\pi i(s_1 + 2s_2 + s_3)/4M} + e^{\pi i(s_1 + s_3)/4M} + e^{\pi i(s_1 - s_3)/4M} + e^{-\pi i(s_1 - s_3)/4M} + e^{-\pi i(s_1 + s_3)/4M} + e^{-\pi i(s_1 + 2s_2 + s_3)/4M},$$

$$\langle 200 \rangle = \langle 002 \rangle^* = \langle 100 \rangle^2 - \langle 010 \rangle, \quad \langle 101 \rangle = \langle 100 \rangle \langle 001 \rangle - 1$$

of A_3,

$$\langle 01 \rangle = 4\cos\frac{\pi}{M}(s_1 + s_2)\cos\frac{\pi}{M}s_2,$$

TABLE I. List of all rational elements of finite order in the simple Lie groups of types A_1, A_2, A_3, B_2, and G_2. The full order of each element is given in column N. The last column contains the character-coefficients needed in the character generators. They are calculated from (3.4).

A_1

Rational element	N	$\langle 1 \rangle$
[10]	1	2
[01]	2	−2
[12]	3	−1
[11]	4	0
[21]	6	1

A_2

Rational element	N	$\langle 10 \rangle$
[100]	1	3
[011]	2	−1
[111]	3	0
[211]	4	1
[411]	6	2

A_3

Rational element	N	$\langle 100 \rangle$
[1000]	1	4
[0101]	2	0
[0010]	2	−4
[1101]	3	1
[1020]	3	−2
[1010]	4	0
[2101]	4	2
[0121]	4	−2
[1121]	5	−1
[0111]	6	−1
[2010]	6	2
[4101]	6	3
[2121]	6	0
[0141]	6	−3
[1111]	8	0
[2111]	10	1
[1212]	12	0
[6141]	12	1
[4161]	12	−1

B_2

Rational element	N	$\langle 10 \rangle$
[100]	1	5
[001]	2	−3
[010]	2	5
[101]	3	−1
[120]	3	2
[110]	4	1
[201]	4	1
[021]	4	1
[121]	5	0
[011]	6	−1
[210]	6	2
[401]	6	3
[221]	6	0
[041]	6	3
[111]	8	−1
[211]	10	0
[112]	12	−2
[641]	12	1
[461]	12	1

G_2

Rational element	N	$\langle 10 \rangle$	$\langle 01 \rangle$
[100]	1	7	14
[001]	2	−1	−2
[010]	3	−2	5
[101]	3	1	−1
[201]	4	3	2
[110]	4	−1	2
[401]	6	5	7
[310]	6	2	1
[111]	6	−1	1
[211]	7	0	0
[311]	8	1	0
[112]	8	−1	0
[313]	12	0	−1
[114]	12	−1	−1

$$\langle 10\rangle = 1 + 4\cos\frac{\pi}{M}(s_1 + 2s_2)\cos\frac{\pi}{M}s_1, \tag{3.4d}$$
$$\langle 02\rangle = \langle 01\rangle^2 - \langle 10\rangle - 1$$

of B_2,

$$\langle 10\rangle = 1 + 2\cos\frac{2\pi s_1}{M} + 2\cos\frac{2\pi}{M}(s_1 + s_2) + 2\cos\frac{2\pi}{M}(2s_1 + s_2),$$
$$\langle 01\rangle = 1 + \langle 10\rangle + 2\cos\frac{2\pi}{M}s_2 + 2\cos\frac{2\pi}{M}(3s_1 + s_2) + 2\cos\frac{2\pi}{M}(3s_1 + 2s_2),$$
$$\langle 20\rangle = \langle 10\rangle^2 - \langle 01\rangle - \langle 10\rangle - 1, \tag{3.4e}$$
$$\langle 11\rangle = \langle 10\rangle\langle 01\rangle - \langle 20\rangle - \langle 10\rangle,$$
$$\langle 30\rangle = \langle 10\rangle\langle 20\rangle - \langle 11\rangle - \langle 10\rangle - \langle 20\rangle - \langle 01\rangle,$$
$$\langle 02\rangle = \langle 01\rangle^2 - \langle 30\rangle - \langle 20\rangle - \langle 01\rangle - 1,$$

of G_2.

Upon substitution of the character-coefficients of (3.4) into the character generators of Sec. 2, one gets the generators of characters of EFO in irreducible representations of the corresponding Lie groups. There is a wealth of information contained in these generators, as exemplified in the last section.

4. CHARACTER GENERATORS FOR RATIONAL ELEMENTS

Among the infinity of EFO in G, a particularly interesting finite subset are the rational elements. An element $[\![s_0 s_1 \cdots s_l]\!]$ in G is called rational if all its characters $\langle \lambda_1 \ldots \lambda_l \rangle_{s_0 s_1 \cdots s_l}$ are rational. (Since these character values are algebraic integers they are, in fact, integers.) It follows that the character generator for a rational element must have all coefficients integer in its numerator and denominator. Equivalently, its character values in (3.4) must be integers.

The rational EFO in the five groups considered here were found in Ref. 5. They are reproduced in Table I together with the characters required as coefficients for the character generators. Let us consider each of the five groups separately.

The A_1-character generator, Eq. (2.1), contains the coefficient $\langle 1\rangle_{s_0 s_1}$ only. For all rational elements of A_1, these are shown in Table I. The character generators for A_1-rational elements are given in Table II. Let us point out the last three of them, corresponding to [12], [11], and [21]. They can be rewritten, respectively, as $(1 - A)/(1 - A^3)$, $(1 - A^2)/(1 - A^4)$, and $(1 + A)(1 - A^3)/(1 - A^6)$, which makes obvious the well-known fact (of Refs. 1 and 3) that all characters of these elements take only three values: ± 1 and 0.

TABLE II. Character generators for rational elements of A_1.

Rational element	Character generator	Distinct character values
[10]	$(1 - A)^{-2}$	infinitely many
[01]	$(1 + A)^{-2}$	infinitely many
[12]	$(1 + A + A^2)^{-1}$	± 1, 0
[11]	$(1 + A^2)^{-1}$	± 1, 0
[21]	$(1 - A + A^2)^{-1}$	± 1, 0

The A_2-character generator, Eq. (2.2), requires two coefficients $\langle 10\rangle_{s_0 s_1 s_2}$ and $\langle 01\rangle_{s_0 s_1 s_2}$, which are complex conjugate, in general, and therefore equal for the rational elements. Their values are shown in Table I for all rational elements of A_2. Due to the natural inclusion $A_2 \supset A_1$ specified in terms of representations as $(10) \supset (1) + (0)$, and the one-to-one correspondence between the rational elements of A_{2n} and A_{2n-1} established in Ref. 5, one has

$$\langle 10\rangle = \langle 1\rangle + 1, \tag{4.1}$$

as demonstrated by the corresponding entries in Table I. Table III contains the character generators for rational elements of A_2. As in the previous case, one finds from the character generators of the regular elements [111], [211], and [411] of A_2 that the characters $\langle \lambda_1\lambda_2\rangle_{111}$, $\langle \lambda_1\lambda_2\rangle_{211}$, and $\langle \lambda_1\lambda_2\rangle_{411}$ take only finitely many distinct values on all irreducible representations of A_2. Indeed, expanding the generator, one finds that

$$\langle \lambda_1\lambda_2\rangle_{111} = \pm 1 \text{ or } 0, \quad \langle \lambda_1\lambda_2\rangle_{211} = \pm 1 \text{ or } 0 \tag{4.2}$$

in agreement with Refs. 1 and 3, but also

$$\langle \lambda_1\lambda_2\rangle_{411} = \pm 3, \pm 2, \pm 1, \text{ or } 0. \tag{4.3}$$

The A_3-character generator, Eq. (2.3), contains six character-coefficients. However, due to the relations (3.4), only two of them are independent on rational (actually, real) elements, say, $\langle 100\rangle_{s_0 s_1 s_2 s_3}$ and $\langle 010\rangle_{s_0 s_1 s_2 s_3}$. The 19 rational elements A_3 are in one-to-one correspondence with the 19 rational elements of B_2 given by[5]

$$[\![s_0 s_1 s_2 s_1]\!] \leftrightarrow [\![s_0 s_2 s_1]\!]. \tag{4.4}$$

Due to the inclusion $B_2 \subset A_3$ and the reductions $(100) \supset (01)$ and $(010) \supset (10) + (00)$ of the representations of their B_2-content, one has the equality of characters of rational elements of A_3 and B_2:

TABLE III. Character generators for rational elements of A_2.

Rational element	Character generator	Distinct character values
[100]	$(1 - PQ)/(1 - P)^3(1 - Q)^3$	infinitely many
[011]	$(1 - PQ)/(1 - P^2)(1 + P)(1 - Q^2)(1 + Q)$	infinitely many
[111]	$(1 - PQ)/(1 - P^3)(1 - Q^3)$	± 1, 0
[211]	$(1 - PQ)(1 + P)(1 + Q)/(1 - P^4)(1 - Q^4)$	± 1, 0
[411]	$(1 - PQ)(1 + P)(1 + Q)(1 + P + P^2)(1 + Q + Q^2)/(1 - P^6)(1 - Q^6)$	$\pm 3, \pm 2, \pm 1$, 0

TABLE IV. Character generators of rational elements of finite order in A_3. Notations: $\Omega = 1 - AC + A^2B + BC^2 - AB^2C + A^2B^2C^2$ and $\Phi = AB + BC + A^2B^2C + AB^2C^2$.

Rational element	Character generator	Distinct character values
[1000]	$\{(1+B)(\Omega+6ABC)-4\Phi\}/(1-A)^4(1-B)^5(1-C)^4$	infinitely many
[0101]	$\{\Omega-2ABC\}/(1-A^2)^2(1-B^4)(1-C^2)^2$	infinitely many
[0010]	$\{(1+B)(\Omega+6ABC)+4\Phi\}/(1+A)^4(1-B)^5(1+C)^4$	infinitely many
[1101]	$\{(1+B)\Omega-\Phi\}/(1-A)(1-A^3)(1+B+B^2)(1-B^3)(1-C)(1-C^3)$	infinitely many
[1020]	$\{(1+B)(\Omega+3ABC)+2\Phi\}/(1+A+A^2)^2(1-B)^2(1-B^3)(1+C+C^2)^2$	infinitely many
[1010]	$\{\Omega+2ABC\}/(1+A^2)^2(1-B)^2(1-B^2)(1+C^2)^2$	infinitely many
[2101]	$\{(1+B)(\Omega+2ABC)-2\Phi\}/(1-A)^2(1+A^2)(1-B)(1+B^2)^2(1-C)^2(1+C^2)$	infinitely many
[0121]	$\{(1+B)(\Omega+2ABC)+2\Phi\}/(1+A)^2(1+A^2)(1-B)(1+B^2)^2(1+C)^2(1+C^2)$	infinitely many
[1121]	$\{(1+B)(\Omega+ABC)+\Phi\}(1-A)(1-C)/(1-A^5)(1-B^5)(1-C^5)$	$\pm 1, 0$
[0111]	$\{(1+B)\Omega+\Phi\}/(1+A)(1+A^3)(1+B+B^2)(1-B^3)(1+C)(1+C^3)$	infinitely many
[2010]	$\{(1+B)(\Omega+3ABC)-2\Phi\}/(1-A+A^2)^2(1-B)^2(1-B^3)(1-C+C^2)^2$	infinitely many
[4101]	$\{(1+B)(\Omega+4ABC)-3\Phi\}/(1-A)^2(1-A+A^2)(1-B)(1-B+B^2)^2(1-C)^2(1-C+C^2)$	infinitely many
[2121]	$\{\Omega+ABC\}(1-A^2)(1-C^2)(1+B+B^2)/(1-A^6)(1-B^6)(1-C^6)$	$\pm 1, 0$
[0141]	$\{(1+B)(\Omega+4ABC+3\Phi)\}/(1+A)^2(1+A+A^2)(1-B)(1-B+B^2)^2(1+C)^2(1+C+C^2)$	infinitely many
[1111]	$\Omega/(1+A^4)(1-B^4)(1+C^4)$	$\pm 1, 0$
[2111]	$\{(1+B)(\Omega+ABC)-\Phi\}(1+A)(1+C)/(1+A^5)(1-B^5)(1+C^5)$	$\pm 1, 0$
[1212]	$\{\Omega-ABC\}(1+A^2)(1+C^2)(1-B+B^2)/(1+A^6)(1-B^6)(1+C^6)$	$\pm 1, 0$
[6141]	$\{(1+B)(\Omega+2ABC)-\Phi\}(1-A^2+A^4)(1+A)(1-A^3)(\Sigma_{i=0}^{5}B^i)(1+B^2)$ $\times(1-C^2+C^4)(1+C)(1-C^3)/(1-A^{12})(1-B^{12})(1-C^{12})$	$\pm 4, \pm 3, \pm 2, \pm 1, 0$
[4161]	$\{(1+B)(\Omega+2ABC)+\Phi\}(1-A^2+A^4)(1-A)(1+A^3)(\Sigma_{i=0}^{5}B^i)(1+B^2)$ $\times(1-C^2+C^4)(1-C)(1+C^3)/(1-A^{12})(1-B^{12})(1-C^{12})$	$\pm 4, \pm 3, \pm 2, \pm 1, 0$

$$\langle 100\rangle_{s_0s_1s_2s_1} = \langle 01\rangle_{s_1s_2s_1}, \tag{4.5}$$

$$\langle 010\rangle_{s_0s_1s_2s_1} = \langle 10\rangle_{s_1s_2s_1} + 1. \tag{4.6}$$

In Table I we have shown only one of the two equal characters of rational elements. The character generators of all 19 rational EFO of A_3 are given in Table IV.

The B_2-character generator, Eq. (2.4), contains two independent coefficients: $\langle 10\rangle_{s_0s_1s_2}$ and $\langle 01\rangle_{s_0s_1s_2}$. The first one is given in Table I, the second is found from (4.5) in the A_3 part of Table I. All 19 character generators for rational elements of B_2 are given in Table V.

The G_2-character generator, Eq. (2.5), contains the coefficients $\langle 10\rangle_{s_0s_1s_2}$ and $\langle 01\rangle_{s_0s_1s_2}$ given in Table I. The generators for characters of rational elements of finite order in G_2 are given in Table VI.

The last columns in Tables II–VI contain the character values for a given element whenever their number is finite. One notices that the set of distinct character values is finite precisely for the regular EFO in any simple Lie group.[5] It was shown in Ref. 3 that the set of character values of nonregular EFO is unbounded. Let us be reminded here that an EFO is regular iff its centralizer in the Lie group is of mini-

TABLE V. Character generators for rational elements of B_2.

Rational element	Character generator	Distinct character values
[100]	$\{(1+B)(1+A^2B)-4AB\}/(1-A)^4(1-B)^4$	infinitely many
[001]	$(1+A^2B)/(1-A^2)^2(1+B)^3$	infinitely many
[010]	$\{(1+B)(1+A^2B)+4AB\}/(1+A)^4(1-B)^4$	infinitely many
[101]	$\{(1+B)(1+A^2B)-AB\}/(1-A)(1-A^3)(1+B+B^2)^2$	infinitely many
[120]	$\{(1+B)(1+A^2B)+2AB\}/(1+A+A^2)^2(1-B^3)(1-B)$	infinitely many
[110]	$(1+B)(1+A^2B)/(1+A^2)^2(1-B^2)^2$	infinitely many
[201]	$\{(1+B)(1+A^2B)-2AB\}/(1+A^2)(1-A)^2(1+B^2)$	infinitely many
[021]	$\{(1+B)(1+A^2B)+2AB\}/(1-A^2)(1+A)^2(1+B^2)^2$	infinitely many
[121]	$\{(1+B)(1+A^2B)+AB\}(1-A)(1-B)/(1-A^5)(1-B^5)$	$\pm 1, 0$
[011]	$\{(1+B)(1+A^2B)+AB\}/(1+A)(1+A^3)(1+B+B^2)^2$	infinitely many
[210]	$\{(1+B)(1+A^2B)-2AB\}/(1-A+A^2)^2(1+B+B^2)(1-B)^2$	infinitely many
[401]	$\{(1+B)(1+A^2B)-3AB\}/(1-A)^2(1-A+A^2)(1-B+B^2)^2$	infinitely many
[221]	$(1-A^2)(1+A^2B)/(1-A^6)(1+B^3)$	$\pm 1, 0$
[041]	$\{(1+B)(1+A^2B)+3AB\}/(1+A)^2(1+A+A^2)(1-B+B^2)^2$	infinitely many
[111]	$(1+A^2B)(1-B)/(1+A^4)(1-B^4)$	$\pm 1, 0$
[211]	$\{(1+B)(1+A^2B)-AB\}(1+A)(1-B)/(1+A^5)(1-B^5)$	$\pm 1, 0$
[112]	$(1-B)(1+A^2B)(1+A^2)(1-B+B^2)/(1+A^6)(1-B^6)$	$\pm 2, \pm 1, 0$
[641]	$\{(1+B)(1+A^2B)-AB\}(1+B^2)(1+A)(1-A^3)(1-A^2+A^4)/(1-A^{12})(1+B^6)$	$\pm 2, \pm 1, 0$
[461]	$\{(1+B)(1+A^2B)+AB\}(1+B^2)(1-A)(1+A^3)(1-A^2+A^4)/(1-A^{12})(1+B^6)$	$\pm 2, \pm 1, 0$

TABLE VI. Character generators of rational elements of finite order in G_2.

Rational element	Character generator	Distinct character values
[100]	$\{1+A+8B-26AB+8B^2+15A^2B-41AB^2+B^3-6A^3B+78A^2B^2-6AB^3+A^4B-41A^3B^2$ $+15A^2B^3+8A^4B^2-26A^3B^3+8A^4B^3+A^3B^4+A^4B^4\}/(1-A)^6(1-B)^6$	infinitely many
[001]	$\{1+A-2AB-A^2B-AB^2+B^3+2A^3B-2A^2B^2+2AB^3+A^4B-A^3B^2-A^2B^3-2A^3B^3+A^3B^4+A^4B^4\}/$ $(1+A)^2(1-A^2)^2(1+B)^2(1-B^2)^2$	infinitely many
[010]	$\{1+A-B+AB-B^2+6A^2B-5AB^2+B^3+3A^3B-4A^2B^2+AB^3+A^4B-5A^3B^2+6A^2B^3-A^4B^2$ $+A^3B^3-A^4B^3+A^3B^4+A^4B^4\}/(1+A+A^2)^3(1-B)^6$	infinitely many
[101]	$\{1+A+2B+AB+2B^2+AB^2+B^3+A^4B+A^3B^2+2A^4B^2+A^3B^3+2A^4B^3+A^3B^4+A^4B^4\}/$ $(1-A^3)^2(1+B+B^2)^3$	infinitely many
[201]	$\{1+A+4B-2AB+4B^2+3A^2B-AB^2+B^3-2A^3B+6A^2B^2-2AB^3+A^4B-5A^3B^2$ $+3A^2B^3+4A^4B^2-2A^3B^3+4A^4B^3+A^3B^4+A^4B^4\}/(1+A^2)^2(1-A)^2(1+B)^2(1+B^2)^2$	infinitely many
[110]	$\{1+A+2AB+3A^2B-AB^2+B^3+2A^3B-2A^2B^2+2AB^3+A^4B-A^3B^2+3A^2B^3+2A^3B^3+A^3B^4+A^4B^4\}/$ $(1+A)^2(1+A^2)^2(1+B^2)^2(1-B)^2$	infinitely many
[401]	$\{1+A+6B-11AB+6B^2+8A^2B+19AB^2+B^3-4A^3B+28A^2B^2-4AB^3+A^4B-19A^3B^2$ $+8A^2B^3+6A^4B^2-11A^3B^3+6A^4B^3+A^3B^4+A^4B^4\}/(1-A+A^2)^2(1-A)^2(1+B^2+B^4)(1-B+B^2)$	infinitely many
[310]	$\{1+A+3B+AB+3B^2+2A^2B-AB^2+B^3-A^3B+4A^2B^2+A^4B-A^3B^2+2A^2B^3$ $+3A^4B^2+A^3B^3+3A^4B^3+A^3B^4+A^4B^4\}/(1-A+A^2)^2(1+A+A^2)(1-B^2)^2(1+B)^2$	infinitely many
[111]	$\{1-A+B-A^4B+A^3B^2-A^4B^2\}(1-B^2)/(1-A^6)(1-B^6)$	$\pm 1, 0$
[211]	$\{1-A^2+AB-AB^2+A^2B^2-B^4-A^5B+A^3B^3-A^4B^3+A^4B^4-A^3B^5+A^5B^5\}/(1-A^7)(1-B^7)$	$\pm 1, 0$
[311]	$\{1+A-B^3-AB^2+A^2B-A^2B^3+A^4B-A^4B^4+A^3B^2-A^3B^4\}(1-A^2)(1+B^2)/(1-A^8)(1-B^8)$	$\pm 1, 0$
[112]	$\{1+A^2B+A^3B+B^3+AB^3-A^2B^2-AB^2+A^3B^4\}(1-A)(1+A^2)(1-B^2)/(1-A^8)(1-B^8)$	$\pm 1, 0$
[313]	$(1-A^4)(1+A^3)(1-B)(1+A^3B)/(1-A^{12})(1-B^4)$	$\pm 1, 0$
[114]	$\{1-AB+A^2B+A^3B-AB^2-A^2B^2+B^3+AB^3-A^2B^3+A^3B^4\}(1-A)(1-B+B^2)(1+B^4)/(1-A^6)(1+B^{12})$	$\pm 2, \pm 1, 0$

mal dimension, i.e., the rank. This is equivalent to the condition $s_i > 0$ for all $i = 0,...,l$.

5. GAUSSIAN ELEMENTS OF FINITE ORDER

The rational EFO considered in Sec. 4 are undoubtedly the most convenient EFO to work with in many cases because of their integer characters. However, for groups with complex-valued characters of EFO, i.e., A_n, $n > 1$, D_{2k+1}, $k > 1$, and E_6, the rational elements have a drawback: they take the same character values on pairs of contragredient representations. That alone is a motivation to study nonrational and/or nonreal EFO.

Next simplest EFO are those whose characters take values from a quadratic number field. Such a character can be written as an integer linear combination of two basis elements, one of them being 1, and the second containing $\sqrt{p}$ or $\sqrt{-p}$, where p is a square free integer. Here we are concerned with the simplest case: the Gaussian numbers $a + ib$ with a,b integer.

Among the five simple groups considered in this paper only A_2 and A_3 have complex-valued characters. The number of conjugacy classes of Gaussian EFO (more generally any EFO of a given quadratic type) is finite in any simple Lie group.[5] In A_2 and A_3 these are shown in Tables VII and VIII together with the relevant character coefficients. We call EFO Gaussian if its characters are found in the Gaussian field and not in its subfield, i.e., rational EFO are not called Gaussian.

The fact that Tables VII and VIII contain all Gaussian EFO in A_2 and A_3 requires a proof. A simple one can be given, for instance, in the following way. In Ref. 5 it was proven that order N of Gaussian EFO satisfies $N = 0$ (mod 4) in any Lie group, and that upper limit $N_{\max}$ for N equals 12 and 20, respectively, for A_2 and A_3. Hence it suffices to compute the character coefficients for all EFO with $N \leqslant N_{\max}$ in order to decide which of them are Gaussian. The list of EFO is readily compiled using (3.1) and the character coefficients are calculated from (3.4b) and (3.4c).

TABLE VII. All conjugacy classes of Gaussian EFO in A_2 and the character coefficients relevant for (3.4).

Gaussian element	N	$\langle 01 \rangle$
[130]	4	$-1+2i$
[103]	4	$-1-2i$
[341]	8	i
[314]	8	$-i$
[174]	12	$-1+i$
[147]	12	$-1-i$

The Gaussian EFO of Table VII are divided into pairs with complex conjugate characters $\langle 01 \rangle$. It follows from (2.2) that the two EFO of each pair have complex conjugate characters on any representation of A_2. Consequently, it suffices to calculate the character generator for one EFO of the pair, say the first one, and to find the other by complex conjugation. Substitution of $\langle 01 \rangle$ into (2.2) and obvious simplifications give the following character generators:

$$[130]: \quad \frac{1-PQ}{(1+P)(1+iP)^2(1+Q)(1-iQ)^2}, \tag{5.1}$$

$$[341]: \quad \frac{1-PQ}{(1+iP)(1+iP^2)(1-iQ)(1-iQ^2)}, \tag{5.2}$$

$$[174]: \quad \frac{1-PQ}{(1+P)(1+iP-P^2)(1+Q)(1-iQ-Q^2)}. \tag{5.3}$$

TABLE VIII. All conjugacy classes of Gaussian EFO in A_3. The last four columns provide information about the character generators as explained in Sec. 5.

Gaussian element	N	$\langle 100 \rangle$	$\langle 010 \rangle$	Related EFO	Phase factor A	B	C
[0001]	4	$4i$	-6	[1000]	i	-1	$-i$
[0100]	4	$-4i$	-6		$-i$	-1	i
[1012]	4	$2i$	-2	[0121]	$-i$	-1	i
[1210]	4	$-2i$	-2		i	-1	$-i$
[3113]	8	$1+i$	0	[3113]	1	1	1
[1133]	8	$-1+i$	0		i	-1	$-i$
[3311]	8	$1-i$	0		$-i$	-1	i
[1331]	8	$-1-i$	0		-1	1	-1
[1410]	12	$-3i$	-4	[4101]	$-i$	-1	i
[1014]	12	$3i$	-4		i	-1	$-i$
[0201]	12	$-2i$	-3	[1020]	i	-1	$-i$
[0102]	12	$2i$	-3		$-i$	-1	i
[1011]	12	i	0	[1101]	i	-1	$-i$
[1110]	12	$-i$	0		$-i$	-1	i
[1416]	12	i	-2	[6141]	i	-1	$-i$
[1614]	12	$-i$	-2		$-i$	-1	i
[1211]	20	$-i$	-1	[1121]	i	-1	$-i$
[1112]	20	i	-1		$-i$	-1	i

The last two EFO are regular; hence they have only finitely many distinct character values. This becomes obvious if the character generators are rewritten in an equivalent form:

$$[341]: \frac{(1-PQ)(1-iP)(1-iP^2)(1-P^2)(1+iQ)(1+iQ^2)(1-Q^2)}{(1-P^8)(1-Q^8)}, \tag{5.2'}$$

$$[174]: \frac{(1-PQ)(1-P)(1+P^2)(1+P^2+P^4)(1-iP-P^2)(1-Q)(1+Q^2)(1+Q^2+Q^4)(1+iQ-Q^2)}{(1-P^{12})(1-Q^{12})}. \tag{5.3'}$$

The distinct character values are then found as coefficients of the numerators in (5.2') and (5.3') which have been multiplied out. Thus the characters of [341] take nine distinct values: 0, ± 1, $\pm i$, $\pm 1 \pm i$, and those of [174] take 17 values: 0, ± 1, ± 2, ± 3, $\pm 2i$, $\pm 1 \pm i$, $\pm 1 \pm 2i$.

The first two columns of Table VIII contain similar information about Gaussian EFO in A_3. The fifth column of the table indicates an EFO related by permutation of the labels s_i of the nodes of the extended Coxeter–Dynkin diagram. In all but one case the related EFO is a rational one and its character generator is therefore found in Table IV. Only for [3113] the character generator has to be newly calculated from (2.3); it is

$$[3113]: \quad \{(1-B^2)\Omega - (1-B^2)\Phi + i(1+B)(AB - BC + AB^2C^2 - A^2B^2C)\}$$
$$\times(1+A)(1+iA)(1+iA^2)(1-B^2)(1+C)(1-iC)(1-iC^2)/(1-A^8)(1-B^8)(1-C^8), \tag{5.4}$$

where Ω and Φ are the same as in Table IV.

Then the character generators of a Gaussian element are obtained from that of the EFO of the fifth column by multiplying its variables A, B, and C by the phase factors shown in columns 6, 7, and 8, respectively. Thus, for instance, from the character generator of the identity EFO [1000] of Table IV one obtains by the substitution $A \to iA$, $B \to -B$, $C \to -iC$ the character generator for the Gaussian element [0001]:

$$\frac{\{(1-B)(1-AC+A^2B+BC^2-AB^2C+A^2B^2C^2)+4i(AB-BC-A^2B^2C+AB^2C^2)\}}{(1-iA)^4(1+B)^5(1+iC)^4}. \tag{5.5}$$

Consequently the characters of the element [0001] have absolute value equal to the dimension of the A_3-representation, and phase equal to one of the fourth roots of 1. The same obviously holds for the characters of the Gaussian element [0100]. Also, let us point out that the center of the group A_3 consists of the elements [1000], [0100], [0010], and [0001].

In Table VIII there are eight EFO which are regular. These elements have only finitely many distinct character values in irreducible representations of A_3. Multiplying out the numerators of the corresponding character generators, one finds the values each character can take. Namely, the following:

$$[3113],[1331],[3311],[1133]: 0, \pm 1, \pm 1 \pm i, \pm 2, \pm 2i,$$
$$[1614],[1416]: 0, \pm 1, \pm i, \pm 2, \pm 2i, \pm 3, \pm 3i, \pm 4, \pm 4i,$$
$$[1211],[1112]: 0, \pm 1, \pm i. \tag{5.6}$$

6. COMMENTS AND EXAMPLES

Let us first point out some of the general properties of EFO on examples of EFO in the five simple Lie groups here. The character generators make them particularly visible.

(1) EFO with only real character values. Clearly, in order that all characters of a given EFO are real, it is necessary and sufficient that the character-coefficients of the character generators are real. That is the case for all EFO in A_1, B_2, and G_2. It is also true for precisely those EFO of A_2 and A_3, which are given by $[s_0s_1s_1]$ and $[s_0s_1s_2s_1]$, respectively. More generally, one concludes that all characters of an element $[s_0s_1\cdots s_l] \in G$ take their values from the same (cyclotomic) number field as do the corresponding character-coefficients.

(2) In order that $[s_0s_1\cdots s_l]$ is a real EFO in a simple Lie group G, it is necessary and sufficient[5] that $(s_1s_2\cdots s_l)$ denotes a selfcontragredient representation of the group G.

(3) There is a one-to-one correspondence between the real EFO of A_2 and A_1 given by

$$[s_0s_1s_1] \leftrightarrow [s_0, 2s_1]/\gcd(s_0, 2s_1) \tag{6.1}$$

and between the real EFO of A_3 and B_2 given by (4.4), or, more generally, between the real elements of A_{2n-1} and those of C_n (Ref. 5):

$$[s_0s_1\cdots s_n\cdots s_1] \leftrightarrow [s_0s_1\cdots s_n]. \tag{6.2}$$

(4) All 19 rational elements of A_4 can be found from those of A_3 by means of the correspondence with the rational EFO's of A_3:

$$[s_0s_1s_2s_1] \leftrightarrow [s_0, s_1, \tfrac{1}{2}s_2, \tfrac{1}{2}s_2, s_1] \quad \text{if } s_2 \text{ is even}, \tag{6.3}$$

or

$$[s_0s_1s_2s_1] \leftrightarrow [2s_0, 2s_1, s_2, s_2, 2s_1] \quad \text{if } s_2 \text{ is odd}. \tag{6.4}$$

(5) Among the 59 rational elements of A_5 (they were found in Ref. 5), 19 can be simply found because of the correspondence between the rational EFO's of A_4 and A_5:

$$[s_0s_1s_2s_2s_1] \leftrightarrow [s_0s_1s_20s_2s_1]. \tag{6.5}$$

(6) Consider a pair of simple Lie groups $G \supset G'$, and a pair $\mathbf{s}$ and $\mathbf{s}'$ of EFO such that $\mathbf{s} \in G$ and $\mathbf{s}' \in G'$ and $\mathbf{s} \leftrightarrow \mathbf{s}'$ through one of the relations (6.1)–(6.5). Then from the reduction $\lambda \supset \lambda'$ of representations λ of G to λ' of G' follows the equality of characters

$$\langle \lambda \rangle_s = \langle \lambda' \rangle_{s'}. \tag{6.6}$$

The reduction for most cases of interest is found, for instance, in Ref. 14. Examples are, for instance, Eqs. (4.1) and (4.4)–(4.6).

TABLE IX. Nonzero characters of the two regular Gaussian EFO of A_2. At the intersection of pth column and qth row is the character of the EFO in the A_2 representation (p,q) (mod N), where $N = 8$ for [341] and $N = 12$ for [174].

[341]

q \ p	0	1	2	3	4	5	6
0	1	$-i$	$-1-i$	$-1+i$	i	1	
1	i		1		$-i$		-1
2	$-1+i$	1	1	$-1-i$		i	$-i$
3	$-1-i$		$-1+i$		$1+i$		$1-i$
4	$-i$	i		$1-i$	-1	-1	$1+i$
5	1		$-i$		-1		i
6		-1	i	$1+i$	$1-i$	$-i$	-1

[174]

q \ p	0	1	2	3	4	5	6	7	8	9	10
0	1	$-1-i$	$1+i$	$-1-2i$	$2i$	$-2i$	$-1+2i$	$1-i$	$-1+i$	1	
1	$-1+i$	1	$-1+i$	2	-1	2	$-1-i$	1	$-1-i$		-1
2	$1-i$	$-1-i$		$-1-i$	$-1+i$		$-1+i$	$1+i$		$1+i$	$1-i$
3	$-1+2i$	2	$-1+i$	3	$-1-i$	2	$-1-2i$		$-1-i$	-1	$-1+i$
4	$-2i$	-1	$-1-i$	$-1+i$	-1	$2i$		$1+2i$	$1-i$	$1+i$	$1-2i$
5	$2i$	2		2	$-2i$		$-2i$	-2		-2	$2i$
6	$-1-2i$	$-1+i$	$-1-i$	$-1+2i$		$2i$	1	$1+i$	$1-i$	1	$-2i$
7	$1+i$	1	$1-i$		$1-2i$	-2	$1-i$	-3	$1+i$	-2	$1+2i$
8	$-1-i$	$-1+i$		$-1+i$	$1+i$		$1+i$	$1-i$		$1-i$	$-1-i$
9	1		$1-i$	-1	$1-i$	-2	1	-2	$1+i$	-1	$1+i$
10		-1	$1+i$	$-1-i$	$1+2i$	$-2i$	$2i$	$1-2i$	$-1+i$	$1-i$	-1

(7) One of the obvious applications of characters of EFO's is the decomposition of tensor products of representations of the corresponding Lie group. Namely, given the tensor product

$$\phi_1 \otimes \phi_2 = \bigoplus_{i=3}^{k} \phi_i \tag{6.7}$$

of representations ϕ of G, one has the corresponding equality of characters

$$\langle s^1 \rangle \times \langle s^2 \rangle = \sum_{i=3}^{k} \langle s^i \rangle, \tag{6.8}$$

where $\langle s^i \rangle$ is the character of an element $\mathbf{s} \in G$ on the representation ϕ_i of G.

(8) Let us consider the element [10···0] in a simple Lie group G. It is the element of order $N = M = 1$, i.e., the identity element of the group. Consequently, its character in every representation of the group is equal to the dimension of the representation of G. The character generator corresponding to it is the generator of dimensions[6] of representations of G. In particular, let G be the exceptional simple Lie group of type G_2. The character generator of the element [100] is found in Table VI. The first terms of the power expansion are

$$1 + 7A + 14B + 64AB + \cdots. \tag{6.9}$$

The coefficient of a power $A^a B^b$ is the dimension of the irreducible representation (ab) of G_2.

(9) As the next example consider the two Gaussian EFO in A_2 which are regular. Their character generators are given by (5.2) and (5.3), and also by (5.2′) and (5.3′). Multiplying out the numerators of (5.2′) and (5.3′) one finds the character values in all A_2 representations (p,q) with $p < N$ and $q < N$, i.e., $N = 8$ for [341] and $N = 12$ for [174]. Comparing the degrees of the numerators and denominators one readily concludes that $\langle p,q \rangle = \langle p',q' \rangle$ for $p = p'$ and $q = q'$ (mod N). Therefore Table IX contains the characters for *all* representations of A_2.

(10) The conjugacy classes determined by [11···1] and [21···1] in the adjoint group are the unique classes of regular EFO of order $N = h$ and $N = h + 1$, respectively, where h is the Coxeter number of the Lie group G. In Ref. 3 it was shown that their characters are zero on an irreducible representation Λ iff $2(\Lambda + \rho,\alpha)/(\alpha,\alpha) = 0$ (mod N) for a root α. Here ρ denotes the half-sum of all positive roots of G.

(11) Consider the example of the EFO [1ss] in A_2. The element is real because $s_1 = s_2$ and, in general, it is not rational. For $s > 0$ it is regular so that its characters take only finitely many distinct values in irreducible representations of A_2. One has

$$\langle 10 \rangle_{1ss} = \langle 01 \rangle_{1ss} \tag{6.10}$$

for all s. The character coefficient $\langle 10 \rangle_{1ss}$ is given in (3.4b). It is

$$\langle 10 \rangle_{1ss} = 1 + 2\cos 2\pi s/(2s+1). \tag{6.11}$$

For $s = 0$ or 1 it is integer. Consequently, the elements [100] and [111] are rational. If, however, $s > 1$, say $s = 2$, one has

$$\langle 10\rangle_{122} = \tfrac{1}{2}(1-\sqrt{5}) \equiv \tau. \tag{6.12}$$

It follows immediately that every character $\langle \lambda_1\lambda_2\rangle_{122}$ is of the form $a + b\sqrt{5}$ with a and b integer or half-odd. Moreover, the character generator can be brought to the form

$$\frac{(1-PQ)(1+\tau P+P^2)(1+\tau Q+Q^2)}{(1-P^5)(1-Q^5)}, \tag{6.13}$$

from which it is clear that $\langle \lambda_1\lambda_2\rangle_{122}$ can take only one of the following values: $0, \pm 1, \pm\tau$.

(12) The character generators for some of the elements, for instance $[11]$ in A_1, $[011]$ in A_2, or $[1010]$ and $[1001]$ in A_3 can be interpreted as generators of signatures of representations of the noncompact real forms SU(1,1), SU(2,1), or SU(2,2) and SU(3,1) of the corresponding groups. An independent computation of these signatures can be found in Ref. 15.

(13) Finite subgroups of Lie groups are of considerable interest in mathematics and, in a different way, also in physics. By definition, any finite group F consists of EFO. If one has $F \subset G$, where G is a Lie group, it should be possible to identify elements of F with certain elements of G. However, here we have studied only G-conjugacy classes of EFO; therefore only the G-conjugates of elements of F can be identified in our description (for instance, by the values of their characters in sufficiently many representations of G). To construct actual *nonabelian* finite subgroups of Lie groups is a problem which has no known general solution in mathematics, although simple particular cases can undoubtedly be solved.

ACKNOWLEDGMENT

We are grateful to W. McKay for computation of characters of EFO's for particular representations and for evaluation of numerators of character generators of regular EFO.

[1] B. Kostant, "On Macdonald's η-function formula, the Laplacian and generalized exponents," Adv. Math. **20**, 179–212 (1976).

[2] V. Kac, "Automorphisms of finite order of semisimple Lie algebras," Func. Anal. Appl. **3**, 252–254 (1969).

[3] V. Kac, "Simple Lie groups and the Legendre symbol," in *Lecture Notes in Mathematics*, Vol. 848 (Springer, New York, 1980).

[4] J. Conway and L. Queen, "Computing the character table of a Lie group" (to be published).

[5] R. V. Moody and J. Patera, "Characters of elements of finite order in Lie groups," submitted to SIAM J. Algebraic Discrete Methods.

[6] J. Patera and R. T. Sharp, "Generating functions for characters of group representations and their applications," in *Lecture Notes in Physics*, Vol. 94 (Springer, New York, 1979), pp. 175–183.

[7] R. P. Stanley, "The SU(n) character generator," J. Math. Phys. **21**, 2321–2326 (1980).

[8] K. Baclawski, "Character generators for unitary and symplectic groups," J. Math. Phys. **24**, 1688–1694 (1983).

[9] R. C. King, "The character generator for Sp($2k$)," C. R. Math. Rep. Acad. Sci. Canada **III**, 149–153 (1981).

[10] R. C. King, "The character generators of SO($2k$) and SO($2k+1$)" (to be published).

[11] R. Gaskell and R. T. Sharp, "Generating functions for G_2 characters and subgroup branching rules," J. Math. Phys. **22**, 2736–9 (1981); R. Gaskell, "Character generators for compact semisimple Lie groups" (to be published).

[12] P. E. Desmier, J. Patera, and R. T. Sharp, "Finite subgroup bases for compact Lie groups," Physica **114A**, 336–340 (1982).

[13] R. T. Sharp and D. Lee, "SU(4) van der Waerden invariant," Rev. Mexicana Fís. **20**, 203–215 (1971).

[14] W. G. McKay and J. Patera, *Tables of dimensions, indices and branching rules for representations of simple Lie algebras* (Marcel Dekker, New York, 1981).

[15] J. Patera and R. T. Sharp, "Signatures of all finite representations of SU(p,q), $p+q<4$," Kinam **4**, 93–98 (1982).

Special properties of the irreducible representations of the proper Lorentz group

K. N. Srinivasa Rao,[a] A. V. Gopala Rao, and B. S. Narahari
Department of Physics, University of Mysore, Mysore 570 006, India

(Received 18 May 1982; accepted for publication 5 November 1982)

It is shown that the finite- and infinite-dimensional irreducible representations (j_0, c) of the proper Lorentz group SO(3,1) may be classified into the two categories, namely, the *complex-orthogonal* and the *symplectic* representations; while all the integral-j_0 representations are equivalent to complex-orthogonal ones, the remaining representations for which j_0 is a half-odd integer are symplectic in nature. This implies in particular that all the representations belonging to the *complementary series* and the subclass of integral-j_0 representations belonging to the *principal series* are equivalent to *real-orthogonal* representations. The rest of the principal series of representations for which j_0 is a half-odd integer are symplectic in addition to being unitary and this in turn implies that the D^j representation of SO(3) with half-odd integral j is a subgroup of the unitary symplectic group USp$(2j+1)$. The infinitesimal operators for the integral-j_0 representations are constructed in a suitable basis wherein these are seen to be complex skew-symmetric in general and real skew-symmetric in particular for the unitary representations, exhibiting explicitly the aforementioned properties of the integral-j_0 representations. Also, by introducing a suitable *real* basis, the finite-dimensional $(j_0 = 0, c = n)$ representations, where n is an integer, are shown to be *real-pseudo-orthogonal* with the signature $(n(n+1)/2, n(n-1)/2)$. In any general complex basis, these representations $(0, n)$ are also shown to be *pseudo-unitary* with the same signature $(n(n+1)/2, n(n-1)/2)$. Further it is shown that no other finite-dimensional irreducible representation of SO(3,1) possesses either of these two special properties.

PACS numbers: 02.20.Qs, 02.20.Rt

I. INTRODUCTION

It is well known that the self-representation $D^{\frac{1}{2}\frac{1}{2}}$ of the proper Lorentz group SO(3,1) is complex orthogonal in Minkowski coordinates with $x^4 \equiv ict$. One also knows[1,2] that SO(3,C) is isomorphic to SO(3,1), which means that the representations D^{01} and D^{10} are also equivalent to complex-orthogonal representations. In other words, the representation matrices (D) of the irreducible representations $D^{\frac{1}{2}\frac{1}{2}}$, D^{01}, and D^{10} of SO(3,1) are known to satisfy the relation

$$\tilde{D}D = D\tilde{D} = E, \tag{1.1}$$

in some suitable basis, where $\tilde{D}$ is the transpose of D and E is a unit matrix of the appropriate dimension. On the other hand, if D is a complex unimodular matrix of two dimensions belonging to the group SL(2,C) which provides the $D^{\frac{1}{2}0}$ (or $D^{0\frac{1}{2}}$) representation of SO(3,1), we also have the similar result[3] that

$$\tilde{D}GD = G = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{1.2}$$

This simply means that the SL(2,C) representation of SO(3,1) is *symplectic* (see any one of the Refs. 4–8 for the definition of the symplectic groups).

In a notation[9–11] that is most convenient for describing both the finite- and infinite-dimensional representations of SO(3,1), where each irreducible representation is characterized by an index pair (j_0, c), where *j_0 is a positive integer or a half-odd integer and c is an arbitrary complex number*, the above-mentioned results would mean that the self-representation (0,2) and the representations (1,2) and (1, −2) of SO(3,1) are equivalent to complex-orthogonal ones whereas the representations $(\frac{1}{2}, \frac{3}{2})$ and $(\frac{1}{2}, -\frac{3}{2})$ are symplectic.

In this paper we show that *all the representations (j_0, c) of SO(3,1), both finite- and infinite-dimensional, are equivalent either to symplectic or to complex-orthogonal representations.*

In the rest of this section, we give a very brief description of the irreducible representations (j_0, c) of SO(3,1) for the sake of completeness and to fix the notation followed in this paper. The carrier space $B(j_0, c)$ of the irreducible representation (j_0, c) of the proper Lorentz group is characterized by the nonnegative integral or half-odd-integral number j_0 and may be analyzed[9–11] as the direct sum of a sequence of finite-dimensional subspaces M_j. The positive number j designating the subspace M_j is called its weight and is also either integral or half-odd-integral with j_0. Each of these M_j is invariant with respect to the irreducible representation D^j of the subgroup of rotations and occurs exactly once in the decomposition of $B(j_0, c)$ into a direct sum of subspaces so that

$$B(j_0, c) = M_{j_0} \dotplus M_{j_0+1} \dotplus M_{j_0+2} \dotplus \cdots, \tag{1.3}$$

where j_0 is the smallest of the weights participating in the irreducible representation (j_0, c). If (j_0, c) is finite-dimensional, then the above sequence terminates at some maximum weight $j = j_0 + n$, $n = 1,2,3,\ldots$. Otherwise, this sequence is nonterminating. If $\{\xi_{j,m}\}$ is a (canonical) basis in M_j with $(2j+1)$ basis vecotrs corresponding to $m = -j, -j+1, \ldots, j-1, j$, then the whole space $B(j_0, c)$ is spanned by the set of all vectors $\xi_{j,m}$, where $j = j_0, j_0+1, j_0+2, \ldots,$

[a] Formerly Professor of Theoretical Physics, now retired from service.

 0022-2488/83/102397-07$02.50

and $m = -j, -j+1, \ldots, j-1, j$, and we shall refer to this basis as the ξ-basis in $B(j_0, c)$.

The six basic infinitesimal generators of a representation of SO(3,1) satisfy the commutation relations

$$[A_{(\alpha)}, A_{(\beta)}] = -[B_{(\alpha)}, B_{(\beta)}] = \epsilon_{\alpha\beta\gamma} A_{(\gamma)},$$
$$[A_{(\alpha)}, B_{(\beta)}] = \epsilon_{\alpha\beta\gamma} B_{(\gamma)}, \tag{1.4}$$

where the Greek suffixes α, β, γ range over 1, 2, 3, and $\epsilon_{\alpha\beta\gamma}$ is the usual antisymmetric permutation symbol. The operators $A_{(\alpha)} \equiv A_{(\beta\gamma)}$ (α, β, γ cyclic) correspond to spatial rotations in the x^β–x^γ plane and the $B_{(\alpha)} \equiv B_{(\alpha 4)}$ correspond to boosts in the x^α–x^4 plane of the Minkowski space–time spanned by coordinates $x^1 \equiv x$, $x^2 \equiv y$, $x^3 \equiv z$, and $x^4 \equiv ct$. The general solution of Eqs. (1.4), corresponding to an irreducible representation (j_0, c), where j_0 is a positive integral of half-odd-integral number and c is an arbitrary complex number as already mentioned, is given in the canonical ξ-basis $\{\xi_{j,m}\}$ by the formulas[9–11]

$$H_\pm \xi_{j,m} = \{(j \pm m + 1)(j \mp m)\}^{1/2} \xi_{j,m\pm 1}, \tag{1.5}$$

$$H_3 \xi_{j,m} = m \xi_{j,m},$$
$$\begin{aligned} F_\pm \xi_{j,m} = {} & \pm \{(j \mp m - 1)(j \mp m)\}^{1/2} C_j \xi_{j-1,m\pm 1} \\ & - \{(j \pm m + 1)(j \mp m)\}^{1/2} A_j \xi_{j,m\pm 1} \\ & \pm \{(j \pm m + 1)(j \pm m + 2)\}^{1/2} C_{j+1} \xi_{j+1,m\pm 1}, \end{aligned} \tag{1.6}$$

$$\begin{aligned} F_3 \xi_{j,m} = {} & \{(j-m)(j+m)\}^{1/2} C_j \xi_{j-1,m} - m A_j \xi_{j,m} \\ & - \{(j+m+1)(j-m+1)\}^{1/2} C_{j+1} \xi_{j+1,m}, \end{aligned}$$

where

$$A_j = i j_0 c / j(j+1), \quad A_0 \equiv ic, \tag{1.7}$$

$$C_j = \frac{i}{j} \left\{ \frac{(j^2 - j_0^2)(j^2 - c^2)}{4j^2 - 1} \right\}^{1/2}, \quad C_0 = 0, \tag{1.8}$$

$$H_\pm \equiv (iA_{(1)} \mp A_{(2)}), \quad H_3 \equiv iA_{(3)}, \tag{1.9}$$

and

$$F_\pm \equiv (iB_{(1)} \mp B_{(2)}), \quad F_3 \equiv iB_{(3)}. \tag{1.10}$$

If the numbers j_0 and c characterizing the irreducible representation (j_0, c) are not simultaneously integral or half-odd-integral, then (j_0, c) is *infinite-dimensional*, and the indices j and m take all the values in the ranges

$$\begin{aligned} & m = -j, -j+1, \ldots, j-1, j, \\ & j = j_0, j_0+1, j_0+2, \ldots . \end{aligned} \tag{1.11}$$

If however, $c^2 = (j_0 + n)^2$ for some positive integer n, so that j_0 and c are simultaneously integral or half-odd-integral, then the representation (j_0, c) is finite-dimensional with m and j taking the run of values

$$\begin{aligned} & m = -j, -j+1, \ldots, j-1, j, \\ & j = j_0, j_0+1, j_0+2, \ldots, |c| - 1. \end{aligned} \tag{1.12}$$

The dimension of such a finite-dimensional representation is given by the formula

$$\dim(j_0, c) = c^2 - j_0^2 = n(2j_0 + n). \tag{1.13}$$

The irreducible representation (j_0, c) is *unitary* if either

c is purely imaginary and j_0 is an arbitrary non-negative integral or half-odd-integral number (1.14a)

or

c is a real number in the interval $0 < |c| \leqslant 1$ and $j_0 = 0$. (1.14b)

The unitary irreducible representations corresponding to the Eq. (1.14a) are called the *principal series* of representations and those corresponding to Eq. (1.14b) are called the *complementary series* of representations. The representations characterized by $(0, c)$ and $(0, -c)$ are equivalent. With the exception of the trivial one-dimensional representation (0,1) all the unitary representations are of infinite dimension.

II. THE NATURE OF THE IRREDUCIBLE REPRESENTATION (j_0, c)

Let $D \equiv D(j_0, c)$ be a matrix belonging to the irreducible representation (j_0, c). We now proceed to prove that every D satisfies the relation

$$\tilde{D} G D = G, \tag{2.1}$$

for an appropriate "metric matrix" G, or equivalently that

$$\tilde{I} G + G I = 0, \tag{2.2}$$

where I is the infinitesimal transformation corresponding to D in the sense that $D = \exp(I)$. Since every such I may be expressed as a linear combination of $A_{(\alpha)}$ and $B_{(\alpha)}$, it is sufficient for this purpose to show that $A_{(\alpha)}$ and $B_{(\alpha)}$ and hence their linear combinations F's and H's given by Eqs. (1.5)–(1.10) satisfy Eq. (2.2) with some specified G. The matrix elements in the ξ-basis corresponding to the operators H_+, H_-, etc., may be easily identified from the relations (1.5) and (1.6), and we obtain

$$H_\pm(j', m'; j, m) = a(j, \mp m) \delta_{j',j} \delta_{m',m\pm 1}, \tag{2.3a}$$

$$H_3(j', m'; j, m) = m \delta_{j',j} \delta_{m',m}, \tag{2.3b}$$

$$\begin{aligned} F_\pm(j', m'; j, m) = {} & \delta_{m',m\pm 1} \{ \pm C_j b(j, \mp m) \delta_{j',j-1} \\ & - A_j a(j, \mp m) \delta_{j',j} \pm C_{j+1} b(j, \pm m + 2) \delta_{j',j+1} \}, \end{aligned} \tag{2.4a}$$

$$\begin{aligned} F_3(j', m'; j, m) = {} & \delta_{m',m} \{ C_j k(j, m) \delta_{j',j-1} - m A_j \delta_{j',j} \\ & - C_{j+1} k(j+1, m) \delta_{j',j+1} \}, \end{aligned} \tag{2.4b}$$

where we have introduced the notation

$$\begin{aligned} & a(j, m) \equiv \{(j+m)(j-m+1)\}^{1/2} = a(j, 1-m), \\ & b(j, m) \equiv \{(j+m)(j+m-1)\}^{1/2} = b(j+1, m-1) \\ & \qquad = b(j+2, m-2) \quad \text{etc.}, \\ & k(j, m) \equiv \{(j+m)(j-m)\}^{1/2} = k(j, -m). \end{aligned} \tag{2.5}$$

It may be observed that the rows and columns of the matrices H_+, H_-, etc., are labeled by pairs of indices so that, for example, $H_3(j', m'; j, m)$ is the element of the matrix H_3 occurring in the (j', m')th row and the (j, m)th column.[11] Further, in each pair (j, m), the second index m takes the run of values $-j, -j+1, \ldots, j-1, j$. Denoting similarly the matrix elements of G in the ξ-basis as $G(j', m'; j, m)$ and using

it in Eq. (2.2) with $I = H_3$, we get

$$(m' + m)G(j', m'; j, m) = 0, \tag{2.6}$$

which shows that all elements of G with $m' \neq -m$ must vanish so that

$$\begin{aligned} G(j', m'; j, m) &= \delta_{m', -m} G(j', m'; j, -m') \\ &= \delta_{m', -m} G(j', -m; j, m). \end{aligned} \tag{2.7}$$

To determine the coefficients of $\delta_{m', -m}$ above, we use the remaining operators H_+, F_3, etc., in Eq. (2.2). In doing so, we may observe that it is *sufficient* to determine $G(j', m'; j, -m')$ for all m' in its range $-j' \leqslant m' \leqslant j'$, if $j' < j$, and $G(j', -m; j, m)$ for all m in its range $-j \leqslant m \leqslant j$, if $j' > j$. On using Eq. (2.6) in Eq. (2.2) with $I = H_+$, we get

$$\begin{aligned} &\delta_{m', -m-1}\{a(j', -m')G(j', -m; j, m) \\ &\quad + a(j, -m)G(j', m'; j, -m')\} = 0. \end{aligned} \tag{2.8}$$

When $j' > j$, $m = j$, and $m' = -j - 1$, this yields

$$G(j', -j; j, j) = 0, \quad \text{for all } j' > j. \tag{2.9}$$

Also, when $j' > j$, it is not difficult to see that Eq. (2.8) yields the general formula

$$G(j', -m; j, m) = \frac{(-1)^{j-m} a(j, -m)a(j, -m-1)\cdots a(j, -j+1)G(j', -j; j, j)}{a(j', -m)a(j', -m-1)\cdots a(j', -j+1)}, \tag{2.10}$$

which, in view of Eq. (2.10), shows immediately that

$$G(j', -m; j, m) = 0, \quad j' > j \geqslant |m|. \tag{2.11}$$

Similarly, when $j' < j$, we obtain, by setting $m' = j'$ and $m = -j' - 1$ in Eq. (2.8),

$$G(j', j'; j, -j') = 0, \quad \text{for all } j' < j. \tag{2.12a}$$

Using this result in the general formula

$$G(j', m'; j, -m') = \frac{(-1)^{j'-m'} a(j', -m')a(j', -m'-1)\cdots a(j', -j'+1)G(j', j'; j, -j')}{a(j, -m')a(j, -m'-1)\cdots a(j, j'+1)}, \tag{2.12b}$$

which may be obtained from Eq. (2.8), we obtain

$$G(j', m'; j, -m') = 0, \quad j > j' \geqslant |m'|. \tag{2.13}$$

In view of the remarks following Eq. (2.7), it is easy to observe that Eqs. (2.10) and (2.13) together imply that the coefficients of $\delta_{m', -m}$ in Eq. (2.7) must vanish if $j' \neq j$, so that

$$G(j', m'; j, m) = \delta_{j', j}\delta_{m', -m} G(j, -m; j, m). \tag{2.14}$$

To determine the nonvanishing elements $G(j, -m; j, m)$, we use the relation

$$G(j, -m; j, m) + G(j, -m-1; j, m+1) = 0, \tag{2.15}$$

implied by Eq. (2.8) when $j = j'$, and

$$\begin{aligned} &\delta_{m', -m}\{C_{j+1} k(j+1, m)\delta_{j', j+1} + mA_j \delta_{j', j} \\ &\quad - C_j k(j, m)\delta_{j', j-1}\} \\ &\quad \times \{G(j, -m; j, m) - G(j', -m; j, m)\} = 0, \end{aligned}$$

which is obtained by using Eq. (2.14) in Eq. (2.2) with $I = F_3$ and the property Eq. (2.15) of $k(j, m)$. When $j' = j + 1$, the latter relation requires, for all $m = -m'$ lying in the range $-j \leqslant m \leqslant j$,

$$G(j, -m; j, m) - G(j+1, -m; j+1, m) = 0, \tag{2.16}$$

which shows that $G(j, -m; j, m)$ is independent of j. The general solution of the Eqs. (2.15) and (2.16) may be expressed as

$$G(j, -m; j, m) = Z_0(-1)^m, \tag{2.17}$$

where Z_0 is an arbitrary complex constant. It may be observed that only those elements of G appearing in Eq. (2.17) are nonzero. Substituting Eq. (2.17) in Eq. (2.14), we thus have finally

$$G(j', m'; j, m) = Z_0(-1)^m \delta_{j', j}\delta_{m', -m}. \tag{2.18a}$$

It is easy to check that this G satisfies Eq. (2.2) for the remaining infinitesimal operators H_-, F_+, and F_- also, so that Eq. (2.18a) provides the only solution of Eq. (2.2) for the metric G.

Taking the transpose of the matrix G, we observe that

$$\begin{aligned} \tilde{G}(j', m'; j, m) \equiv G(j, m; j', m') &= Z_0(-1)^{m'} \delta_{j, j'}\delta_{m, -m'} \\ &= Z_0(-1)^{m-2m}\delta_{j', j}\delta_{m', -m} \\ &= (-1)^{2m} G(j', m'; j, m) \end{aligned}$$

i.e.,

$$\tilde{G} = (-1)^{2j_0} G, \tag{2.19}$$

since $(-1)^{2m} = (-1)^{2j_0}$ for each m. This shows that G *is symmetric or skew-symmetric according as* j_0 *is integral or half-odd-integral.* Further, since the G matrix is determined only up to a multiplicative constant Z_0, we may choose it so that G is *real* both when j_0 is integral or half-odd-integral. For simplicity, we choose

$$\begin{aligned} &Z_0 = 1, \quad j_0 \text{ integral}, \\ &Z_0 = i, \quad j_0 \text{ half-odd-integral}. \end{aligned} \tag{2.18b}$$

The fact that the representation matrices $D(j_0, c)$ belonging to the irreducible representation (j_0, c), with j_0 half-odd-integral, preserve an antisymmetric G in the sense of Eq. (2.1) immediately establishes the following theorm.

Theorem 1: *All irreducible representations* (j_0, c) *of the proper Lorentz group* SO(3, 1), *corresponding to half-odd-integral* j_0 *are symplectic in nature.*

From this it follows in particular that the infinite-dimensional representations belonging to the principal series and having a half-odd-integral j_0 are symplectic in addition to being unitary.

In the ξ-basis, G has the block-diagonal structure

$$G = G_{j_0} + G_{j_0+1} + G_{j_0+2} + \cdots, \tag{2.20}$$

where the submatrix G_j has elements given by

$$G_j(m', m) = Z_0(-1)^m \delta_{m',-m}, \quad -j \leqslant m', m \leqslant j. \tag{2.21}$$

[The notation $G_j(m',m)$ stands for the element in the m'th row and mth column of the submatrix G_j.] It may be observed that when j_0 is integral, all the G_j are real-symmetric matrices of odd dimension $(2j+1)$ with elements equal to ± 1 appearing along the main antidiagonal. For example, we have for $j = 2$,

$$G_2 = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & -1 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & -1 & \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot & \cdot & \cdot \end{bmatrix}.$$

Similarly, when j_0 is half-odd-integral, all the G_j are real skew-symmetric matrices of even dimension $(2j+1)$ with elements equal to ± 1 appearing along the main antidiagonal. For example, we have for $j = 3/2$

$$G_{3/2} = \begin{bmatrix} \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & -1 & \cdot \\ \cdot & 1 & \cdot & \cdot \\ -1 & \cdot & \cdot & \cdot \end{bmatrix}.$$

So far, we have discussed G with respect to the canonical ξ-basis only. A change of basis in $B(j_0, c)$ evidently changes the matrix G, and with an appropriate change of basis G may be transformed to have a simpler (canonical) form. If $\eta = \tilde{T}\xi$ is a transformation from the ξ-basis to a new "η-basis" in $B(j_0, c)$, then the representation matrices $D = D(j_0, c)$ in the ξ-basis are transformed into matrices D' of the η-basis according to

$$D' = T^{-1} D T, \quad \eta = \tilde{T}\xi. \tag{2.22a}$$

Therefore, the relation $\tilde{D}GD = G$ implies

$$\tilde{T}^{-1}\tilde{D}'\tilde{T}GTD'T^{-1} = G,$$

so that we have

$$\tilde{D}'G'D' = G', \tag{2.22b}$$

where

$$G' = \tilde{T}GT \tag{2.22c}$$

is the metric matrix in the η-basis. Evidently, the symmetry (or skew-symmetry) of G is preserved under this transformation.

In the case of a real skew-symmetric G corresponding to symplectic representations, a T generated by a sequence of elementary transformations[12] may be used (if so desired) to throw G into the standard symplectic form[7]

$$G \to G' = \begin{pmatrix} \cdot & 1 \\ -1 & \cdot \end{pmatrix} + \begin{pmatrix} \cdot & 1 \\ -1 & \cdot \end{pmatrix} + \cdots. \tag{2.23}$$

On the other hand, in the case of the representations (j_0, c) which preserve a real symmetric G, i.e., when j_0 is integral, a new η-basis in $B(j_0, c)$ may always be chosen such that G is transformed into a unit matrix. To show this explicitly, we consider a class of basis transformations of the form

$$\eta_{j',m'} = \tilde{T}(j', m'; j, m)\xi_{j,m} = T(j, m; j', m')\xi_{j,m}, \quad (j, j', m, m' \text{ all integral}), \tag{2.24a}$$

where the transformation matrix T is given by

$$T(j, m; j', m') = p(m)(\delta_{m,m'} + i\delta_{m,-m'})\delta_{j,j'}. \tag{2.24b}$$

The complex coefficients $p(m)$ which appear here are defined for all m in $-j \leqslant m \leqslant j$, and are assumed to satisfy the relations

$$p(m)\,p(-m) = \tfrac{1}{2} i(-1)^{m+1}, \tag{2.25}$$

$$p(m)\,p^*(m) = \tfrac{1}{2}, \tag{2.26}$$

where $p^*(m)$ is the complex conjugate of $p(m)$. As a consequence of these two properties of $p(m)$, it follows that

$$p(m) = i(-1)^{m+1} p^*(-m),$$

and this in turn implies the useful relation

$$p(m)\,p^*(m-1) = -\{p(-m)\,p^*(-m+1)\}^*. \tag{2.27}$$

The defining equations (2.25) and (2.26) do not determine the coefficients $p(m)$ uniquely; in fact, a whole family of solutions exist for these equaitons. Of these, we may specially note the solution

$$p(m) = \tfrac{1}{2}(1-i)\exp(i|m|\pi/2), \tag{2.28}$$

for which $p(m)\,p^*(m-1) = \pm i/2$ according as $m > 0$ or $m \leqslant 0$.

It may be observed that the transformation given in Eq. (2.24) does not mix up the basis vectors belonging to different subspaces M_j of $B(j_0, c)$ because of the factor $\delta_{j,j'}$, and hence the matrix T may be expressed as the direct sum

$$T = T_{j_0} + T_{j_0+1} + T_{j_0+2} + \cdots, \tag{2.29a}$$

where any submatrix T_j has elements

$$T(j, m; j, m') \equiv T_j(m, m') = p(m)\{\delta_{m,m'} + i\delta_{m,-m'}\}, \quad -j \leqslant m \leqslant j. \tag{2.29b}$$

Using the property Eq. (2.25) of $p(m)$, it is easy to check that the T given by Eq. (2.24b) is *unitary*, i.e.,

$$T(j, m; j'', m'')T^\dagger(j'', m''; j', m') = T^\dagger(j, m; j'', m'')T(j'', m''; j', m') = E(j, m; j', m'), \tag{2.30}$$

where

$$T^\dagger(j, m; j', m') \equiv \tilde{T}^*(j, m; j'm') = T^*(j', m'; j, m) \tag{2.31}$$

and

$$E(j, m; j', m') \equiv \delta_{j,j'}\delta_{m,m'} \tag{2.32}$$

are the elements of the unit matrix E. Similarly, using Eq. (2.24b) and (2.22c), it may be checked that the metric matrix G' in the new η-basis is given by

$$G'(j, m; j', m') = E(j, m; j', m'), \quad G' = E. \tag{2.33}$$

Therefore, from Eq. (2.22b), it now follows that the representation matrices $D' \equiv D'(j_0, c)$ satisfy (for integral j_0 only)

$$\tilde{D}'D' = D'\tilde{D}' = E.$$

We thus have the following theorem:

Theorem 2: *All irreducible representations* (j_0, c) *of* SO(3, 1) *corresponding to integral values of* j_0 *are equivalent to complex-orthogonal representations.*

In particular, for the finite-dimensional representations of SO(3,1) corresponding to j_0 and c both being integral, the above result implies that the group of matrices $D(j_0, c)$ is a subgroup of the complex-orthogonal group SO(N, C) where $N = c^2 - j_0^2$ is the dimension of the representation. Here we may also recall the known result that for $N = 3$, i.e., $c = 2$, $j_0 = 1$, $D(1,2)$ is actually isomorphic to SO(3,C).

Secondly, we observe that since the transformation T is unitary, all those representations which were unitary in the canonical ξ-basis remain unitary in the η-basis also. Combined with the property of complex orthogonality in the η-basis, this means that all the unitary representations (j_0, c) of SO(3,1) corresponding to integral j_0, i.e., the entire complementary series and those of the principal series with integral j_0, are actually *real-orthogonal* in the η–basis. It may be observed that this special property is really an extension of the known result[13–15] that the (unitary) D^j representation of the subgroup SO(3) are equivalent to real-orthogonal ones if j is integral. Here it may also be appropriate to note that the remaining (unitary) D^j representations of SO(3), corresponding to half-odd-integral j, are symplectic by virtue of Theorem 1. Hence the matrices of a genuine spinorial representation D^j of the rotation group form a subgroup of the *unitary symplectic group*[6,8] Usp(2 j + 1).

We now proceed to construct explicitly the infinitesimal operators $A'_{(\alpha)}$ and $B'_{(\alpha)}$ corresponding to *integral* j_0 in the η-basis. For this purpose, it is convenient to introduce two new sets of coefficients $q_1(m)$ and $q_2(m)$, which are related to $p(m)$ as follows:

$$q_1(m) \equiv p(m)\,p^*(m-1) + p(-m)\,p^*(-m+1), \tag{2.34a}$$

$$q_2(m) \equiv p(m)\,p^*(m-1) - p(-m)\,p^*(-m+1). \tag{2.34b}$$

Using Eqs. (2.27) it is easy to check the following properties of $q_1(m)$ and $q_2(m)$-

$$q_1^*(\pm m) = -q_1(\pm m) = q_1(\mp m + 1), \tag{2.35a}$$

$$q_2^*(\pm m) = q_2(\pm m) = -q_2(\mp m + 1). \tag{2.35b}$$

Observe in particular that $q_1(m)$ are purely imaginary whereas the $q_2(m)$ are real.

Now, if I is any matrix in the ξ-basis and I' its image in the η-basis given by Eq. (2.24), then we have, on using Eq. (2.22c),

$$\begin{aligned} I'(j', m'; j, m) = {} & p^*(m')\,p(m) I(j', m'; j, m) \\ & + p^*(-m')\,p(-m) I(j', -m'; j, -m) \\ & + i\,p^*(m')\,p(-m) I(j', m'; j, -m) \\ & - i\,p^*(-m')\,p(m) I(j', -m'; j, m). \end{aligned} \tag{2.36}$$

Transforming the H's and the F's using the relation and then using Eqs. (1.9), (1.10), and (2.3)–(2.5), we get

$$A'_{(3)}(j', m'; j, m) = -m\delta_{j',j}\delta_{m',-m}, \tag{2.37a}$$

$$\begin{aligned} A'_{(1)}(j', m'; j, m) = {} & -\tfrac{1}{2}i\delta_{j',j}\{a(j, m)q_1(m)\delta_{m',m-1} \\ & + a(j, -m)q_1(-m)\delta_{m',m+1} \\ & - ia(j, m)q_2(m)\delta_{m',-m+1} \\ & + ia(j, -m)q_2(-m)\delta_{m',-m-1}\}, \end{aligned} \tag{2.37b}$$

$$\begin{aligned} A'_{(2)}(j', m'; j, m) = {} & \tfrac{1}{2}\delta_{j',j}\{a(j, m)q_2(m)\delta_{m',m-1} \\ & + a(j, -m)q_2(-m)\delta_{m',m+1} \\ & - ia(j, m)q_1(m)\delta_{m',-m+1} \\ & + ia(j, -m)q_1(-m)\delta_{m',-m-1}\}, \end{aligned} \tag{2.37c}$$

$$\begin{aligned} B'_{(3)}(j', m'; j, m) = {} & mA_j\delta_{j',j}\delta_{m',-m} \\ & - i\delta_{m',m}\{C_j k(j, m)\delta_{j',j-1} \\ & - C_{j+1}k(j+1, m)\delta_{j',j+1}\}, \end{aligned} \tag{2.38a}$$

$$\begin{aligned} B'_{(1)}(j', m'; j, m) = {} & \tfrac{1}{2}i\delta_{m',m-1}\{q_1(m)A_j a(j, m)\delta_{j',j} \\ & + q_2(m)[C_j b(j, m)\delta_{j',j-1} \\ & + C_{j+1}b(j, -m+2)\delta_{j',j+1}]\} \\ & + \tfrac{1}{2}i\delta_{m',m+1}\{q_1(-m)A_j a(j, -m)\delta_{j',j} \\ & + q_2(-m)[C_j b(j, -m)\delta_{j',j-1} \\ & + C_{j+1}b(j, m+2)\delta_{j',j+1}]\} \\ & + \tfrac{1}{2}\delta_{m',-m+1}\{q_2(m)A_j a(j, m)\delta_{j',j} \\ & + q_1(m)[C_j b(j, m)\delta_{j',j-1} \\ & + C_{j+1}b(j, -m+2)\delta_{j',j+1}]\} \\ & - \tfrac{1}{2}\delta_{m',-m-1}\{q_2(-m)A_j a(j, -m)\delta_{j',j} \\ & + q_1(-m)[C_j b(j, -m)\delta_{j',j-1} \\ & + C_{j+1}b(j, m+2)\delta_{j',j+1}]\}, \end{aligned} \tag{2.38b}$$

$$\begin{aligned} B'_{(2)}(j', m'; j, m) = {} & -\tfrac{1}{2}\delta_{m',m-1}\{q_2(m)A_j a(j, m)\delta_{j',j} \\ & + q_1(m)[C_j b(j, m)\delta_{j',j-1} \\ & + C_{j+1}b(j, -m+2)\delta_{j',j+1}]\} \\ & - \tfrac{1}{2}\delta_{m',m+1}\{q_2(-m)A_j a(j, -m)\delta_{j',j} \\ & + q_1(-m)[C_j b(j, -m)\delta_{j',j-1} \\ & + C_{j+1}b(j, m+2)\delta_{j',j+1}]\} \\ & + \tfrac{1}{2}i\delta_{m',-m+1}\{q_1(m)A_j a(j, m)\delta_{j',j} \\ & + q_2(m)[C_j b(j, m)\delta_{j',j-1} \\ & + C_{j+1}b(j, -m+2)\delta_{j',j+1}]\} \\ & - \tfrac{1}{2}i\delta_{m',-m-1}\{q_1(-m)A_j a(j, -m)\delta_{j',j} \\ & + q_2(-m)[C_j b(j, -m)\delta_{j',j-1} \\ & + C_{j+1}b(j, m+2)\delta_{j',j+1}]\}. \end{aligned} \tag{2.38c}$$

These matrices are seen to simplify considerably if we choose for $p(m)$ the special solution given in Eq. (2.28) as then $q_2(m) = 0$ and $q_1(m) = \pm i$ according as $m > 0$ or $m \leqslant 0$.

Using the fact that $iq_1(m)$ and $q_2(m)$ are real [for all solutions $p(m)$ of Eqs. (2.25) and (2.26)] and the properties of the real coefficients $a(j, m)$, $b(j, m)$, and $k(j, m)$ defined in Eqs. (2.5), it is easy to check directly, by using Eqs. (2.37) and

(2.38), the following properties of the $A'_{(\alpha)}$ and $B'_{(\alpha)}$:

(i) All the six generators $A'_{(\alpha)}$ and $B'_{(\alpha)}$ are skew-symmetric.

(ii) The three rotation generators $A'_{(\alpha)}$ are real.

(iii) In the case of unitary representations (with integral j_0 only) the boost generators $B'_{(\alpha)}$ are also real due to the fact that A_j are all real and C_j are pure-imaginary.

(iv) The $B'_{(\alpha)}$ are pure-imaginary in the case of finite-dimensional representations (only) as then the A_j are all pure-imaginary and the C_j are all real.

These properties provide an explicit verification of the special properties already obtained for the representations (j_0, c) corresponding to integral values of j_0. The property (i) shows that the matrices $D'(j_0, c)$ are complex-orthogonal when j_0 is integral. Properties (ii) and (iii) show that the unitary representations (j_0, c) are equivalent to real-orthogonal ones if j_0 is integral.

The property (iv) shows that all the (finite-dimensional unitary) integral-j_0 representations of the four-dimensional rotation group SO(4) are equivalent to real-orthogonal representations. This special property of the representations of SO(4) is implicit in the result that SO(4) is a direct product of two groups each of which is isomorphic to SO(3), and that the integral-j_0 representations of SO(3) are equivalent to real-orthogonal ones. To see that the properties (i) and (iv), of the generators of the finite-dimensional representations of SO(3,1), imply the above-mentioned property of SO(4), it is sufficient to observe that the *real skew-symmetric* matrices $\mathscr{A}_{(\alpha)}$ and $\mathscr{B}_{(\alpha)}$ defined by $\mathscr{A}_{(\alpha)} \equiv A'_{(\alpha)}$ and $\mathscr{B}_{(\alpha)} \equiv iB'_{(\alpha)}$ would be the matrices representing the corresponding infinitesimal generators of SO(4).

Lastly we examine the finite-dimensional irreducible representations of SO(3,1) with regard to the two special properties, namely (i) real-pseudo-orthogonality and (ii) pseudo-unitarity.

It is well known (see, for example, Lomont[16]) that among the finite-dimensional irreducible representations $D^{JJ'}$ of SO(3,1), where the positive numbers J and J' are either integral or half-odd-integral, only the D^{JJ} representations are equivalent to real ones. Equivalently, this means that the $(j_0 = 0, c = n)$ representations, where $n = (2J + 1)$ is a positive integer, are the only finite-dimensional irreducible representations of SO(3,1) which are equivalent to real ones. In addition to being real, the self-representation (0,2) is known to be pseudo-orthogonal with the signature (3,1). Hence it is of interest to find out whether the other real finite-dimensional representations (0,n) are also pseudo-orthogonal with some specific signature. We now show that *the n^2-dimensional (0, n) representation where $n = 1,2,3,\ldots$ are the only finite-dimensional irreducible real-pseudo-orthogonal representations of SO (3,1) and that the corresponding signature is (N_+, N_-), where $N_+ = n(n + 1)/2$ and $N_- = n(n - 1)/2$.*

We prove the result by transforming to a new ρ-basis from the η-basis by

$$\rho_{j',m'} = \tilde{S}(j', m'; j, m)\eta_{j,m} = S(j, m; j', m')\eta_{j,m}, \quad (2.39)$$

where the transformation matrix S is given by

$$S(j', m'; j, m) = (i)^{j+1}\delta_{j',j}\delta_{m',m}, \quad (2.40)$$

and the indices j, j', m, and m' are all integers (we are considering only those representations for which j_0 is an integer).

Using Eq. (2.40), we see that the infinitesimal generators I'' in the ρ-basis are related to the corresponding generators I' of the η-basis by $I'' = S^{-1}I'S$, i.e.,

$$I''(j', m'; j, m) = (i)^j(-i)^{j'}I'(j', m'; j, m). \quad (2.41)$$

In particular, for the (0, n) representations for which the coefficients C_j are all real (as $j < c = n$), $A_0 = in$, and $A_j = 0$ for all $j \neq 0$ [see Eq. (1.7) and (1.8)], a direct check using the above equation shows that all the six generators $A''_{(\alpha)}$ and $B''_{(\alpha)}$ are *real*. Next we observe [see Eq. (2.33)] that the metric G'' in the ρ-basis is given by $G'' = \tilde{S}G'S = \tilde{S}S$ so that

$$G''(j', m'; j, m) = (-1)^{j+1}\delta_{j',j}\delta_{m'm}, \quad (2.42)$$

which is a real matrix (as j is an integer) having ± 1 along the main diagonal. For the (0, n) representations, it is easy to see that G'' is the n^2- dimensional diagonal matrix

$$G'' = -E_0 + E_3 + (-E_5) + \cdots + (-1)^n E_{2n-1} = \sum_{j=0}^{n-1} + (-1)^{j+1}E_{2j+1}, \quad (2.43)$$

where E_k is the k-dimensional unit matrix. The signature (i.e., "the total number of plus ones minus the total number of minus ones") of this G'' is evidently $+n$ when n is *even* and $-n$ when n is odd. However, since a metric is determined only up to a constant multiplicative factor, we may take the metric preserved by the (0, n) representation to be $(-1)^n G''$ so that the signature is $+n$ irrespective of whether n is even or odd. The signature of this metric may also be expressed as (N_+, N_-), where $N_+ = n(n + 1)/2$ and $N_- = n(n - 1)/2$. It is thus proved that in the ρ-basis, the (0,n) representation is real pseudo-orthogonal with the signature (N_+, N_-). We may also note that the (0,n) representation matrices form a subgroup of the real group $SO(N_+, N_-, R)$.

The question of pseudo-unitarity of the representations of SO(3,1) has been completely solved by Gel'fand, *et al.*,[9] and the irreducible representation (j_0, c) is shown to be pseudo-unitary (in the sense that it admits an invariant Hermitian form) only under the conditions that either

c is purely imaingary and j_0 is any integer or half-odd-integer (2.44a)

or

c is real and $j_0 = 0$. (2.44b)

It may be observed that these conditions of pseudo-unitarity include the unitarity conditions, i.e., conditions under which (j_0, c) would admit a positive-definite Hermitian form, given in Eqs. (1.14a) and (1.14b). It is well known that the finite-dimensional irreducible representations of SO(3,1) [with the exception of the trivial (0,1) representation] are not unitary. However, the condition (2.44b) of pseudo-unitarity is satisfied by the (0,n) representations which are therefore evidently the *only* finite-dimensional pseudo-unitary representations of SO(3,1). In fact, in the ξ-basis, it is easy to check that all the six generators $A_{(\alpha)}$ and $B_{(\alpha)}$ of the (0,n) reresentation are "pseudo-skew-Hermitian" in the sense that they satisfy

$$I^{\dagger}\Gamma = -\Gamma I, \tag{2.45}$$

where Γ is the indefinite Hermitian metric

$$\Gamma(j', m'; j, m) = (-1)^{j+1}\delta_{j',j}\delta_{m',m}. \tag{2.46}$$

Thus in the ξ-basis the $(0,n)$ representation is seen to be pseudo-unitary with respect to the metric Γ. The signature of Γ is easily seen to be $(N_+, N_-) = (n(n+1)/2, n(n-1)/2)$. We may also note that the $(0,n)$ representation matrices form a subgroup of the pseudo-unitary group $SU(N_+, N_-, C)$ of dimension $(N_+ + N_-) = n^2$. However, since the $(0,n)$ representations are equivalent to real ones, the above-mentioned pseudo-unitarity reduces to the already discussed real-pseudo-orthogonality in the (real) ρ-basis.

The pseudo-unitarity conditions given in Eqs. (2.44a) and (2.44b) also show that none of the finite-dimensional symplectic representations (for which j_0 and c are simultaneously half-odd-integral) is of the unitary-symplectic type with some specific signature (see Wybourne[6] or Gilmore[8] for a definition of these groups).

ACKNOWLEDGMENTS

We are grateful to the referee for his suggestion that it would be of interest to pick out the real-pseudo-orthogonal and unitary-symplectic representations (if any) among the finite-dimensional representations of SO(3,1) and to determine the corresponding signatures. This has resulted in the last few paragraphs of the paper. One of us (B.S.N.) is grateful to the UGC (India) for the award of a fellowship under the Faculty Improvement Program and to Professor B. Sanjeevaiah for facilities and encouragement.

[1] J. Ehers and W. Kundt, in *Gravitation, An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962), p. 53.
[2] K. N. Srinivasa Rao, D. Saroja, and A. V. Gopala Rao, J. Math. Phys. **22**, 2167–79 (1981).
[3] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964), p. 27.
[4] H. Weyl, *The Classical Groups* (Princeton U. P., Princeton, N. J., 1939), p. 165.
[5] P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, Mass., 1965), p. 686.
[6] B. G. Wybourne, *Classical Groups for Physicists* (Wiley-Interscience, New York, 1974), p. 13.
[7] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1964), p. 405.
[8] R. Gilmore, *Lie Groups, Lie Algebras, and Some of Their Applications* (Wiley-Interscience, New York, 1974), pp. 43–46.
[9] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (Pergamon, New York, 1963), pp. 186–201.
[10] M. A. Naimark, *Linear Representations of the Lorentz Group* (Pergamon, New York, 1964), pp. 103–120.
[11] M. Carmeli, *Group Theory and General Relativity* (McGraw-Hill, New York, 1977), pp. 30–3.
[12] I. M. Gel'fand, *Lectures on Linear Algebra* (Interscience, New York, 1961), pp. 157–8.
[13] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics* (Wiley, New York, 1974), p. 338.
[14] E. P. Wigner, *Group Theory* (Academic, New York, 1959), pp. 285–9.
[15] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics—Theory and Applications* (Addison-Wesley, Reading, Mass., 1981), p. 316.
[16] J. S. Lomont, *Applications of Finite Groups* (Academic, New York, 1959), p. 321.

A class of finite-dimensional Lie algebras, the Casimir operators of which are not of finite type

F. Berdjis
Department of Physics, University of Illinois, P.O. Box 4348, Chicago, Illinois 60680

(Received 15 April 1983; accepted for publication 3 June 1983)

A class of complex finite-dimensional Lie algebras is constructed, the center of the universal enveloping algebra of each element of which is not finitely generated. For the construction of these Lie algebras, we use the counterexample of Nagata answering Hilbert's 14th problem in the negative.

PACS numbers: 02.20.Sv

A large number of papers investigate Casimir operators because of their usefulness in pure mathematics as well as in their applications in physics (see the references of Ref. 1). One is interested in knowing a minimal generating set of Casimir operators, and it was hoped that the cardinality of such a set would be finite for all Lie algebras of finite dimension[2,3] (see also Ref. 4 on Hilbert's 14th problem). In this paper, a whole class of finite-dimensional Lie algebras is constructed, the elements of which have the property that the set of their Casimir elements (so we call the elements of the center of the universal enveloping algebra of a Lie algebra) is not finitely generated. In Ref. 5 (pp. 165–166), one such Lie algebra is given in a remark without proof. This Lie algebra is identical with the lowest dimensional Lie algebra from our class of Lie algebras.

Let us now introduce our notation: $S(V)$ means the symmetric algebra over the vector space V, and $S_G(V)$ is the algebra of the invariants of a group G, acting as a transformation group in V, in $S(V)$, i.e., $S_G(V) := \{s \in S(V) | \tilde{g}(s) = s, g \in G\}$, where $\tilde{g}$ is the unique extension of $g \in G$ to an algebra automorphism of $S(V)$. For a Lie algebra L we denote by $S_{\mathrm{Int}}(L)$ the algebra of the invariants of the adjoint group $\mathrm{Int}(L)$ of L in $S(L)$. Finally $Z(L)$ denotes the set of the Casimir elements of L, i.e., $Z(L)$ is the center of the universal enveloping algebra $U(L)$ of L.

At first we prove a general lemma and then, using Nagata's counterexample to Hilbert's 14th problem, construct Lie algebras which fulfill the assumptions of the lemma. In the following is $\mathbb{K}$ the field of either real ($\mathbb{R}$) or complex ($\mathbb{C}$) numbers.

Lemma 1: Let V be a finite-dimensional vector space over $\mathbb{K}$, G a connected and nilpotent Lie subgroup of the group of the automorphisms of V, L the Lie algebra of G, $d: L \to \mathrm{End}(V)$ the (faithful) representation of L in the endomorphisms of V which is induced by the action of G on V, and $\hat{L} := V \oplus L$ ($\oplus$ means direct sum of vector spaces) the Lie algebra defined by the Lie product $[v_1 + x_1, v_2 + x_2]_{\hat{L}} := d(x_1)(v_2) - d(x_2)(v_1) + [x_1, x_2]_L, v_1, v_2 \in V, x_1, x_2 \in L$ (i.e., $\hat{L} = V \dot{+} L$, where $\dot{+}$ means semidirect sum of Lie algebras and V is made into an abelian ideal in $\hat{L}$). Then we have the following: If $S_G(V)$ is not of finite type, then $S_{\mathrm{Int}}(\hat{L})$ is not of finite type.

Proof: We will prove the lemma in three steps.

(a) Let $\tilde{S} := \{s \in S(V) | \tilde{d}_X(s) = 0, x \in L\}$, where $\tilde{d}_X$ is the unique extension of $d(x)$ as derivation on $S(V)$, and $\exp_G : d(L) \to G$ the exponential mapping of $d(L)$ into G. Let $s \in S_G(V)$. Then $\exp_G(td(x))(s) = s$, $x \in L$. By differentiation after t and setting $t = 0$, we get: $\tilde{d}_X(\tilde{s}) = 0$. Therefore $S_G(V) \subseteq \tilde{S}$. Now let $\tilde{s} \in \tilde{S}$. Then $\exp(\tilde{d}_x)(\tilde{s}) = \exp_G(d(x))(\tilde{s}) = \tilde{s}$. As G is connected and nilpotent, we have $\exp_G(d(L)) = G$ (for the proof, see Ref. 6, p. 229) and, therefore, $\tilde{S} \subseteq S_G(V)$. From the statement before, we have then $\tilde{S} = S_G(V)$.

(b) We want to show that $S_G(V) \subseteq S_{\mathrm{Int}}(\hat{L})$. Clearly, $S_G(V) \subseteq S(V) \subseteq S(\hat{L})$. Let $\mathrm{ad}_{\hat{L}}$ be the adjoint representation of $\hat{L}$ (in $\hat{L}$) and $\widetilde{\mathrm{ad}_{\hat{L}}(\hat{x})}$ be the unique extension of $ad_{\hat{L}}(\hat{x})$, $\hat{x} \in \hat{L}$, as derivation on $S(\hat{L})$ (it is well known that $S_{\mathrm{Int}}(\hat{L}) = \{s \in S(\hat{L}) | \widetilde{\mathrm{ad}_{\hat{L}}(\hat{x})}(s) = 0, \hat{x} \in \hat{L}\}$). Then we have: $\widetilde{\mathrm{ad}_{\hat{L}}(v)}(s) = 0$, $v \in V$, $s \in S(V)$, and because of $\mathrm{ad}_{\hat{L}}(x)(v) = [x, v]_{\hat{L}} = d(x)(v)$, $v \in V$, $x \in L$, it holds that $\widetilde{\mathrm{ad}_{\hat{L}}(x)} = \tilde{d}_X$. Therefore, we have for $\hat{x} = v + x \in \hat{L}$, $v \in V$, $x \in L$, and $s \in S_G(V)$: $\widetilde{\mathrm{ad}_{\hat{L}}(\hat{x})}(s) = \widetilde{\mathrm{ad}_{\hat{L}}(v)}(s) + \tilde{d}_X(s) = 0 + 0 = 0$ by the first part of (a). Therefore $S_G(V)$ is contained in $S_{\mathrm{Int}}(\hat{L})$ and it holds that $S_G(V) = S(V) \cap S_{\mathrm{Int}}(\hat{L})$.

(c) Let $\{v_1, \ldots, v_n\}$ and $\{x_1, \ldots, x_m\}$, $n, m \in \mathbb{N}$, be a basis of V and L, respectively, and $\{s_\iota\}_{\iota \in I}$ a generating set of $S_{\mathrm{Int}}(\hat{L})$. $s_\iota = s_\iota(v_1, \ldots, v_n, x_1, \ldots, x_m)$ is a polynomial in the basis $\{v_1, \ldots, v_n, x_1, \ldots, x_m\}$ of $\hat{L}$. Let us assume that card $(I) < \infty$. As with s_ι, every homogeneous part of it is an element of $S_{\mathrm{Int}}(\hat{L})$; we can conclude that there exists a minimal generating set Σ of $S_{\mathrm{Int}}(\hat{L})$ consisting of homogeneous elements with card $(\Sigma) < \infty$. We define $\Sigma_1 := \Sigma \cap S_G(V)$ and $\Sigma_2 := \Sigma \setminus \Sigma_1$ ($\setminus$ means the set-theoretical difference). Clearly $\Sigma_1 \cap \Sigma_2 = \emptyset$. Let $f \in \Sigma_2 \cap S(V)$. Then it follows for $x \in L$: $0 = \widetilde{\mathrm{ad}_{\hat{L}}(x)}(f) = \tilde{d}_X(f)$, and by (a), $f \in S_G(V)$, which is a contradiction. Therefore $\Sigma_2 \cap S(V) = \emptyset$, and we have for any $f \in \Sigma_2$: $f = f_1 + f_2$, where $f_1 \in S(V)$, f_1 is homogeneous, and f_2 is a homogeneous polynomial in $v_1, \ldots, v_n, x_1, \ldots, x_m$ such that in every monomial of it at minimum one x_k, $l \leqslant k \leqslant m$, appears as a factor. Now, let again $x \in L$. Then, $0 = \widetilde{\mathrm{ad}_{\hat{L}}(x)}(f) = \tilde{d}_X(f_1) + \widetilde{\mathrm{ad}_{\hat{L}}(x)}(f_2)$, where $\tilde{d}_X(f_1) \in S(V)$ and $\widetilde{\mathrm{ad}_{\hat{L}}(x)}(f_2)$ is either zero or a polynomial in $v_1, \ldots, v_n, x_1, \ldots, x_m$ such that again in every monomial of it at minimum one x_k appears as a factor. From this, it follows that $\tilde{d}_X(f_1) = 0$ and $\widetilde{\mathrm{ad}_{\hat{L}}(x)}(f_2) = 0$. Therefore we have, by (a) and (b), $f_1 \in S_G(V) \subseteq S_{\mathrm{Int}}(\hat{L})$ and $f_2 = f - f_1 \in S_{\mathrm{Int}}(\hat{L})$. Now, we construct a new integrity basis for $S_{\mathrm{Int}}(\hat{L})$.

 0022-2488/83/102404-03$02.50

$\hat{\Sigma} := \Sigma_1 \cup \Sigma_{2,1} \cup \Sigma_{2,2}$, where $\Sigma_{2,1} := \{f_1 | f_1 + f_2 = f, f \in \Sigma_2\}$, $\Sigma_{2,2} := \{f_2 | f_1 + f_2 = f, f \in \Sigma_2\}$, and f_1 and f_2 are the summands of $f \in \Sigma_2$ as before. The sets Σ_1, $\Sigma_{2,1}$, and $\Sigma_{2,2}$ have all cardinality less than infinity and it is clear that all elements of $S_G(V)$ must be generated by $\Sigma_1 \cup \Sigma_{2,1}$ because algebraic combinations with elements $f_2 \in \Sigma_{2,2}$ cannot be elements of $S(V)$ (up to constant elements). But this is a contradiction to the supposition that $S_G(V)$ is not of finite type. Therefore $S_{\text{Int}}(\hat{L})$ is not finitely generated. ■

Now we introduce the groups which were constructed by M. Nagata.[7]

Definition: Let $r \in \mathbb{N}$, $r \geqslant 4$, and a_{ij}, $1 \leqslant i \leqslant 3$, $1 \leqslant j \leqslant r$, elements of $\mathbb{K}$ which are algebraically independent over the field of rational numbers $\mathbb{Q}$. Then we define the group $G := \{\sigma \in \mathrm{GL}(2r, \mathbb{K}) |$ the entries σ_{ij} of σ fullfill the following algebraic equations (1) to (8):

$$\sigma_{ij} = 0, \quad j > i \tag{1}$$

$$\sigma_{ij} = 0, \quad 1 \leqslant j < i \leqslant r \tag{2}$$

$$\sigma_{ij} = 0, \quad r+1 \leqslant j < i \leqslant 2r \tag{3}$$

$$\sigma_{ij} = 0, \quad 1 \leqslant j \leqslant r, \quad i > j + r \tag{4}$$

$$\sigma_{i+r,j} = 0, \quad 1 \leqslant i < j \leqslant r \tag{5}$$

$$\prod_{i=1}^{r} \sigma_{ii} = 1, \tag{6}$$

$$\sigma_{ii} - \sigma_{i+r,i+r} = 0, \quad 1 \leqslant i \leqslant r \tag{7}$$

$$\sum_{i=1}^{r} a_{ki} \prod_{\substack{l=1 \\ l \neq i}}^{r} \sigma_{ll} \sigma_{i+r,i} = 0, \quad 1 \leqslant k \leqslant 3\}. \tag{8}$$

Remark: Let $(C_1, C_2, \ldots, C_r) \in \mathbb{K}^r$ with $\Pi_{i=1}^r C_i = 1$ and $(b_1, b_2, \ldots, b_r) \in \mathbb{K}^r$ with $\sum_{j=1}^{r} a_{ij} b_j = 0$, $1 \leqslant i \leqslant 3$. Then $\sigma \in G$ is of following form:

$$\sigma = \left[\begin{array}{ccc|ccc} c_1 & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & c_r & & & \\ \hline c_1 b_1 & & 0 & c_1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & c_r b_r & 0 & & c_r \end{array}\right]$$

and we have

$$\sigma^{-1} = \left[\begin{array}{ccc|ccc} \frac{1}{c_1} & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \frac{1}{c_r} & & & \\ \hline \frac{-b_1}{c_1} & & 0 & \frac{1}{c_1} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \frac{-b_r}{c_r} & 0 & & \frac{1}{c_r} \end{array}\right]$$

The composition of $\sigma, \sigma'^{-1} \in G$ is

$$\sigma \cdot \sigma'^{-1} = \left[\begin{array}{ccc|ccc} \frac{c_1}{c_1'} & & 0 & & & \\ & \ddots & & & 0 & \\ 0 & & \frac{c_r}{c_r'} & & & \\ \hline \frac{c_1}{c_1'}(b_1 - b_1') & & 0 & \frac{c_1}{c_1'} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \frac{c_r}{c_r'}(b_r - b_r') & 0 & & \frac{c_r}{c_r'} \end{array}\right]$$

the entries of which fulfill, of course, again Eqs. (1)–(8) in the definition.

Lemma 2: Let G as in the definition. Then we have the following.

(a) G is an abelian algebraic subgroup of $\mathrm{GL}(2r, \mathbb{K})$.

(b) G is a Lie group.

(c) For $\mathbb{K} = \mathbb{C}$, G is connected.

Proof: (a) By Eqs. (1)–(8) in the definition, G is an algebraic group. The commutativity of G is obvious.

(b) Let G carry the induced topology of $\mathrm{GL}(2r, \mathbb{K})$. Then G is as an algebraic subgroup of $\mathrm{GL}(2r, \mathbb{K})$ a closed subset of $\mathrm{GL}(2r, \mathbb{K})$. In Ref. 6 (p. 105), it is proved that such a group is a Lie group.

(c) We prove in this case that G is arcwise connected, from which it follows that G is connected. G is topologically isomorphic to $\Omega \times \mathbb{C}^{r-3}$, where $\Omega := \{(C_1, \ldots, C_r) \in \mathbb{C}^r | \Pi_{i=1}^r C_i = 1\}$. Therefore it suffices to prove that Ω is arcwise connected. Let $(C_1, \ldots, C_r)$ and $(C_1', \ldots, C_r')$ be two points of Ω. We construct a path $\omega(t)$ in Ω with $\omega(0) = (C_1, \ldots, C_r)$ and $\omega(1) = (C_1', \ldots, C_r')$. In polar coordinates we have $(C_1, C_2, \ldots, C_r) = (R_1, \ldots, R_r, \phi_1, \ldots, \phi_r)$, $(C_1', \ldots, C_r') = (R_1', \ldots, R_r', \phi_1', \ldots, \phi_r')$ with $R_1 \cdots R_r = R_1' \cdots R_r' = 1$, and $\Sigma_{i=1}^r \phi_i = \Sigma_{i=1}^r \phi_i' = O \bmod 2\pi$.

Now let $\omega(t) := (R_1(t), \ldots, R_r(t), \phi_1(t), \ldots, \phi_r(t))$ with

$$R_i(t) := (1 + t(R_i'/R_i - 1))\, R_i \text{ for } 1 \leqslant i \leqslant r-1,$$

$$R_r(t) := \frac{1}{\prod_{i=1}^{r-1} (1 + t(R_i'/R_i - 1))} R_r,$$

$$\phi_i(t) := \phi_i + t(\phi_i' - \phi_i) \text{ for } 1 \leqslant i \leqslant r-1,$$

and

$$\phi_r(t) := \phi_r - t\left(\sum_{i=1}^{r-1} \phi_i' - \sum_{i=1}^{r-1} \phi_i\right).$$

It is easily seen that the denominator in the definition of $R_r(t)$ does not vanish for $t \in [0, 1]$ and that $\Pi_{i=1}^r R_i(t) = 1$ and $\Sigma_{i=1}^r \phi_i(t) = O \bmod 2\pi$, i.e., $\omega(t) \in \Omega$ with initial point $(C_1, \ldots, C_r)$ and endpoint $(C_1', \ldots, C_r')$. Therefore Ω is connected and our proof is complete.

For the following theorem, see Ref. 7.

Theorem: Let $r = s^2$, $s \in \mathbb{N}$, $s \geqslant 4$, V a $2r$-dimensional

vector space over $\mathbb{K}$, and G as in the definition. Then $S_G(V)$ is not of finite type.

And now our statement:

Lemma 3: Let $\mathbb{K} = \mathbb{C}$, r, G, and V as in the theorem, L the Lie algebra of G [cf. Lemma 2b], and $\hat{L}$ as in Lemma 1. Then $Z(\hat{L})$ is not of finite type.

Proof: We have, by Lemma 1 and 2: $S_{\text{Int}}(\hat{L})$ is not of finite type, because G as abelian group is nilpotent. In Ref. 1, Lemma 1 is proved that a set $\Sigma \subset S_{\text{Int}}(\hat{L})$ is a generating set of $S_{\text{Int}}(\hat{L})$ if and only if $\lambda(\Sigma)$ is a generating set of $Z(\hat{L})$, where λ is a vector space isomorphism of $S(\hat{L})$ onto $U(\hat{L})$ with $\lambda(S_{\text{Int}}(\hat{L})) = Z(\hat{L})$ (the proof is given there for a finite set Σ, but it is obvious by inspection that the proof remains valid for arbitrary cardinality of Σ).

[1]F. Berdjis, J. Math. Phys. **22**, 1851 (1981).
[2]E. G. Beltrametti and A. Blasi, Phys. Lett. **20**, 62 (1966).
[3]U. Mutze, Z. Phys. **229**, 224 (1969).
[4]D. Mumford, Proc. Sympos. Pure Math. **28**, 431 (1976).
[5]J. Dixmier, *Enveloping Algebras* (Akademie-Verlag, Berlin, 1977).
[6]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[7]M. Nagata, Am. J. Math. **81**, 766 (1959).

Dirac general covariance and tetrads. I. Clifford and Lie bundles and torsion

John R. Urani
Physics Department, University of Missouri–Kansas City, Kansas City, Missouri 64110

George P. Barker
Mathematics Department, University of Missouri–Kansas City, Kansas City, Missouri 64110

(Received 4 February 1983; accepted for publication 26 May 1983)

The isomorphic map of Clifford and Lie bundles to arbitrary coordinate atlases using a global orthonormal tetrad field on a parallelizable space-time is used to construct a fully covariant Dirac spinor theory. The Klein–Gordon equation exhibits a natural spin-torsion coupling of the Einstein–Cartan form, the torsion coming from the tetrad field. The tetrad connection coefficients are explicitly derived in addition to their relationship to the usual Levi–Cività coefficients. Various topological conditions for vanishing torsion are given. The Dirac and adjoint Dirac equations are obtained from a simple Lagrangian and the structure of the adjoint equation is discussed.

PACS numbers: 02.40.Re

I. INTRODUCTION

Global geometrical methods in general relativity have, over the past twenty years, resulted in some spectacular advances. The singularity theorems[1] are, of course, a prime example of global structure implied by a physical theory.

Very important results on the existence of a spin structure on a space-time iff the underlying manifold admits a global tetrad field are found in the seminal works of Geroch.[2] Parallelizable space-times (those which admit a global tetrad) then also admit a global orthonormal tetrad, orthonormal via the Lorentzian metric $g(x)$. Physically, this corresponds to choosing a Lorentz frame at each x in M relative to which the metric has components $\eta_{ab} = (- \; - \; - \; +)$. Mathematically, we may describe this choice of a Lorentz frame at each x in M as a smooth cross section of the orthonormal frame bundle, $O(M)$. Important recent applications of orthonormal (on) tetrad[3–7] techniques have been reported. In particular, Weinberg[3] gives an excellent discussion of the physical principles underlying the tetrad approach.

Of importance to this paper, too, are the early results of Weyl and of Schrödinger on generalizations of the Dirac equation.[8]

In this paper, we consider the full implications of the tetrad map of Clifford and Lie algebra elements to arbitrary coordinate charts on a space-time (M, g). Since we assume the existence of a spin structure[2] on the space-time, we then may take the on tetrad on the manifold to be globally defined. We assume at least C^3 differentiability for $g(x)$ and at least C^2 for the tetrad field.

It is shown that by using the local on tetrad to express Dirac (Clifford) algebra units γ^a relative to a local coordinate chart, a Lie algebra isomorphism of the proper Lorentz group $SO^+(3, 1)$ is obtained. The isomorphic nature of the tetrad map and the general scalar nature of the Dirac wave amplitude, $\psi(x)$, under changes of manifold coordinate atlas, lead to a generally covariant Dirac theory whose Klein–Gordon equation exhibits torsion in the classic Einstein–Cartan sense.[9–11] The torsion arises naturally from a connection on the space-time induced by the tetrad field cross section of $O(M)$. In the subsequent paper, acceleration covariance of the theory is demonstrated when a smooth change in tetrad field, representing a pointwise change of Lorentz frame, is induced by a global timelike congruence of observer trajectories.[12,13]

II. THE TETRAD MAP

We let (M, g) be a parallelizable space-time[14] and choose a global on tetrad field, $K_a(x), a = 1, 2, 3, 4$, for M. For each x in M we then have $g(K_a, K_b) = \eta_{ab} = (- \; - \; - \; +)$. This tetrad field then specifies a Lorentz frame at each x in M which is, by definition, a basis for the tangent space T_x at each x in M. The tetrad 1-forms $\omega^a(x)$ dual to the tetrad fields at each x form a basis for the cotangent space T_x^* for each x in M. We shall denote the tetrad on 1-forms by $K^{-1a}(x)$. We then have[3] $\omega^a(x) = K^{-1a}(x) = \eta^{ab} g(K_b(x), \,)$.

Arbitrary tensors are then expressible in components relative to the tetrad field and 1–form bases by suitable tensor products and projections.[7] For example, an arbitrary (1, 0) tensor $T(x)$ is expressed in tetrad components via $T_a(x) = g(K_a(x), T(x))$ which is, in a local coordinate chart, expressed as $T_a(x) = g_{\mu\nu}(x) K^{\mu}_{a}(x) T^{\nu}(x)$.

Similarly, since an on tetrad at x is a Lorentz frame at x, strictly Lorentzian quantities which are defined with respect to these on frames may be expressed relative to a local coordinate chart via the use of the tetrad. Formally, such Lorentzian quantities transform under a change of on tetrad at x via the full Lorentz group $O(3, 1)$, the structure group of the fiber over x of on frames in the on frame bundle $O(M)$. This on frame bundle is trivial in the case of the parallelizable manifold we are considering.[14]

A Lorentzian quantity of great interest is the set of Dirac elements $\gamma^a, a = 1, 2, 3, 4$ which are Lorentz invariant. We find that the Dirac units γ^a may be usefully expressed in local coordinates via

$$\gamma^{\mu}(x) = K^{\mu}_{a}(x) \gamma^a,$$

giving an isomorphic Clifford anticommutator

$$[\gamma^{\mu}(x), \gamma^{\nu}(x)]_+ = 2g^{\mu\nu}(x) I, \tag{2.1}$$

where $\eta(- \; - \; - \; +)$ is simply replaced by g.

 0022-2488/83/102407-04$02.50

We shall state this and subsequent results in terms of bundles. We have, so far, considered T_M, the tangent bundle over M, and, implicitly, the bundles $T^r_s(M)$ of (r, s) tensors over M. The on tetrad field is a cross section of $O^+(M)$, the oriented on frame bundle over M, a principle fiber bundle with structure group $L_P \sim \mathrm{SO}^+(3, 1)$, the proper Lorentz group. We have also considered the Clifford bundle[15,16] over M defined by assigning the full Dirac algebra to each point $x \epsilon M$. In general coordinate charts, the isomorphism between the flat R^4 Dirac algebra and Eq. (2.1) results from the fact that the matrix $K^{\mu}_{a}(x) \epsilon GL(4, R)$ for each $x \epsilon M$.

We consider the $O^+(M)$ bundle again and note that the global trivialization of $O^+(M)$ (since M is parallelizable) allows the identification of a trivial Lie bundle over M. This is simply the bundle over M of $\mathrm{SO}^+(3, 1)$ fiber (vertical) tangent spaces at the identity element of each fiber. Since the frame bundle is trivial, we may choose the usual $\mathrm{SO}^+(3, 1)$ generators[17], $M^{ab} = -M^{ba}$, at each $x \epsilon M$, to generate the Lie algebra of $\mathrm{SO}^+(3, 1)$. As with the Clifford elements γ^a we may tetrad map to a local coordinate chart to obtain

$$M^{\mu\nu}(x) = K^{\mu}_{a}(x) K^{\nu}_{b}(x) M^{ab}. \tag{2.2}$$

The local-coordinate generators of Eq. (2.2), in fact, satisfy a commutator Lie algebra isomorphic to the flat R^4 Lie algebra of L_P but with $\eta = (- \; - \; - \; +)$ replaced by $g^{\mu\nu}(x)$ for each $x \epsilon M$. The isomorphism is again due to the nonsingular nature of $K^{\mu}_{a}(x)$.

The existence of a spin structure on M is based on the existence of a 2-1 bundle morphism $h: S(M) \to O^+(M)$ where $S(M)$ is the bundle of "spin frames" [$\mathbb{C}^2$ bases over M with structure group SL $(2, \mathbb{C})$]. We now consider the Dirac spin-frame bundle $D(M)$ of bases for $\mathbb{C}^4$ with structure group $\mathrm{SL}(2, C) \oplus \mathrm{SL}(2, C)$ represented by $D^{(1/2,0)} \oplus D^{(0,1/2)}$. The Lie algebra generators in this trivial bundle are proportional to $\sigma^{ab} = (-i/2)[\gamma^a, \gamma^b,]$ obtained from the Clifford elements.[17]

The tetrad map of the σ^{ab} to local coordinates, via

$$\sigma^{\mu\nu}(x) = K^{\mu}_{a}(x) K^{\nu}_{b}(x) \sigma^{ab}, \tag{2.3}$$

results in a Lie algebra isomorphism identical to that of the tetrad map of the self-representation generators of $\mathrm{SO}^+(3.1)$ given by Eq. (2.2).

The essential result is that the flat-R^4 Clifford and Lie algebra elements, when expressed via local tetrads relative to tangent space bases derived from local coordinate charts, are general tensor components. For example, $\gamma^{\mu}(x)\partial_{\mu} \equiv D(x)$ is a vector field operator in the general sense. Similarly, $\sigma^{\mu\nu}(x)$ are (2, 0) antisymmetric general tensor components for each $x \epsilon M$.

We define the Dirac spinor field[18] ψ as a map $\psi: M \to \mathbb{C}^4$ by $x \to \psi(x)$ or, more elegantly, as a smooth cross section of the vector bundle $(O(M) X_G \mathbb{C}^4)$ where $G = D^{(1/2,0)} \oplus D^{(0,1/2)}$.

Under a local change of coordinate charts (containing $x \epsilon M$), the image $\psi(x) \epsilon \mathbb{C}^4$ is unchanged and $\psi(x)$ is a general scalar field[3] on M. But, under changes of local on tetrad at x, the $\mathrm{SO}^+(3, 1) - \mathrm{SL}(2, C)$ homomorphism induces $D^{(1/2,0)} \oplus D^{(0,1/2)}$ coordinate maps of $\psi(x)$. Hence $\psi(x)$ is a spinor[3] under changes of local Lorentz frame (on tetrad) at $x \epsilon M$.

The Dirac equation may then be written, given a choice of tetrad field, $K_a(x)$, $a = 1, 2, 3, 4$, on M, as

$$(\gamma^{\mu}(x)\partial_{\mu} + \mathrm{im})\psi(x) = R(x). \tag{2.4}$$

Unless a change in local tetrad is made, no connection coefficients are needed.[3] The operator $D(x) \equiv \gamma^{\mu}(x)\partial_{\mu}$ is a vector field operator and the Dirac equation is a general scalar and a Lorentz spinor.

The Klein–Gordon equation is constructed, as usual, by operating with $(D(x) - \mathrm{im})$ to yield

$$[\gamma^{\mu}(x)\gamma^{\nu}(x)\partial_{\mu}\partial_{\nu} + \gamma^{\mu}(x)(\partial_{\mu}\gamma^{\nu}(x))\partial_{\nu} + m^2]\psi(x) = \mathrm{S(x)}, \tag{2.5}$$

where $S(x) = (D(x) - \mathrm{im}) R(x)$.

We next take symmetric and antisymmetric parts of the first two terms in Eq. (2.5).

Assuming that $\psi(x)$ is of class $C^k(M)$, $k \geqslant 2$, only the symmetric part of $\gamma^{\mu}(x)\gamma^{\nu}(x)\partial_{\mu}\partial_{\nu}\psi$ contributes in the form $g^{\mu\nu}(x)\partial_{\mu}\partial_{\nu}\psi$. Some detail is necessary in treating the second term $\gamma^{\nu}(\partial_{\mu}\gamma^{\nu})\partial_{\nu}\psi$ of Eq. (2.5).

We write, using the orthonormality of the tetrad fields,

$$\begin{aligned}
\gamma^{\mu}(x)\partial_{\mu}\gamma^{\nu}(x) &= \gamma^{\mu}(x)\partial_{\mu}K^{\nu}_{a}(x)\gamma^{a} \\
&= \gamma^{\mu}(x)(\partial_{\mu}K^{\nu}_{b}(x))K^{-1b}_{a}(x)K^{a}_{a}(x)\gamma^{a} \\
&= -\gamma^{\mu}(x)[K^{\nu}_{b}(x)\partial_{\mu}K^{-1b}_{a}(x)]\gamma^{a}(x) \\
&= -\gamma^{\mu}(x)[\tilde{\Gamma}^{\nu}_{\mu a}(x)]\gamma^{a}(x),
\end{aligned} \tag{2.6}$$

where $K^{-1b}(x)$ are the dual tetrad 1-forms. The affine connection coefficients, $\tilde{\Gamma} = K\partial K^{-1}$, define the connection[19,20] associated with the given tetrad field cross section of $O^+(M)$. This connection defines parallel transport relative to the horizontal subspaces of $T_{O^+(M)}$ defined by the tetrad field. This is discussed in detail following further manipulations.

We finally write the second term of Eq. (2.5) as

$$-\gamma^{\mu}(x)\gamma^{a}(x)\tilde{\Gamma}^{\nu}_{\mu a}(x) = -g^{\mu a}(x)\tilde{\Gamma}^{\nu}_{(\mu a)}(x) - i\sigma^{\mu a}(x)\tilde{\Upsilon}^{\nu}_{\mu a}(x),$$

by taking symmetric and antisymmetric parts.

Here, $\tilde{\Upsilon}^{\nu}_{\mu a} = \tilde{\Gamma}^{\nu}_{\mu a} - \tilde{\Gamma}^{\nu}_{a\mu}$ is the local coordinate form[21] of the torsion of the affine (tetrad) connection. The Klein–Gordon equation is, then,

$$[g^{\mu\nu}(\partial_{\mu}\partial_{\nu} - \tilde{\Gamma}^{\alpha}_{(\mu\nu)}\partial_{\alpha}) - i\sigma^{\mu\nu}(x)\tilde{\Upsilon}^{\alpha}_{\mu\nu}(x)\partial_{\alpha} + m^2]\psi(x) = S(x).$$

We see that this generally covariant Klein–Gordon equation contains the tetrad affine connection rather than the Levi-Cività metric connection. However, we show in Sec. III that $g^{\mu\nu}\Gamma^{\alpha}_{\mu\nu} = g^{\mu\nu}\tilde{\Gamma}^{\alpha}_{\mu\nu}$ and, hence, Eq. (2.7) is the usual Klein–Gordon equation with a torsion term from the tetrad connection. The torsion of the tetrad connection, here coupled naturally to the spin operator $\sigma(x)$ in the classic Einstein–Cartan form,[9] is an intrinsic structure on a parallelizable manifold with a global tetrad field. The vanishing of the torsion requires a rather stringent condition on the cohomology structure of the space-time. This is described in the next section which considers details of the tetrad connection coefficients.

III. THE TETRAD CONNECTION

The tetrad connection coefficients, $\tilde{\Gamma}$, relative to a given coordinate atlas, are clearly associated with parallel transport relative to the horizontal subspaces[19] of $T_{O^+(M)}$ defined by the given tetrad field cross section of $O^+(M)$.

This is immediately seen by a direct manipulation in local coordinates since, using the usual tetrad orthonormality relations,[3] we obtain

$$\partial_\alpha K^\mu_a = \partial_\alpha K^\mu_b \delta^b_a = \partial_\alpha K^\mu_b (K^{-1b}_\nu K^\nu_a) = -K^\mu_b (\partial_\alpha K^{-1b}_\nu) K^\nu_a.$$

This may be written as

$$\partial_\alpha K^\mu_a + \tilde{\Gamma}^\mu_{\alpha\nu} K^\nu_a = 0, \tag{3.1}$$

where the connection coefficients $\tilde{\Gamma}$ are defined as $\tilde{\Gamma} = K_b \partial K^{-1b}$. Now, Eq. (3.1) may be written in invariant form, for all a, b values, as

$$\tilde{\nabla}_{K_b} K_a = 0, \tag{3.2}$$

which expresses the horizontal parallel transport of the tetrad fields relative to their own connection.

It is easily shown that $\gamma^\mu(x)$ [Eq. (2.6)], $\sigma^{\mu\nu}(x)$ and any other SO(3, 1) (Lorentz) tensors mapped to the local coordinate basis for T_x by the local tetrad map are general GL(4, R) tensors which parallel transport via $\tilde{\nabla}$, the tetrad connection.

It is also informative to express the Levi-Cività metric connection Γ in terms of the tetrad connection coefficients $\tilde{\Gamma}$. In local coordinates, using the usual expression[3] for the metric connection along with the tetrad orthonormality relation $\eta^{ab} K^\mu_a K^\nu_b = g^{\mu\nu}$ we obtain

$$\Gamma^\mu_{\alpha\beta} = g^{\mu\lambda} g_{\theta(\alpha} \tilde{\Upsilon}^\theta_{\beta)\lambda} + \tilde{\Gamma}^\mu_{(\alpha\beta)}, \tag{3.3}$$

where ( ) brackets denote symmetrized indices.

The combination of Eqs. (3.1), (3.3), and the definition of torsion in terms of $\tilde{\Gamma}$, in fact, constitute a restatement, in terms of the tetrad field's affine connection, of the "vierbein postulate" commonly used in quantum gravity formulations.[5,22] Several useful contractions of Eq. (3.3) are $g^{\mu\nu}\tilde{\Gamma}^\alpha_{\mu\nu} = g^{\mu\nu}\Gamma^\alpha_{\mu\nu}$, $\tilde{\Gamma}^\alpha_{\mu\alpha} = \Gamma^\alpha_{\mu\alpha}$, $\tilde{\Upsilon}^\mu_{\mu\alpha} = -\tilde{\Upsilon}^\mu_{\alpha\mu} = 0$ and $g^{\mu\nu}\tilde{\Upsilon}^\alpha_{\mu\nu} = 0$.

In Sec. IV, the orthonormal congruence property of the tetrad fields is used to obtain the contracted relation $\Gamma^\mu_{\mu\alpha} = \tilde{\Gamma}^\mu_{\mu\alpha}$, a great simplification.

From Eq. (3.3), it is clear that when the torsion vanishes and hence $\tilde{\Gamma}$ is symmetric, then $\Gamma = \tilde{\Gamma}$. This presents several interesting situations.

We note that

$$\tilde{\Upsilon}^\mu_{\alpha\lambda} = K^\mu_b (\partial_\alpha K^{-1b}_\lambda - \partial_\lambda K^{-1b}_\alpha) = (K_b \otimes dK^{-1b})^\mu_{\alpha\lambda}.$$

The torsion vanishes iff all tetrad 1-forms K^{-1b} are closed. We consider several topological possibilities.

For example, if each tetrad 1-form is exact, then there exists a global chart for M such that $K^{-1a} = dy^a$, $a = 1, 2, 3, 4$ on this chart. Relative to this chart, g may be diagonalized to η for each $x \in M$. Hence the space-time is effectively Lorentzian R^4.

We note that the space-time is time orientable and spacially orientable if it is parallelizable and, hence, we consider the case where $K_4(x)$, the fourth tetrad field, is defined by the global time function,[21] $t(x)$. In this particular case we have the result that $dt(x) = (K^{-1})^4(x)$ is the normal 1-form on the level surfaces of $t(x)$. The torsion tensor then arises due to any nonclosed forms among the first three tetrad 1–forms and

$$\tilde{\Upsilon} = \sum_{b=1}^{3} K_b \otimes (dK^{-1})^b.$$

The torsion structure is then most easily treated on orientable globally hyperbolic[21,23] space-times whose topology is $M = S \times R^1$. Here, S is a spacelike orientable (hence parallelizable[24]) 3-D manifold which may be taken to be a Cauchy surface for each global-time value $t \epsilon R^1$. The question of vanishing torsion then reduces to the existence of closed, inexact, nowhere-zero global 1-forms on S. If the first cohomology module $H^1(S) = 0$, there must be torsion or $S \sim R^3$ in the orientable globally hyperbolic case.

In the more general case of a parallelizable space-time M, the condition $H^1(M) \neq 0$ is a necessary but not sufficient condition for vanishing torsion. Torsion is the rule rather than the exception for general cases.

IV. VARIATIONAL METHODS

The usual Dirac action[3,4] is

$$I = -\int \bar{\psi}(\gamma^\mu(x)\partial_\mu + \mathrm{im})\psi\omega,$$

where ω is the invariant volume form $(-g)^{1/2} d^4x$, and $\bar{\psi} = \psi^\dagger \gamma^A$. Note that γ^A is the constant Dirac matrix, not $\gamma^A(x) = K^A_a(x)\gamma^a$.

From $\delta\bar{\psi}$, the Dirac equation results, in the free case, being

$$(\gamma^\mu(x)\partial_\mu + \mathrm{im})\psi = 0.$$

From $\delta\psi$ results the free-adjoint equation,

$$\nabla_\mu(\bar{\psi}\gamma^\mu(x)) - \mathrm{im}\,\bar{\psi} = 0, \tag{4.1}$$

where $\nabla = \partial + \Gamma$ involves the Levi-Cività connection. Since ψ is a general scalar, $\partial_\mu \psi = \psi_{;\mu}$, and the current $i\bar{\psi}\gamma^\mu(x)\psi = J^\mu(x)$ is conserved relative to the metric connection, namely, $J^\mu{}_{;\mu}(x) = 0$.

The structure of the adjoint equation is interesting since the $\gamma(x)$ matrices undergo tetrad parallel transport, $\tilde{\nabla}_{K_a}\gamma(x) = 0$. In a local coordinate chart, the adjoint equation is, using Eq. (3.1),

$$(\partial_\mu \bar{\psi})\gamma^\mu(x) - \mathrm{im}\,\bar{\psi} + (\Gamma^\mu_{\mu\alpha} - \tilde{\Gamma}^\mu_{\mu\alpha})\bar{\psi}\gamma^\alpha(x) = 0. \tag{4.2}$$

Again, general covariance is manifest since $\bar{\psi}$, as a general scalar, gives $\partial_\mu \psi = \bar{\psi}_{;\mu}$ and $(\Gamma - \tilde{\Gamma})$ is a tensor.

Equation (4.2) may be simplified by using the orthonormal *congruence* property of the tetrad fields which implies that $K^\mu_{a;\mu} = 0$ or equivalently, for the 1-forms, that $d{*}K^{-1a} = 0$, where $*$ is the Hodge dual map. Thus from Eq. (3.1) and $\nabla \cdot K_a = 0$, we have that $\Gamma^\mu_{\mu\alpha} = \tilde{\Gamma}^\mu_{\mu\alpha}$ and the adjoint Dirac equation is simply

$$(\partial_\mu \bar{\psi})\gamma^\mu(x) - \mathrm{im}\,\bar{\psi} = 0,$$

the obvious generalization of the flat R^4 equation. Again, a set of connection coefficients results if a change of tetrad is used.

V. CONCLUSIONS

In this paper, the isomorphic map of Clifford and Lie bundles [over parallelizable space-times (M, g)] to arbitrary coordinate atlases via an on tetrad was used to develop a fully covariant Dirac spinor theory. The Klein–Gordon equation contained not only a tetrad affine connection associated with the given $O^+(M)$ cross section, but also a spin-torsion term of the Einstein–Cartan variety. The torsion tensor associated with a smooth parallelizable space-time can vanish only if $(M, g) \sim (R^4, \eta)$ or $H^1(M) \neq 0$, the cohomology criterion being necessary but not sufficient. In work to be reported, the vanishing of torsion is shown to be necessary and sufficient for certain classes of Klein–Gordon solutions to exist.

It might be asked why torsion was not seen in previous generalizations of the flat R^4 Dirac theory. It is generally the case that the coordinate derivative ∂_μ is tacitly replaced by the Levi-Civitá covariant derivative ∇_μ in most treatments. Since this connection is torsion-free, no intrinsic torsion could be obtained by the usual methods. Classical scalar field theories use the metric connection to construct their Klein–Gordon equations and no intrinsic torsion is seen. Electrodynamics is a theory of differential forms which is metric independent and, hence, no torsion enters. Einstein gravity is an intrinsically symmetric theory and any torsion must be inserted as a subsidiary structure on the space-time.[9]

Torsion enters spinor theories naturally through the tetrad map of the Dirac algebra elements. The spin-$\frac{3}{2}$ case is treated in work to be reported.

ACKNOWLEDGMENTS

JRU wishes to thank B. DeFacio and J. Beem of the University of Missouri, and R. Kantowski of the University of Oklahoma for valuable discussions.

[1]S. W. Hawking and R. Penrose, Proc. R. Soc. London Ser. A: **314**, 529 (1970).
[2]R. Geroch, J. Math. Phys. **11**, 343 (1970); **11**, 437 (1970). See also J. Milnor, in *Differential and Combinatorial Topology* edited by S. Cairns (Princeton U. P., Princeton, NJ, 1965), p. 55.
[3]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), Chap. 12.
[4]S. Chandrasekhar, *General Relativity, An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979), Chap. 7.
[5]S. Deser and C. J. Isham, Phys. Rev. D. **14**, 2505 (1976).
[6]S. Deser, "Exclusion of Static Solutions in Gravity-Matter Coupling" in *Field Theory, Quantization and Statistical Physics*, edited by E. Tirapegui (D. Reidel, Boston, 1981), p. 77.
[7]S. B. Edgar, Gen. Relativ. Gravit. **12**, 347 (1980).
[8]See H. Weyl, Phys. Rev. **77**, 699 (1950) and Z. Phys. **56**, 330 (1929). Also E. Schrödinger, Sitzungsber. Akad. Phys. **57**, 261 (1929) and C. Møller, in *Proceedings of the International School of Physics "Enrico Fermi," Evidence for Gravitational Theories* edited by C. Møller (Academic, London, 1962), p. 252.
[9]F. W. Heyl, P. von der Heyde, and G. D. Kerlick, Rev. Mod. Phys. **48**, 393 (1976).
[10]M. Rosembaum, *et al.*, Phys. Rev. D **26**, 761 (1982).
[11]P. C. Aichelburg and T. Dereli, Phys. Rev. D **18**, 1754 (1978).
[12]B. DeFacio and D. G. Retzloff, J. Math. Phys. **21**, 751 (1980).
[13]J. R. Urani and M. H. Kemp, J. Math. Phys. **23**, 227 (1982); **23**, 423 (1982); and J. Math. Phys. (to appear).
[14]C. T. J. Dodson, *Catagories, Bundles and Space-time Topology* (Shiva, Orpington, 1980).
[15]R. Ablamowicz, Z. Oziewicz, and R. Rzewuski, J. Math. Phys. **23**, 231 (1982).
[16]N. Salingaros, J. Math. Phys. **23**, 1 (1982).
[17]S. Schweber, *Relativistic Quantum Field Theory* (Harper and Row, New York, 1961), p. 79.
[18]S. J. Avis and C. J. Isham, in *Recent Developments in Gravitation, Cargese, 1978* edited by M. Levy and S. Deser (Plenum, New York, 1979).
[19]D. Bleeker, *Gauge Theory and Variational Principles* (Addison-Wesley, Reading, 1981), p. 29.
[20]K. Nomizu, *Lie Groups and Differential Geometry* (Mathematical Society of Japan, 1956).
[21]J. Beem and P. Ehrlich, *Global Lorentzian Geometry* (Marcel Dekker, New York, 1981).
[22]P. Van Nieuwenhuizen, *Recent Developments in Gravitation, Cargese, 1978*, edited by M. Levy and S. Deser (Plenum, New York, 1979), p. 523.
[23]R. Geroch, J. Math. Phys. **9**, 1739 (1968). Also, Ref. 2.
[24]N. Steenrod, *Topology of Fiber Bundles* (Princeton U. P., Princeton, NJ, 1955), p. 203.

Modeling quantum behavior with standard (nonquantum) probability theory

Robert F. Bordley
Societal Analysis Department, General Motors Research Laboratories, Warren, Michigan 48090

(Received 30 November 1982; accepted for publication 6 May 1983)

This paper argues that quantum behavior can be modeled using standard probability theory. To show this, such a model is constructed in which the Lagrangians associated with different paths are random. (This random Lagrangian formulation is equivalent to constrained entropy maximization.) We assume that the random error term varies as a harmonic oscillator over time. (We attribute this to certain properties of measuring devices.) The result is a formula which provides a good qualitative description of the n-slit interference experiment—indeed the formula is quite similar to the formulas of quantum mechanics. Hence standard probability theory models can describe interference effects so that a quantum probability theory is unnecessary.

PACS numbers: 02.50. + s, 03.65 − w, 05.20. − y

INTRODUCTION

Almost all theorists agree that quantum behavior is not deterministic; hence probability theory must be used to model it. Many theorists[1-3] go even further and argue that the standard probability theory is incapable of modeling quantum behavior. In its place, a generalized theory of probability, quantum probability theory, has been developed.

Generalizing the standard probability theory is a drastic step. The works of Kolmogorov[4] and DeFinetti[5] have provided the standard probability theory with strong mathematical/philosophical foundations. Since the standard probability theory has proved adequate in most other fields, we should avoid generalizing the standard probability theory unless it is necessitated by experiment.

Quantum probability theorists argue that the standard theory of probability cannot model such well-known quantum effects as the n-slit interference experiment.[6] This paper argues, contrary to the quantum probability theory view, that the standard probability theory *can* model quantum behavior. To support our contention, this paper develops a standard probability theory model of the n-slit interference experiment.

To develop such a model, we first look to the principle of entropy maximization. Suppose a particle has a probability distribution over all possible trajectories ω, which maximizes its entropy: $-\Sigma P(\omega)\log[P(\omega)]$, subject to two constraints:

(a) The probability of various trajectories ω is nonnegative,

(b) The Lagrangian of each trajectory, averaged over $P(\omega)$, gives us the particle's classical Lagrangian (this provides some correspondence with classical mechanics).

It has been shown[7] that such a model is equivalent to the following "random Lagrangian" model:

A particle follows that trajectory which minimizes the time integral of the Lagrangian. However, the experimenter cannot measure the exact value of the Lagrangian for all possible trajectories. Hence there is some measurement error so that the experimenter can only predict the probability with which the particle follows trajectories. This measurement error is distributed as a double exponential.

The equivalence between constrained entropy maximization and the random Lagrangian model suggests that the random Lagrangian formulation[8] might be useful in understanding particle behavior. We only need to specify how the random measurement error on a path at time t relates to the random measurement error on the path at some later time t'.

To specify how measurement error changes over time, suppose we consider the experimenter trying to measure the position of the particle at time t. Suppose the actual position is $X(t)$ while the experimenter measures it as $X^m(t)$. Thus we have an error of $X^m(t) - X(t)$. If the experimenter's methods adjust for error over time, they will tend to decrease the value of this error: $X^m(t) - X(t)$. Suppose we model this adjustment for error as a force: $F = -k[X^m(t) - X(t)]$. Clearly this will cause the error in position to vary sinusoidally over time—first the experimenter gives too high a value for position, then too low, then too high, etc. We will incorporate this adjustment force into the random component of the particle's Lagrangian. (This force, of course, will be random since we do not know the amplitude of the variation nor what the error is at time $t = 0$.)

Following these suggestions then, we assume that the random component of the random Lagrangian is a harmonic oscillator with unknown amplitudes and starting phase. This will give rise to a correlation between the Lagrangians of different paths which will depend upon the difference in path lengths. We will develop a random Lagrangian model for a particle choosing a path in the n-slit interference expeirment. When we insert this correlation between paths into our model, we get a formula which describes the n-slit interference experiment. Indeed this formula is very similar to the formula derived by taking a quantum probability theory approach. Thus a model developed in terms of the standard probability theory describes the results of the n-slit interference experiment. Hence the n-slit interference experiment does not invalidate the use of standard probability theory in quantum mechanics; indeed it provides an opportunity to demonstrate the adequacy of the standard probability theory.

Now there are other ways in which we could have specified how measurement error varies over time. These will give

 0022-2488/83/102411-11$02.50

us different results. Thus we are not arguing that the particular model developed in this paper should replace quantum mechanics. Instead we are arguing that a standard (nonquantum) probability theory model, like the one developed in this paper, can describe the interference effects of the n-slit interference experiment. This refutes the quantum probability theorist assertion about the inadequacy of standard probability theory and suggests another approach for those interested in stochastic formulations of quantum mechanics (see Bohm[9], Comisar,[10] Braffort and Tzara,[11] and Nelson[12]).

In the quantum mechanical theory, an entity is viewed as having a dual wave/particle nature. In my model, the entity is a particle but its Lagrangian, by virtue of its random harmonic oscillator error term, is somewhat wavelike. Thus my formulation does not eliminate quantum-type effects; instead it simply shifts them from the probability theory used to model the physics to the physics being modeled. This has a number of advantages. First it is philosophically appealing to a number of theorists. Second it may suggest testable hypotheses. Third it may offer new ways to generalize quantum mechanics to relativistic contexts.

This paper consists of four sections. The first section reviews the n-slit interference experiment and why many theorists feel the standard probability theory cannot describe it. The second section begins building the random Lagrangian model. The third section considers that model for various types of correlation. In the fourth section, we postulate that the random component of the measured Lagrangian is a harmonic oscillator and deduce the resulting correlation. We insert this correlation into our random Lagrangian model and derive a formula which describes the results of the n-slit interference experiment. Indeed the formula is very similar to the formula we get from the traditional quantum probability theory model.

1. THE n-SLIT INTERFERENCE EXPERIMENT

Imagine the following idealized experiment with electrons. Electrons are produced at a source S and move toward a wall with two slits (see Fig. 1). If we install a detection screen behind the wall, we can record whether or not the electron hits a point x along the wall. If we close the first slit, slit 1, then the probability with which the electron hits various positions x along the wall is given by a bell-shaped distribution with the maximum at the point $x = +\frac{1}{2}d$, the point on the screen directly across from slit 2. Likewise if we open slit 1 and close slit 2, then $P(x)$ has a bell-shaped distribution with maximum at the point $x = -\frac{1}{2}d$ (see Fig. 2). We call the probability distribution, when we close slit 1, $p(x/2)$, the probability the particle hits point (x) on the detection screen given it went through slit 2. Likewise we call the probability distribution when we open slit 1 and close slit 2, $p(x/1)$.

Now suppose we open both slits. Then the probability distribution $P(x)$ becomes a bell-shaped distribution with maximum at $x = 0$ and has "superimposed interference fringes" (see Fig. 3). We call this probability distribution for two open slits $p(x/1,2)$, the probability the particle reaches x given it can travel through slit 1 or slit 2.

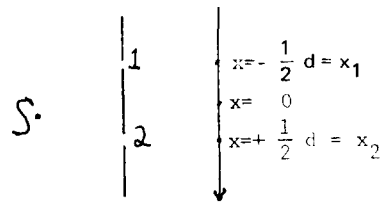


FIG. 1. Layout of the 2-slit interference experiment.

Now, as the quantum probability theorists correctly note, there should be some relationship among $p(x/1,2)$, $p(x/1)$, and $p(x/2)$. In fact, using the standard theory of probability, we can write

$$p(x/1,2) = p(x/1,(1,2))p(1/(1,2)) + p(x/2,(1,2))p(2/(1,2)). \tag{1.1}$$

The quantum probability theorists then implicitly assume that

$$p(x/1,(1,2)) = p(x/1) \quad \text{and} \quad p(x/2,(1,2)) = p(x/2). \tag{1.2}$$

This implies that

$$p(x/1,2) = w_1 p(x/1) + (1 - w_1)p(x/2), \quad \text{where}$$
$$w_1 = p(1/(1,2)). \tag{1.3}$$

According to (1.3), the distribution given in Fig. 3 is a weighted average of the two distributions given in Fig. 2. But this is clearly false.

The quantum probability theorists conclude that standard probability theory, i.e., Eq. (1.1), is the problem. We will argue that (1.2) and not (1.1) is the problem. This paper constructs a model in which the Lagrangians associated with paths from slit 1 to point x are correlated with the Lagrangians associated with paths from slit 2 to point x. Now $p(x/1,(1,2))$ is the probability a particle takes a path from slit 1 to point x given that the Lagrangians associated with all paths through slit 2 are greater than the Lagrangian of some path going through slit 1. The distributions of the Lagrangians for paths going through slit 2, given they all exceed the Lagrangian of some path through slit 1, are different from the overall distributions of the Lagrangians for paths going through slit 2.

Hence, since the Lagrangian for a path from slit 1 to point x is correlated with these slit 2 Lagrangians, its distribution is different in the case in which we know that all the slit 2 Lagrangians exceed some Lagrangians for a path through slit 1 vs the case in which the slit 2 Lagrangians could be larger or smaller than the slit 1 Lagrangians. Thus $p(x/1,(1,2))$ is different from $p(x/1)$, the probability, given we don't even allow for paths through slit 1, and hence have no information about what those slit 2 Lagrangians might have been. Thus assumption (1.2) need not be true.

Later sections of this paper construct the model outlined above. We consider four special cases of this model: in the first two cases, assumption (1.2) will be true; in the last two cases, it will be false. We will find that for the case which best describes the n-slit interference experiment, assumption (1.2) is false. Hence we cannot reject assumption (1.1); hence

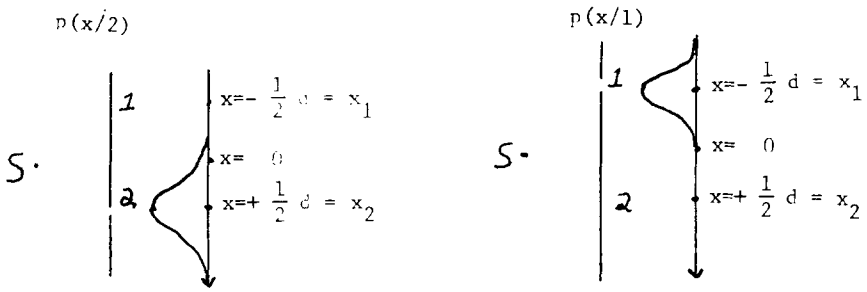


FIG. 2. Results with only one slit open.

the argument against the standard probability theory fails.

Define $\alpha_k = p(x/k,(1,2))/p(x/k)$ so that α_k measures the degree to which assumption (1.2) fails. Then we can write (1.1) as

$$p(x/1,2) = \sum_{k=1}^{2} p(k/(1,2))\alpha_k p(x/k). \tag{1.4}$$

For the general n-slit interference experiment, we have

$$p(x/1,2,...,n) = \sum_{k=1}^{n} p(k/1,2,...,n)\alpha_k p(x/k). \tag{1.5}$$

Now $\alpha_k \neq 1$ is equivalent to saying that (1.2) is false. For the quantum probability theorists to reject standard probability theory, they need $\alpha_k = 1$. This paper develops models in which $\alpha_k \neq 1$ in general.

2. THE GENERAL MODEL

Let P be the set of all paths by which the particle can get from the source to the detection screen. Define (k,θ) to be that path in which the particle moves in a straight line from the source to slit k and then moves at an angle θ from the perpendicular to the detection screen (see Fig. 4). Feynman and Hibbs[13] similarly considered the set of all paths, P. We define the smaller set of paths, $P^R = \{(k,\theta) | k = 1,2,...,n; -\pi/2 \leqslant \theta \leqslant \pi/2\}$. For simplicity, we ignore all paths in P which are not in P^R. (Thus we exclude "wobbly" paths from source to screen as well as "ricocheting" paths. These could be included in a more thorough treatment.)

Let $L(k,\theta)$ be the time integral of the Lagrangian associated with path (k,θ). We occasionally refer to this as the Lagrangian of path (k,θ). By Hamilton's principle, the particle chooses that path (k,θ) minimizing $L(k,\theta)$ over all (k,θ) in P.

Although $L(k,\theta)$ has a specific value for each path the particle may take, we assume—consistent with the spirit of the Uncertainty Principle—that the experimenter cannot measure it exactly. Hence because of random measurement error, $L(k,\theta)$ is a random variable. Following Domencich and McFadden[14] and Litinas,[15] we assume[16] that

$$L(k,\theta) = V(k,\theta) - \epsilon(k,\theta),$$

where $\epsilon(k,\theta)$ is a random variable with the distribution

$$\Pr(\epsilon(k,\theta) \leqslant a) = e^{-e^{-(a/s(k,\theta))}}.$$

This is the double exponential or extreme-value distribution. It deviates somewhat from the normal distribution, although it is a good approximation for many purposes. It is more analytically tractable for determining the distribution of the minimum of $L(k,\theta)$ than the normal distribution.

The probability of a particle taking a path which goes through slit k and then proceeds at an angle θ is the probability the path (k,θ) has the smallest Lagrangian of all paths, i.e.,

$$P(k,\theta) = \Pr(L(k,\theta) \leqslant \min_{j,\xi}(L(j,\xi))).$$

If we require that the particle go through slit k, then the probability of a particle taking slit k and proceeding at an angle θ is the probability the path (k,θ) has the smallest Lagrangian of all paths going through slit k, i.e.,

$$P_k(k,\theta) = \Pr(L(k,\theta) \leqslant \min_{\xi}(L(k,\xi))).$$

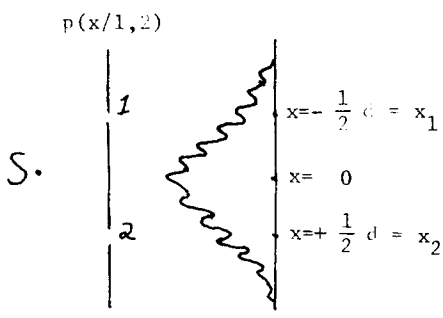


FIG. 3. Results with both slits open.

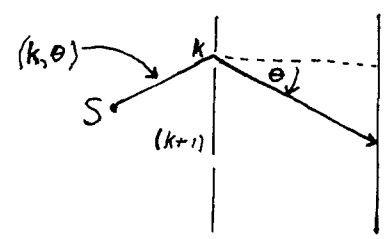


FIG. 4. Path (k, θ).

Now if we let $(k,\theta_k(x))$ be that path which goes to slit k and then proceeds from slit k at an angle causing it to hit point x, we can write down the equations for $P(x)$ given n slits are open and $P(x)$ given only slit k is open. Thus given n slits are open, $p(x/1,2,...,n)$ is the probability that one of the n paths reaching x: $(1,\theta_1(x)),...,(n,\theta_n(x))$ has the smallest Lagrangian, i.e.,

$$p(x/1,2,...,n) = \Pr\Big(\min_k(L(k,\theta_k(x))) \leqslant \min_{j,\xi}(L(j,\xi))\Big).$$

Since there is a zero probability of two different paths obtaining the same minimum, this becomes

$$p(x/1,2,...,n) = \sum_k \Pr\Big(L(k,\theta_k(x)) \leqslant \min_{j,\xi}(L(j,\xi))\Big). \quad (2.1)$$

If only slit k is open, then $p(x/k)$ is the probability that $(k,\theta_k(x))$ has the smallest Lagrangian of all paths going through slit k, i.e.,

$$p(x/k) = \Pr\Big(L(k,\theta_k(x)) \leqslant \min_{\xi}(L(k,\xi))\Big). \quad (2.2)$$

We can also determine the probability the particle goes through slit k given that n slits are open. It is the probability that one of the paths going through slit k has the smallest Lagrangian, i.e.,

$$p(k/1,2,...,n) = \Pr\Big(\min_{\delta}(L(k,\delta)) \leqslant \min_{j,\xi}(L(j,\xi))\Big). \quad (2.3)$$

Thus we can write Eq. (1.5) as

$$p(x/1,2,...,n) = \sum_{k=1}^{n} p(k/1,2,...,n)\alpha_k p(x/k), \quad (2.4)$$

where α_k is given by

$$\alpha_k = \frac{\Pr(L(k,\theta_k(x)) \leqslant \min_{j,\xi}(L(j,\xi)))}{\Pr(\min_{\delta}(L(k,\delta)) \leqslant \min_{j,\xi}(L(j,\xi)))\Pr(L(k,\theta_k(x)) \leqslant \min_{\xi}(L(k,\xi)))}. \quad (2.5)$$

The next section specifies various correlations among the $\epsilon(k,\theta)$'s, which give rise to different values for the α_k's.

3. FOUR CASES OF CORRELATION

We consider four cases

Case I: independence

Suppose all $\epsilon(k,\theta)$'s are independent. Then

$$\Pr\Big(L(k,\theta) \leqslant \min_{j,\xi}(L(j,\xi))\Big) = \frac{e^{-V(k,\theta)}}{\sum_{k=1}^{n}\int_{-\pi/2}^{\pi/2} e^{-V(k,\xi)}\,d\xi}.$$

Proof: See Domencich and McFadden[14] and Litinas.[15]

Using Eqs. (2.1)–(2.3), we can compute the probability distribution for x, given all n slits are open and given only one slit is open. We use (2.5) to compute the resulting value of α_k. We find $\alpha_k = 1$. So we would expect assumption (1.2) and thus (1.3) to be correct if the random Lagrangians were uncorrelated.

Case II: slit correlation

Suppose there is a correlation $\rho(k)$ among the random Lagrangians in all paths going through the same slit k (for $k = 1,2,...,n$). However, paths going through different slits have uncorrelated Lagrangians. Then we find

$$\Pr\Big(L(k,\theta) \leqslant \min_{j,\xi}(L(j,\xi))\Big) = \frac{e^{-\bar{V}(k,\theta)}\big(\int e^{-\bar{V}(k,\xi)}\,d\xi\big)^{-\sigma(k)}}{\sum_{j=1}^{n}\int e^{-\bar{V}(j,\theta)}\big(\int e^{-\bar{V}(j,\xi)}\,d\xi\big)^{-\sigma(k)}\,d\theta},$$

where $\bar{V}(k,\theta) = V(k,\theta)/[1-\sigma(k)]$ and $\sigma(k) = 1-[1-\rho(k)]^{1/2}$

Proof: See Litinas[15] and Block and Marschak.[17]

We can compute the probability distribution for the particle reaching x, given all n slits are open and given only one slit is open using (2.1)–(2.3). If we then use Eq. (2.5) to find α_k, we find that $\alpha_k = 1$ for all k. Thus correlation among the Lagrangians in and of itself will not cause Eq. (1.2) to be violated.

We now consider a case of correlation which does violate assumption (1.2).

Case III: angular correlation

Suppose there is a correlation $\rho(\theta)$ among the Lagrangians of all paths which emerge from the slit at the same angle θ towards the perpendicular. In other words, $\epsilon(j,\theta)$ and $\epsilon(k,\theta)$ are correlated for $j \neq k$. On the other hand, there is no correlation among Lagrangians for paths emanating from the same slit. With these conditions, we get a formula similar to that of Case II, namely,

$$\Pr(L(k,\theta) \leqslant \min_{j,\xi}(L(j,\xi))) = \frac{e^{-V^*(k,\theta)}\big(\sum_{j=1}^{n} e^{-V^*(j,\theta)}\big)^{-\sigma(\theta)}}{\sum_{k=1}^{n}\int e^{-V^*(k,\theta)}\big(\sum_{j=1}^{n} e^{-V^*(j,\theta)}\big)^{-\sigma(\theta)}\,d\theta},$$

where $V^*(k,\theta) = V(k,\theta)/[1-\sigma(\theta)]$ and $\sigma(\theta) = 1-[1-\rho(\theta)]^{1/2}$.

Proof: See Litinas[15] and Block and Marschak.[17]

If we now use Eqs. (2.1)–(2.5), we will find that we do not get $\alpha_k = 1$. This is because $p(x/k,1,2,...,n)$ is a function not only of $V^*(k,\theta_k(x))$ but of all the other $V^*(j,\theta_j(x))$, $j \neq k$, values with which there is correlation. On the other hand, $p(x/k)$ is given by

$$p(x/k) = e^{-V(k,\theta_k(x))}/\int e^{-V(k,\theta)}\,d\theta$$

and thus $p(x/k)$ does not depend upon $V^*(j,\theta_j(x))$ for $j \neq k$.

If we compute α_k, we find that it is given by

$$\alpha_k = \delta(k,\theta_k(x))/E(\delta),$$

where $\delta(k,\theta)$, the correlation shift parameter is given by

$$\delta(k,\theta) = \Big(1 + \sum_{j\neq k} e^{-(V(j,\theta) - V(k,\theta)/[1-\sigma(\theta)])}\Big)^{-\sigma(\theta)}$$

and where

$$E(\delta) = \frac{\sum_{k=1}^{n} \int \delta(k,\theta) e^{-V(k,\theta)}\, d\theta}{\sum_{k=1}^{n} \int e^{-V(k,\theta)}\, d\theta}.$$

The correlation shift parameter measures the impact of correlation. The amount by which this affects the probability of the path is the ratio of this correlation shift parameter to the average correlation shifts affecting all paths.

Now α_k will only equal 1, in general, if $\delta(k,\theta)$ is the same for all angles θ, i.e., if $V(j,\theta)$ and $\rho(\theta)$ were the same for all angles θ. Since this is usually not the case, $\alpha_k \neq 1$. Thus assumption (1.2) is false, and the argument against the standard theory of probability is invalid.

We can make a number of other observations, which are detailed in Appendix A. These show that positive correlation tends to decrease the probability of a path being taken while negative correlation increases the path's probability of being taken.

We now look at one final case of correlation which will also violate assumption (1.2).

Case IV: bivariate position correlation

In Case IV, there is a correlation $\rho(x;i,j)$ between the paths $(i,\theta_i(x))$ and $(j,\theta_j(x))$ for all $i,j = 1,2,\ldots,n$ and for all x. To develop the formulas for this case, we first develop the solution for the case in which $\rho(x;i,j) = \rho(x)$ for all i and j. Then we approximate the solution to get a formula for the case in which $\rho(x;i,j)$ varies for different i and j.

Let D be the distance from the slits to the detection screen. Let x_k denote that point on the detection screen directly across from slit k (see Fig. 5.) Then we know that $\theta_k(x)$ satisfies $\tan[\theta_k(x)] = +(x - x_k)/D$. Hence for the path (k,θ), an integration of $d\theta$ from $-\pi/2$ to $+\pi/2$ corresponds to an integration of $f_k(x)\,dx$ from $-\infty$ to $+\infty$ (where $f_k(x) = \{1 + [(x - x_k)/D]^2\}^{-1}$). The density $f_k(x)$ is bell-shaped like the density in Fig. 2.

We now suppose that all paths (k,x) which terminate at the same point x have Lagrangians which are correlated with correlation coefficient $\rho(x)$. Thus $\epsilon(j,x)$ and $\epsilon(k,x)$ are correlated for $j \neq k$. Under these conditions, we have

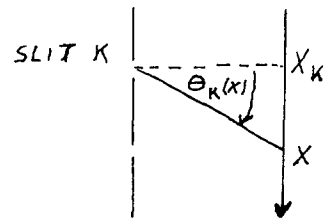


FIG. 5. Angle $\theta_k(x)$.

$$\Pr\Big(L(k,x) \leqslant \min_{j,y}(L(j,y))\Big) = \frac{e^{-V^+(k,x)}\left(\sum_{j=1}^{n} e^{-V^+(j,x)}\right)^{-\sigma(x)} f_k(x)}{\sum_{k=1}^{n} \int e^{-V^+(k,x)}\left(\sum_{j=1}^{n} e^{-V^+(j,x)}\right)^{-\sigma(x)} f_k(x)\,dx},$$

where $V^+(k,x) = V(k,x)/[1 - \sigma(x)]$ and $\sigma(x) = 1 - [1 - \rho(x)]^{1/2}$.

Proof: See Litinas[15] and Block and Marschak.[17] In the case in which $V(k,x) = V$ for all k and x, we find that

$$\alpha_k = \frac{n^{-\sigma(x)}}{E(n^{-\sigma(x)})} \left(\text{where } E(n^{-\sigma(x)}) = \frac{\sum_{k=1}^{n} \int (n)^{-\sigma(x)} f_k(x)\,sx}{\sum_{k=1}^{n} \int f_k(x)\,dx}\right)$$

which is similar to what we had in Case III. Thus this case also violates assumption (1.2).

These are the formulas for position correlation. To use them to develop formulas for bivariate correlation, we first consider the $n = 2$ case. In this case, the correlation shift parameter $\delta(1,x)$ is given by

$$\delta(1,x) = (1 + e^{[V(1,x) - V(2,x)]/(1 - \sigma(x))})^{-\sigma(x)}.$$

We first note that $p(x/2)/p(x/1) = e^{V(1,x) - V(2,x)}$. Hence $\delta(1,x) = \{1 + [p(x/2)/p(x/1)]^{[1/(1-\sigma(x))]}\}^{-\sigma(x)}$. We now expand $\delta(1,x)$ in a Taylor series about $\sigma(x) = 0$ and get

$$\delta(1,x) = 1 - \sigma(x)\ln\left(1 + \frac{p(x/2)}{p(x/1)}\right) + (\text{terms to be neglected}).$$

Thus

$$\alpha_1 = \frac{\delta(1,x)}{E_{(1,2)}(\delta(x))} = \frac{1 - \sigma(x)\ln\{1 + [p(x/2)/p(x/1)]\}}{1 - \int \sigma(x)\{\ln[1 + p(x/2)/p(x/1)]g_1(x) + \ln[1 + p(x/1)/p(x/2)]g_2(x)\}\,dx},$$

where

$$g_k(x) = \frac{e^{V(k,x)} f_k(x)}{\int e^{V(k,x)} f_k(x)\,dx}.$$

We let $\beta_1(1,2) = 1/E_{(1,2)}(\delta(x))$ so that $\alpha_1 = \beta_1(1,2)\delta(1,x)$. This gives us the following equation for $p(x/1,2)$:

$$\begin{aligned} p(x/1,2) &= \sum_{j=1}^{2} p(x/j)p(j/1,2)\alpha_j \\ &= \tfrac{1}{2}(\beta_1(1,2)p(x/1) + \beta_2(1,2)p(x/2) \\ &\quad - \sigma(x)[\beta_1(1,2)\ln(1 + [p(x/2)/p(x/1)])p(x/1) \\ &\quad + \beta_2(1,2)\ln\{1 + [p(x/1)/p(x/2)]p(x/2)\}]). \end{aligned} \tag{3.1}$$

Now the obvious way to generalize this model is to look at the correlation shift parameter for an arbitrary n, i.e.,

$$\delta(1,x) = \left(1 + \sum_{\substack{j=1 \\ j \neq 1}}^{n} e^{[V(1,x) - V(j,x)]/[1 - \sigma(x)]}\right)^{-\sigma(x)}$$

However, this would imply that all paths leading to the same point x have the same correlation with one another. We want to allow the n paths to have different correlations with one another. This requires that we take a different approach.

First define the Lagrangian associated with a slit j to be the smallest Lagrangian of all paths going through slit j, i.e.,

$$L(j) = \min_x [L(j,x)].$$

Now the probability that path (k,x) has the smallest Lagrangian is the probability that (1) slit k is one of the two slits with the smallest Lagrangians of all n slits, (2) of these two slits, slit k has the smaller Lagrangian and the path (k,x) has the smallest Lagrangian of all paths through slit k.

In other words,

$$p((k,x)/(1,2,...,n)) = \sum_{j=1}^{n} \sum_{s<j} p((k,x)/(j,s),(1,...,n)) p((j,s)/(1,...,n)), \tag{3.2}$$

where the notation (j,s) indicates that the particle goes through either slit j or slit s. We will now assume that $p((k,x)/(j,s),(1,...,n)) = p((k,x)/(j,s))$, i.e., that, given we know the particle will go through either slit j or slit s, we can ignore all the other slits.

Let's compare this assumption with the quantum probability theorist assumption, (1.2), which stated that $p((k,x)/k,(1,...,n)) = p((k,x)/k) = p(x/k)$. Their assumption said that, given we know the slit with the smallest Lagrangian, we can ignore all the other slits in determining the probability of path (k,x). Our assumption, (3.2), says that given we know the two slits with the two smallest Lagrangians, we can ignore all the other slits in determining the probability of path (k,x). Thus assumption (3.2) is a fairly straightforward generalization of assumption (1.2).

Now clearly, $p((k,x)/(j,s)) = 0$ if $j \neq k$ and $s \neq k$. Likewise, if we assume that all pairs of slits have an equal chance of being the slits with the two smallest Lagrangians, we get $p((j,s)/1,2,...,n) = 2/n(n-1)$. With these substitutions, (3.2) becomes

$$p((k,x)/(1,...,n)) = \frac{2}{n(n-1)} \sum_{\substack{j=1 \\ j \neq k}}^{n} p((k,x)/(j,k)).$$

Now

$$p((k,x)/(j,k)) = p(x/k,(j,k)) p(k/(j,k)) = \tfrac{1}{2} p(x/k,(j,k)).$$

Since we know that

$$p(x/k,(j,k)) = \beta_k(j,k)\{p(x/k) - \sigma(x;j,k) p(x/k) \times \ln[1 + p(x/j)/p(x/k)]\},$$

we have

$$p((k,x)/(j,k)) = \tfrac{1}{2}\beta_k(j,k)\{p(x/k) - \sigma(x;j,k) p(x/k) \times \ln[1 + p(x/j)/p(x/k)]\}.$$

So we get

$$p((k,x)/(1,...,n)) = \left(\frac{1}{n(n-1)}\right)\left[p(x/k) \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_k(j,k) - p(x/k) \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_k(j,k) \sigma(x;j,k) \ln\left(1 + \frac{p(x/j)}{p(x/k)}\right)\right].$$

But

$$p(x/1,...,n) = \sum_{k=1}^{n} p((k,x)/(1,...,n)) = \frac{1}{n(n-1)}\left[\sum_{k=1}^{n} p(x/k) \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_k(j,k) - \sum_{k=1}^{n} p(x/k) \sum_{\substack{j=1 \\ j \neq k}}^{n} \beta_k(j,k) \sigma(x;j,k) \ln\left(1 + \frac{p(x/j)}{p(x/k)}\right)\right].$$

We can rewrite this as

$$p(x/1,...,n) = \left(\frac{1}{n(n-1)}\right)\left(\sum_{k=1}^{n} p(x/k) \sum_{\substack{j+1 \\ j \neq k}}^{n} \beta_k(j,k) - 2 \sum_{k=1}^{n} \sum_{j<k}^{n} \sigma(x;j,k) \frac{1}{2}\left[\beta_k(j,k) p(x/k) \ln\left(1 + \frac{p(x/j)}{p(x/k)}\right) + \beta_j(j,k) p(x/j) \ln\left(1 + \frac{p(x/k)}{p(x/j)}\right)\right]\right). \tag{3.3}$$

This is the bivariate correlation model.

We now specify a form for the correlation, $\rho(x;j,k)$, which will give us a form for $\sigma(x;j,k)$. When we insert this into our bivariate correlation model, the resulting formula predicts interference fringes.

4. A MODEL OF THE n-SLIT INTERFERENCE EXPERIMENT

We have discussed fairly simple kinds of correlation: all paths emerging from the same slit being correlated with constant correlation; all paths proceeding from the same angle θ to the perpendicular being correlated with constant correlation. We now consider a physically more intuitive form of correlation.

First let $L(j,x;t)$ be the Lagrangian of path (j,x) at time t. [Thus $L(j,x;t)$ equals the kinetic energy minus the potential energy.] Similarly define the known and random components of the Lagrangian, $V(j,x;t)$ and $\epsilon(j,x;t)$, so that

$$L(j,x;t) = V(j,x;t) - \epsilon(j,x;t).$$

If the path from source S through slit j to the point x starts at time $t = 0$ and ends at time $t = T$, we also have

$$L(j,x) = \int_0^T L(j,x;t)\, dt,$$

$$V(j,x) = \int_0^T V(j,x;t)\, dt,$$

$$\epsilon(j,x) = \int_0^T \epsilon(j,x,;t)\, dt.$$

We will use this notation in defining a model for $\epsilon(j,x)$ which, in turn, implies a form for $\rho(x;j,k)$.

Nelson[12] derived the Schrödinger equation from a model in which a particle's motion consists of a classical component reflecting the macroscopic Newtonian physics and a random component associated with frictionless Brownian motion. Our model is somewhat similar.

In our model, we think of the classical component of a particle's motion as resulting from the known component of the particle's Lagrangian, $V(j,x;t)$. We view the unknown component as reflecting the random difference between the known, measured Lagrangian and the actual Lagrangian.

Now our methods of measurement tend to reduce the difference between the measured position of a particle and its actual position. Suppose we represent this as a force, $F = -k[\mathbf{X}(t) - \mathbf{X}^m(t)]$, where $\mathbf{X}(t)$ is the actual position of the particle at time t and $\mathbf{X}^m(t)$ is the measured position. If $\mathbf{X}^m(t)$ changes very slowly relative to $\mathbf{X}(t) - \mathbf{X}^m(t)$, then we can define $\mathbf{z}(t) = [\mathbf{X}(t) - \mathbf{X}^m(t)]$ and assume $\mathbf{z}(t)$ is independent of $\mathbf{X}^m(t)$ and thus of $V(j,x;t)$. Hence $F = -k\mathbf{z}(t)$ gives rise to a simple harmonic oscillator force. The kinetic energy of the oscillator is $\frac{1}{2}kA^2(j,x)\cos^2[(k/m)^{1/2}t + \frac{1}{2}\alpha]$, where $A^2(j,x)$ and $\frac{1}{2}\alpha$ are unknown constants. The potential energy is $\frac{1}{2}kA^2(j,x)\sin^2((k/m)^{1/2}t + \frac{1}{2}\alpha)$. Hence the harmonic oscillator Lagrangian, $\epsilon(j,x;t)$ which is the difference between the kinetic and the potential energy, is given by

$$\begin{aligned}\epsilon(j,x;t) &= \tfrac{1}{2}kA^2(j,x)\{\cos^2[(k/m)^{1/2}t + \tfrac{1}{2}\alpha] \\ &\quad - \sin^2[(k/m)^{1/2}t + \tfrac{1}{2}\alpha]\} \\ &= \tfrac{1}{2}kA^2(j,x)\cos[2(k/m)^{1/2}t + \alpha].\end{aligned}$$

Now, $A(j,x)$ is, of course, related to the energy of the oscillator by $A = (2E/k)^{1/2}$, where E is the energy. Hence, when the particle leaves the source S, a certain amount of energy goes into its harmonic motion. But since we do not know how much energy goes into its harmonic motion, E and thus $A^2(j,x)$ are random variables with possible values ranging from zero to infinity. Likewise we do not know α, the phase in the harmonic motion, at which the particle leaves source S. Since α can assume any value from 0 to 2π, $\cos[2(k/m)^{1/2}t + \alpha]$ can assume any value from $+1$ to -1. Hence $\epsilon(j,x;t)$ can assume any value from $-\infty$ to $+\infty$. By appropriately choosing the probability distribution we assign to $A^2(j,x)$ and α, we can give $\epsilon(j,x;t)$ many possible probability distributions.

We present a graph of $\epsilon(j,x;t)$ in Fig. 6. As Appendix B shows, given this model for $\epsilon(j,x;t)$, the correlation between $\epsilon(j,x)$ and $\epsilon(k,x)$—which is also the correlation between $L(j,x)$ and $L(k,x)$—is given by

$$\rho(x;j,k) = a_0\cos\{C[d_j(x) - d_k(x)]\},$$

$$\text{where } C = \frac{1}{v}\left(\frac{k}{m}\right)^{1/2},$$

where $d_j(x)$ is the length of path (j,x), $d_k(x)$ is the length of path (k,x), v is the speed at which the particle travels along the path—assumed constant for simplicity, a_0 is the correlation between $A^2(j,x)$ and $A^2(k,x)$.

We intend to use this formula for $\rho(x;j,k)$ in (3.1) and (3.2). To do so, recall that $\sigma(x;j,k) = 1 - [1 - \rho(x;j,k)]^{1/2}$. If we make the approximation $\sigma(x;j,k) = [\rho(x;j,k)/2\ln 2]$, the resulting formula—as Appendix C(A) shows—actually improves the degree to which formulas (3.1) and (3.2) approximate the analytical random Lagrangian formulas of Case IV. So we now replace $\sigma(x;j,k)$ by $[1/(2\ln 2)]\rho(x;j,k)$ in (3.1) and (3.2).

Consider the constants β_k. By definition, we have

$$\beta_k(j,k) = \left\{1 - \int_x \frac{\rho(x;j,k)}{2\ln(2)}\left[\ln\left(1 + \frac{p(x/j)}{p(x/k)}\right)g_k(x) + \ln\left(1 + \frac{p(x/k)}{p(x/j)}\right)g_j(x)\right]dx\right\}^{-1}.$$

Now using elementary ideas from physical optics, we know that $[d_j(x) - d_k(x)]$ is approximately equal to xh/D, where h is the distance between slits k and j (see Klein, Ref. 18, p. 187). Thus we can insert $\rho(x;j,k) = a_0\cos[C(h/D)x]$ into the expression for $\beta_k(j,k)$. Since the mass m is small, C is very large and the cosine oscillates between $+1$ and -1 much more rapidly than $\ln[1 + p(j,x)/p(k,x)]g_k(x)$ varies as a function of x. Hence the integral is approximately equal to zero. Thus $\beta_k(j,k) = 1$ and $\Sigma_{\substack{k=1 \\ k\neq j}}^{n}\beta_k(j,k) = (n-1)$ for all j and k. This simplifies formulas (3.1) and (3.3) considerably. Formula (3.1) with this choice of correlation becomes

$$p(x/1,2) = \tfrac{1}{2}\left(p(x/1) + p(x/2) - a_0\left(\frac{p(x/1)\ln[1 + p(x/2)/p(x/1)] + p(x/2)\ln[1 + p(x/1)/p(x/2)]}{2\ln 2}\right)\cos\{C[d_1(x) - d_2(x)]\}\right). \tag{4.1}$$

Formula (3.3) with this choice of correlation becomes

$$p(x/1,2,\ldots,n) = \frac{1}{n}\left\{\sum_{k=1}^{n}p(x/k) - \left(\frac{a_0}{n-1}\right)\left[\sum_{k=1}\sum_{j<k}\left(\frac{p(x/k)\ln[1 + p(x/j)/p(x/k)] + p(x/j)\ln[1 + p(x/k)/p(x/j)]}{2\ln 2}\right)\cos\{C[d_j(x) - d_k(x)]\}\right]\right\}. \tag{4.2}$$

Now suppose that the n slits are fairly close together. Then for many values of x, $[p(x/k)/p(x/j)]$ is fairly close to 1. Appendix C(B) shows that we can approximate $p(x/k)\times\ln[1 + p(x/j)/p(x/k)] + p(x/j)\ln[1 + p(x/k)/p(x/j)]$ by $2\ln 2[p(x/k)p(x/j)]^{1/2}$ with an error of at most 2.7% if $\frac{1}{2} \leqslant p(x/k)/p(x/j) \leqslant 2$. We will call this the "close slit" approximation. If we make the approximation, (4.2) becomes

$$p(x/1,2,\ldots,n) = \frac{1}{n}\left[\sum_{k=1}^{n}p(x/k) - \left(\frac{a_0}{n-1}\right)\sum_{k=1}\sum_{j<k}[p(x/k)p(x/j)]^{1/2}\times\cos\{C[d_j(x) - d_k(x)]\}\right]. \tag{4.3}$$

It is fairly clear that this formula would give the interference

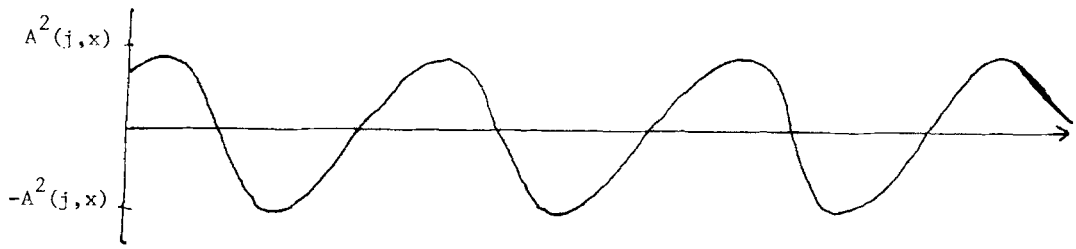


FIG. 6. Graph of $\epsilon(j, x; t)$.

phenomena described in Fig. 3 even if $p(x/k)$ were bell-shaped as in Fig. 2 for all k. This shows that (4.3) and thus (4.2) predict interference effects.

Now let's compare (4.3) with the quantum mechanical solution to this problem. For the quantum mechanical solution, we would define a wave function for each path, $\phi(k,x) - Q(k,x)e^{iB(k,x)}$. We have

$$p(x/(1,\ldots,n)) = \left(\sum_{k=1}^{n} \phi(k,x)\right)\left(\sum_{k=1}^{n} \phi^*(k,x)\right)$$

$$= K_{\text{norm}}\left(\sum_{k=1}^{n} Q^2(k,x) + 2\sum_{k=1}\sum_{j<k} Q(j,x)Q(k,x)\cos[B(j,x) - B(k,x)]\right)$$

$$= K_{\text{norm}}\left(\sum_{k=1} p(x/k) + 2\sum_{k=1}\sum_{j<k} [p(x/j)p(x/k)]^{1/2}\cos[B(j,x) - B(k,x)]\right), \tag{4.4}$$

where K_{norm} is a normalizing constant. Let $B(j,x) = Cd_j(x) = (1/v)(k/m)^{1/2}d_j(x)$. [Thus $B(j,x)$ is proportional to the time needed to traverse path (j,x).] Let $a_0 = -1$. [Thus $A^2(j,x)$ and $A^2(k,x)$ are negatively correlated.] Then the only difference between (4.3) and (4.4) is that (4.3) assigns a weight of $[1/(n-1)]$ to the interference terms while (4.4) assigns a weight of 2 to the interference terms.

Of course, we can get different constants in (4.3) by making different assumptions about the distribution of $\epsilon(j,x)$. The important point is that our model is qualitatively similar to the quantum mechanical one and thus will predict similar qualitative effects. Hence this approach to the behavior of the particle may provide a powerful alternative formulation of quantum theory. We leave developing such an alternative theory to other work.

In summary, this paper has shown that:

(1) Equations derived from the random Lagrangian model with a harmonic oscillator model of measurement error lead to a qualitatively plausible model of the n-slit interference experiment. This establishes that it is possible to model quantum behavior using standard probability theory. Hence the quantum probability theory arguments must be revised.

(2) Many researchers have worked on the idea that one can duplicate quantum behavior by postulating a particle as subject to frictionless Brownian motion.[9–12] (Many other papers discuss related ideas.[19–25]) Our paper provides a distinctly different approach to this basic idea. Instead of postulating frictionless Brownian motion, we postulate measurement error, giving rise to a harmonic oscillator Lagrangian with unknown amplitude and initial phase. The result is a formula which does imply interference effects and is, in many ways, quite similar to the formulas of quantum mechanics. We also saw that this approach was very closely related to the principle of entropy maximization.

CONCLUSION

Most physicists currently feel that quantum behavior such as the n-slit interference experiment requires an abandonment of the standard probability theory. This seems rather peculiar because other fields, while recognizing the necessity of modeling uncertainty, have found the standard probability theory perfectly adequate for their uses. This paper argues that quantum physics does not, in fact, require an abandonment of standard probability theory; instead we merely require a more sophisticated use of standard probability theory.

Our argument proceeds in two steps:

(1) We first examine the assertion that standard probability theory cannot model phenomena in which $p(x/1,2) \neq p(x/1)p(1/1,2) + p(x/2)p(2/1,2)$. To show that this assertion is false, we develop a model in which the particle follows that path minimizing the time integral of the Lagrangian. However, although the particle's behavior is deterministic, the experimenter is unable to measure the Lagrangian exactly and hence can only predict the probability with which the particle follows various paths. This model is equivalent to a constrained entropy maximization principle.

We show that when there is correlation among the La-

grangians, the random Lagrangian model generally predicts that $p(x/1,2) \neq p(x/1)p(1/1,2) + p(x/2)p(2/1,2)$. This refutes the assertion that a standard probability theory model cannot model quantum behavior.

(2) We now proceed to develop a random Lagrangian model which provides a qualitatively reasonable model of the n-slit interference experiment—an especially popular example of quantum effects. To do this, we need to develop a model for the random component of the Lagrangian. We go back to work by Nelson, Tzara, Braffort, Comisar, and Bohm, which suggests that we can think of a particle as having two kinds of motion: (1) a classical motion obeying Newton's laws and (2) an oscillatory Brownian motion component. These authors have shown that such a conceptualization leads to results consistent with the Schrödinger equation.

We use their conceptualization to form our model of the random component of the Lagrangian. We assume this random component is a harmonic oscillator Lagrangian where the amplitude and starting phase of the oscillator are unknown. This implies that the correlation between two path Lagrangians is proportional to the cosine of the difference in path lengths. When we insert this correlation into our random Lagrangian model, we get a formula which provides a model of the n-slit interference experiment. Indeed under certain conditions, for example, the distance between slits being fairly small, the formula is very similar to the formula we would get from the traditional quantum probability theory approach.

Thus we have established that a standard probability theory model can model quantum effects—at least in the case of the n-slit interference experiment—contrary to the arguments of quantum probability theorists. Since the formulas given by my model are not identical to those given by the traditional quantum probability theory approach, there are some obvious ways to test the formulas to see which is more descriptively valid. But this paper is not concerned with establishing which one of the many possible standard probability theory formulations of quantum behavior is "best." We are only concerned with showing that such a formulation is possible. We leave other speculations to later work.

APPENDIX A

We will show that positive correlation decreases a path's probability of being followed while negative correlations tends to increase a path's probability.

(1) Suppose $L(k,\theta)$ is positively correlated with other Lagrangians whose average value tends to be pretty small. Then there are two cases:

(a) $L(k,\theta)$ happens to be pretty small. However, because it is positively correlated with other paths, these other paths also have Lagrangians which are smaller than average. Since these other paths have very small Lagrangians on average anyway, it is very probable that their Lagrangians will be less than $L(k,\theta)$. Thus the probability of the particle following path (k,θ), given $L(k,\theta)$ is small, is less than if $L(k,\theta)$ were uncorrelated with these other paths.

(b) $L(k,\theta)$ tends to be large. In this case, those other paths with which it is correlated have larger than average Lagrangians too. However, since their Lagrangians are small on average, they may still be less than $L(k,\theta)$.

Thus it would appear that being positively correlated with paths of small $V(j,\xi)$ tends to decrease the probability of the particle following path (k,θ). Likewise we would expect that being negatively correlated with such paths would increase the path's probability. We can confirm these conjectures by looking at the formula for $\delta(k,\theta)$.

(2) If $V(k,\theta) = V$ for all k and all θ, then α_k is proportional to $(n)^{-\sigma(\theta_k(x))}$. Thus if $\sigma(\theta_k(x))$ is larger than all other correlations, α_k tends to be less than 1. Furthermore, α_k decreases as the number of slits, n, increases. Hence high positive correlation reduces a path's probability. Conversely, if $\sigma(\theta_k(x))$ is smaller than all other correlations, α_k tends to be greater than 1. Thus low negative correlation increases the path's probability.

Why is this the case? If $L(k,\theta_k(x))$ is positively correlated with the Lagrangians associated with paths from other slits, then $L(k,\theta_k(x))$ will tend to be higher than $V(k,\theta_k(x))$ when those other paths tend to be larger than their V's. Given we know that the particle is coming through slit k, we know that the Lagrangian of all paths not coming through slit k are all larger than the Lagrangian of at least one path coming through slit k. Hence the Lagrangian associated with paths not coming through slit k tends to be somewhat larger than normal. Thus $L(k,\theta_k(x))$ tends to be larger than normal. Hence the probability $L(k,\theta_k(x))$ is the smallest Lagrangian is less than usual. Hence the probability the particle takes path $(k,\theta_k(x))$ is less than in the case of no correlations. Hence α_k is less than 1.

Now suppose $L(k,\theta_k(x))$ is negatively correlated with the Lagrangians associated with paths from other slits. Then, given that we know the path goes through slit k, we know that Lagrangians for paths not going through slit k are higher than their average. Thus $L(k,\theta_k(x))$ is smaller than normal. Hence the probability of path $(k,\theta_k(x))$ is larger than in the absence of correlation. Hence α_k exceeds 1.

Thus in either case, negative correlation increases a path's probability whereas positive correlation does just the reverse.

APPENDIX B

Let $d_1(x)$ denote the length of the path from the source S to slit 1 and then to point x on the detection screen [i.e., the length of path $(1,x)$]. If we assume all particles travel at the same speed v, then the time it takes to travel path $(1,x)$ is just $d_1(x)/v$. Let time $t = 0$ be the time at which the particle leaves the source S. Then the Lagrangian of the path $(1,x)$ is just

$$L(1,x) = \int_0^{d_1(x)/v} L(1,x;t)\,dt$$

$$= \int_0^{d_1(x)/v} [V(1,x;t) - \epsilon(l,x;t)]\,dt$$

$$= V(1,x) - \tfrac{1}{2}kA^2(1,x)$$

$$\times \int_0^{d_1(x)/v} \cos\left[2\left(\frac{k}{m}\right)^{1/2} t + \alpha\right] dt$$

$$= V(1,x) - \tfrac{1}{4}(mk)^{1/2}A^2(1,x)$$

$$\times \left\{\sin\left[2\left(\frac{k}{m}\right)^{1/2}\frac{d_1(x)}{v} + \alpha\right] - \sin(\alpha)\right\}$$

$$= V(1,x) - \tfrac{1}{4}(mk)^{1/2}A^2(1,x)(2)$$

$$\times \cos\left[\left(\frac{k}{m}\right)^{1/2}\frac{d_1(x)}{v} + \alpha\right]\sin\left[\left(\frac{k}{m}\right)^{1/2}\frac{d_1(x)}{v}\right]. \quad \text{(B1)}$$

Now the Lagrangian of path $(2,x)$ is, similarly,

$$L(2,x) = V(2,x) - \tfrac{1}{2}(mk)^{1/2}A^2(2,x)$$

$$\times \cos\left[\left(\frac{k}{m}\right)^{1/2}\frac{d_2(x)}{v} + \alpha\right]\sin\left[\left(\frac{k}{m}\right)^{1/2}\frac{d_2(x)}{v}\right]. \quad \text{(B2)}$$

We now want to compute the correlation between the two Lagrangians. Since both $A^2(1,x)$ and α are random variables, we will have to integrate over both of them.

The correlation between $L(1,x)$ and $L(2,x)$ is given by

$$\rho(x;1,2) = \frac{\mathrm{Cov}(L(1,x),L(2,x))}{[\mathrm{Var}(L(1,x))\mathrm{Var}(L(2,x))]^{1/2}}. \quad \text{(B3)}$$

Now

$$\mathrm{Var}(L(1,x)) = E(A^4(1,x))\tfrac{1}{4}(mk)\sin^2[Cd_1(x)]$$

$$\times \int_\alpha \cos^2[Cd_1(x) + \alpha]\, d\alpha$$

$$- E^2(A^2(1,x))\tfrac{1}{4}(mk)\sin^2[Cd_1(x)]$$

$$\times \left(\int_\alpha \cos[Cd_1(x) + \alpha]\, d\alpha\right)^2$$

where $C = (1/v)(k/m)^{1/2}$.

Also

$$\mathrm{Cov}(L(1,x),L(2,x))$$

$$= E(A^2(1,x)A^2(2,x))\left(\frac{mk}{4}\right)\sin[Cd_1(x)]\sin[Cd_2(x)]$$

$$\times \int_\alpha \cos[Cd_1(x) + \alpha]\cos[Cd_2(x) + \alpha]\, d\alpha$$

$$- E(A^2(1,x))E(A^2(2,x))\left(\frac{mk}{4}\right)\sin[Cd_1(x)]\sin[Cd_2(x)]$$

$$\times \int_\alpha \cos[Cd_1(x) + \alpha]\, d\alpha \int_\alpha \cos[Cd_2(x) + \alpha]\, d\alpha.$$

Since $\int_\alpha \cos[Cd_1(x) + \alpha]\, d\alpha = 0$, the variances and the covariances simplify considerably. We get

$$\rho(x;1,2) = \frac{E(A^2(1,x)A^2(2,x))\int_\alpha \cos[Cd_1(x) + \alpha]\cos[Cd_2(x) + \alpha]\, d\alpha}{\{E^2(A^2(1,x))E^2(A^2(2,x))\int_\alpha \cos^2[Cd_1(x) + \alpha]\, d\alpha \int_\alpha \cos^2[Cd_2(x) + \alpha]\, d\alpha\}^{1/2}}.$$

Now

$$\cos[Cd_1(x) + \alpha]\cos[Cd_2(x) + \alpha]$$
$$= \tfrac{1}{2}(\cos\{C[d_1(x) - d_2(x)]\}$$
$$+ \cos[C(d_1(x) + d_2(x)) + 2\alpha]).$$

When we integrate this expression over α, the second term vanishes and we get $\pi\cos\{C[d_1(x) - d_2(x)]\}$. When we integrate $\int_\alpha \cos^2[Cd_j(x) + \alpha]\, d\alpha$, we get π. Finally let a_0 be the correlation between $A^2(1,x)$ and $A^2(2,x)$ and suppose this correlation is independent of $d_1(x)$ and $d_2(x)$. Then we can write

$$\rho(x;1,2) = a_0\cos\{C[d_1(x) - d_2(x)]\},$$

where $C = (1/v)(k/m)^{1/2}$. (B4)

APPENDIX C

(A) We approximate $\sigma(x;j,k) = 1 - [1 - \rho(x;j,k)]^{1/2}$ linearly with $C\rho(x;j,k)$ where C is some constant. To choose C, we recall the original analytical formula for $p(x/1,2)$ in the case in which $p(x/1) = p(x/2)$. Thus we have

$$p(x/1,2) = p(x/1)(2)^{[1-\rho(x;j,k)]^{1/2}-1}.$$

Now our approximation formulas is

$$p(x/1,2) = p(x/1)[1 - \sigma(x;j,k)\ln 2].$$

The following table compares the two formulas at different values of $\rho(x;j,k)$:

$\rho(x;j,k)$	Exact formula	Approximation formula	% error
1	0.50	0.31	38%
0	1	1	0%
−1	1.33	1.29	3%

If possible, we would like to choose C so as to decrease the % error associated with $\rho(x;j,k) = 1$ without increasing the % error too much for $\rho(x;j,k) = -1$. We choose $C = 1/2\ln 2$. This formula gives 0.5 at $\rho(x;j,k) = 1$, 1 at $\rho(x;j,k) = 0$, and 1.5 at $\rho(x;j,k) = -1$. Its % error is 0%, 0%, and 13%. This is the formula which we use in the paper.

(B) Our task is to approximate $p(x/j)\ln[1 + p(x/k)/p(x/j)] + p(x/k)\ln[1 + p(x/j)/p(x/k)]$ by $\alpha[p(x/j)p(x/k)]^{1/2}$. First we define the ratio $r = p(x/k)/p(x/j)$ and rewrite our two formulas as: $p(x/j)\{\ln(1 + r) + r\ln[1 + (1/r)]\}$ and $\alpha p(x/j)(r)^{1/2}$. We choose α so that the two formulas are equal when $r = 1$. Hence $\alpha = 2\ln 2$. We now compare the difference and percent difference between the two formulas as we vary r.

r	$p(x/j)\{\ln(1+r)+r\ln[1+(1/r)]$		$\dfrac{\ln(1+r)+r\ln[1+(1/r)]-2\ln(2)(r)^{1/2}}{\ln(1+r)+r\ln[1+(1/r)]}$
1	0	$p(x/j)$	0%
$\frac{3}{2}$	(0.015)	$p(x/j)$	0.7%
$\frac{5}{3}$	(0.026)	$p(x/j)$	1.5%
2	(0.051)	$p(x/j)$	2.7%
3	(0.152)	$p(x/j)$	6.8%
4	(0.271)	$p(x/j)$	10.8%

Because of the symmetry of the formula, the error for $r = \nu$ is the same as the error for $r = 1/\nu$. Hence we only compute values of r equal to 1 or greater.

[1]P. Dirac, *The Principles of Quantum Mechanics* (Clarendon, Oxford, 1930).

[2]J. Von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1955).

[3]Stanley P. Gudder, *Stochastic Methods in Quantum Mechanics* (Elsevier, New York, 1979).

[4]A. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, New York, 1956).

[5]Bruno DeFinetti, *Theory of Probability* (Wiley, New York, 1970), Vol. I.

[6]Richard P. Feynman, Robert B. Leighton, and Matthew Sands, *The Feynman Lectures on Physics: Quantum Mechanics* (Addison-Wesley, Reading, MA, 1965).

[7]Alex Anas, "Discrete Choice Theory, Information Theory and the Multinomial Logit and Gravity Models," Transportation Research B **17**(1), 13 (1983).

[8]The random Lagrangian model is an application of the economic random utility models to physics. For a discussion of the relevance of economic theory to physics, see R. F. Bordley, "Reformulating Classical and Quantum Mechanics in Terms of a United Set of Consistency Conditions," Int. J. Theoret. Phys. **22**, 9 (1983).

[9]D. Bohm, Phys. Rev. **85**, 160 (1952).

[10]G. G. Comisar, "Brownian Motion Model of Nonrelativistic Quantum Mechanics," Phys. Rev. **138**, B 1332–7 (1965).

[11]P. Braffort, and C. Tzara, "Energie de l'oscillateur harmonique dans le vide," C. R. Acad. Sci. Paris **239**, 1779–80 (1954).

[12]E. Nelson, "Derivation of the Schrödinger Equation from Newtonian Mechanics," Phys. Rev. **150**, 1079–85 (1966).

[13]Richard P. Feynman and A.Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[14]Thomas A. Domencich and Daniel McFadden, *Urban Travel Demand* (North-Holland, Oxford, 1975).

[15]Nicolaos Litinas, "Simplified Methods for Transportation Policy Analysis: A Continuous Modelling Approach," Ph. D. thesis, MIT, 1980.

[16]These references look at the probability a variable u is the maximum of other variables. By a trivial transformation, we can make their results appropriate for finding the probability that a variable L is the minimum of other variables. This transformation underlies all the formulas presented in this paper.

[17]H. D. Block and J. Marschak, "Random Orderings and Stochastic Theories of Response," in *Contributions to Probability and Statistics*, edited by I. Olkin *et al.* (Stanford U.P., Stanford, 1960).

[18]Miles V. Klein, *Optics* (Wiley, New York, 1970).

[19]L. de la Peña-Auerbach, "A simple derivation of the Schrödinger equation from the theory of Markoff processes," Phys. Lett. A **24**, 603–4 (1967).

[20]L. de la Peña-Auerbach, "A new formulation of stochastic theory and quantum mechanics," Phys. Lett. A **27**, 594–5 (1968).

[21]L. de la Peña-Auerbach, E. Braun, and L. S. Garcia-Colin, "Quantum Mechanical Description of a Brownian Motion Particle," J. Math. Phys. **9**, 668–74 (1968).

[22]L. de la Peña-Auerbach, "Stochastic quantum mechanics for particles with spin," Phys. Lett. A **31**, 403–4 (1970).

[23]L. de la Peña-Auerbach and A. M. Cetto. "Self-interaction corrections in a nonrelativistic stochastic theory of quantum mechanics," Phys. Rev. D **3**, 795–800 (1971).

[24]Y. A. Rylov, "Quantum Mechanics as a Theory of Relativistic Brownian Motion," Ann. Phys. **27**, 1–11 (1971).

[25]W. Garczyński, "Stochastic Approach to Quantum Mechanics and to Quantum Field Theory," Acta Phys. Austr. Suppl. **6**, 501–17 (1969).

The geometric foundations of the integrability property of differential equations and physical systems. I. Lie's "function groups"[a)]

Robert Hermann
Association for Physical and Systems Mathematics, Inc., Brookline, Massachusetts 02146

(Received 12 August 1982; accepted for publication 15 April 1983)

This series of papers will attempt to discuss in a systematic way when the dynamical differential equations of a physical system have the "integrability" property. This first paper contains two topics: A description of some general properties of "function groups" and the related geometric structures of Poisson-cosymplectic manifolds; and the Lax representations for differential equations as a sort of "quantization" of Lie's "function groups." The general geometric setting of the "integrability" material in terms of the theory of Ehresmann pseudogroups is also described.

PACS numbers: 02.30.Hq, 02.20.Sv, 02.40. + m

1. INTRODUCTION

Current thought in physics is that there are two extreme types of differential equations (and physical systems): *integrable* and *chaotic*. Examples of the former are what Whittaker[1] calls the "soluble" problems of analytical mechanics, the free quantum fields (and certain simple types of one-dimensional interacting fields, like those associated with the sine–Gordon equation), and certain systems to which the Bäcklund or inverse scattering method applies. In the nature of things, it is much harder to give examples of systems which are at the other, "chaotic" extreme, and which can be analyzed precisely enough to formulate what should be meant mathematically by the very term "chaotic behavior." (The models of Lorenz[2] and Feigenbaum[3] are those which are of greatest current interest.) In this series of papers I will concentrate on the task of discussing various geometric structures which seem to underly the "integrability" property.

Now, in the post-1965 work on "integrability," the term mainly refers to the existence of some algorithms or procedures for finding a complete set of constants of motion and/or an explicitly *constructed* linearization (such as the "inverse scattering method"). In the 19th and early 20th century, Abel, Jacobi, Liouville, Lie, Fuchs, Darboux, Kowalewska, Painlevé, Picard, Schlesinger, Vessiot, and Drach were leading mathematicians who pursued various ramifications of this integrability idea. A key work in this tradition for what I want to do in this series is Lie's material on *function groups*.[4] (Of course, they are not "groups" in the modern sense, but Lie algebras.)

Weierstrass, Kowalewska, and Painlevé had a point of view[5,6] emphasizing certain "analyticity" properties of the general solutions. My aim in this series is to investigate systematically, with the tools of modern differential geometry, these two approaches and their interrelation, emphasizing calculus-on-manifolds, Lie group and algebra theory, Ehresmann pseudogroup-foliation-connection theory, Kodaira–Spencer deformation theory, and so on.

In the work of Kowalewska and Painlevé[5,6] "integrability" involved certain analyticity properties (local or global) of the *general* solution. Thus, the simplest condition of this type might be that the general solution could be written as the *quotient* of two functions which were analytic in the entire space of the independent variables and the initial conditions. Painlevé's 1900 paper[6] is a key one to keep in mind. He considered as an illustrative example the following two-parameter family of nonlinear ordinary differential equations:

$$\frac{d^2y}{dx^2} = ay^2 + bx\,, \tag{1.1}$$

where a and b are parameters, or, in terms of a 3×3 system,

$$\frac{dy_1}{dt} = 1, \quad \frac{dy_2}{dt} = y_3, \quad \frac{dy_3}{dt} = ay_2^2 + by_1\,. \tag{1.2}$$

He determines the values of a and b for which there are six analytic maps

$$\mathbb{C}\times\mathbb{C} \to \mathbb{C}^3\,,$$
$$N_1(t; z^0), \quad N_2(t; z^0), \quad N_3(t; z^0)\,,$$
$$D_1(t; z^0), \quad D_2(t; z^0), \quad D_3(t; z^0)\,,$$
$$t \in \mathbb{C}\,,$$
$$z^0 = (z_1^0, z_2^0, z_3^0) \in \mathbb{C}^3$$

such that

$$y_i(t) = \frac{N_i(t; z_0)}{D_i(t; z_0)}\,, \quad i = 1,2,3, \tag{1.3}$$

is the solution of (2.2), with

$$y_i(0) = z_i^0 \quad \text{for } i = 1,2,3. \tag{1.4}$$

Now, for certain values of the parameters (a, b), (1.2) can be solved in terms of "constants of motion" and elliptic functions, which can be interpreted as meaning that Lie's "function group" ideas can be applied. It is also noteworthy that Lorenz' "chaotic" equations[2] are of the same general algebraic type.

It is my goal in this series of papers to develop the differential geometric machinery to understand such examples from a modern point of view. Where background material is needed, but is not readily available in a straightforward form in the current literature, I will provide explanations in the

[a)] Supported by Ames Research Center (NASA), Grant NSG-2402; U. S. Army Research Office, Contract #ILI61102RH57-02 MATH; NSF MCS8003227.

 0022-2488/83/102422-11$02.50

text or appendices.

This first part concentrates on the "function groups." As pointed out in Refs. 7 and 8, the relevant contemporary mathematics is the theory of *Poisson* or *cosymplectic* structures[9] on manifolds. I will also present background material on Ehresmann pseudogroup theory[10] in the form developed by Plante.[11]

Many of the integrable mechanical and physical systems discussed in recent years are generalizations of the *rigid rotating body*; the key link here is a paper by Arnold.[12] The classic work of Sophie Kowalewska[5] on the analyticity–integrability properties of rigid bodies is an important clue that there are links between the Arnold–Lax and the analyticity approach. This series of papers will also explore the foundational aspects of this development. In this first paper I will deal primarily with Poisson structures on manifolds, which involve twice-covariant tensors on finite-dimensional manifolds. The underlying physics involves systems with a *finite number of degrees of freedom*. I intend in a later paper in this series to work on the field-theoretic generalizations, which will involve higher degree tensors.[13] Certain field theoretic models[14,15] involve Lie algebras which are infinite-dimensional versions of the "function groups" of Lie. These Lie algebras are also related to the "current algebras" of Gell-Mann and the Kac–Moody Lie algebras.

The key mathematical property providing "integrability" of these mechanical systems is the existence of a *Lax representation*.[16] A major theme of this paper is that the construction of such representations can be traced back to the "function groups" and associated geometric ideas, particularly the theory on Lie algebras of vector fields on manifolds and the work on pseudogroups by Ehresmann. I will show in this paper (as announced briefly in Ref. 17) that the search for a Lax representation for certain systems of differential equations can be regarded as a sort of "quantization" of Lie's "function groups."

2. POISSON OPERATIONS ON MANIFOLDS AND THEIR ASSOCIATED TENSOR FIELDS

Let X be an n-dimensional C^∞, paracompact manifold. $\mathscr{F}(X)$ denotes the C^∞ real-valued functions on X. Let

$$T(X) = \bigcup_{x\in X} X_x \tag{2.1}$$

be the *tangent vector bundle* and let

$$T^d(X) = \bigcup_{x\in X} X_x^d \tag{2.2}$$

be the dual bundle, called the *cotangent bundle*. The 1-differential forms $\mathscr{D}^1(X)$ on X are the cross sections of T^d, while the vector fields $\mathscr{V}(X)$ are the cross sections of $T(X)$. Let

$$T(X)\wedge T(X) \tag{2.3}$$

be the exterior product of two copies of the tangent bundle. Let

$$\Gamma(T'(X)\wedge T(X))$$

be its smooth cross sections. By the well-known principles of multilinear algebra, the elements of the fiber of the bundle (2.3) above a point $x\in X$ can be identified (and we shall do so) with the skew-symmetric, bilinear maps

$$\omega_x:\quad X_x^d \times X_x^d \to R\,.$$

A C^∞ cross section of the bundle $T^d(X)\wedge T^d(X)$ can be identified with an $\mathscr{F}(X)$-bilinear skew-symmetric map:

$$\omega:\quad \mathscr{D}^1(X)\times \mathscr{D}^1(X)\to \mathscr{F}(X)\,. \tag{2.4}$$

It will be called a *bivector field* on X.

Let such an ω be given. It defines a certain type of geometric structure. We will now describe in a coordinate-free way certain geometric concepts naturally attached to this structure that were treated in a tensor-analysis framework by Schouten[18] and Nijenhuis.[19]

Definition: Let ω be given as a cross section of the bundle (2.3) defining a map of the type indicated in (2.4). Let $\{\ ,\ \}_\omega$ be the map

$$\mathscr{F}(X)\times \mathscr{F}(X)\to \mathscr{F}(X)$$

defined as follows:

$$\{f_1,f_2\} = \omega(df_1, df_2) \quad \text{for } f_1,f_2\in\mathscr{F}(X)\,. \tag{2.5}$$

$\{\ ,\ \}$ is called the *Poisson operator* associated with the bivector field ω.

Remark: It will be called a *Poisson bracket* only if it satisfies the Jacobi identity, i.e., if the Schouten–Nijenhuis curvature tensor (which will be defined below) is identically zero.

Theorem 2.1: The Poisson operation $\{\ ,\ \}_\omega$ associated with $\omega\in\Gamma(T(X)\wedge T(X))$ satisfies the following identities:

$$\{f_1,f_2\}_\omega = -\{f_2,f_1\}_\omega \quad \text{for } f_1,f_2\in\mathscr{F}(X)\,, \tag{2.6}$$

$$\{f_1,f_2,f_3\}_\omega = \{f_1,f_2\}_\omega f_3 + f_2\{f_1,f_3\}_\omega$$

$$\text{for } f_1,f_2,f_3\in\mathscr{F}(X)\,. \tag{2.7}$$

Let us now be algebraic[19,20] and consider abstractly an R-bilinear operation $\{\ ,\ \}$ on $\mathscr{F}(X)$ satisfying (2.6) and (2.7). For each $f\in\mathscr{F}(X)$, set

$$V_f(f_1) = \{f,f_1\}\,. \tag{2.8}$$

(2.7) says that V_f is a *derivation* of $\mathscr{F}(X)$; it can then be identified with a *vector field* on the manifold X, i.e., a cross section of the tangent bundle $T(X)$.

Theorem 2.2: The mapping

$$f\to V_f$$

associated with the bivector structure on X satisfying (2.6)–(2.7) is a *first-order linear differential operator* from $\mathscr{F}(X)$ to $\mathscr{V}(X)$. The *symbol* of the operator (in the sense of Refs. 20 and 21) is a linear bundle map

$$\sigma:\quad T^d(X)\to T(X) \tag{2.9}$$

such that

$$\theta_1(\sigma(\theta_2)) = -\theta_2(\sigma(\theta_1)) \quad \text{for } \theta_1,\theta_2\in X_x^d,\ x\in X\,. \tag{2.10}$$

Proof: From (2.8) we have

$$V_{\{f_1,f_2\}} = f_1V_{f_2} + f_2V_{f_1}\,. \tag{2.11}$$

This derivation rule characterizes first-order linear differential operators, in the treatment of Chap. 1 of Ref. 20. As defined there, the symbol σ of this operator assigns to each $x\in X$, $\theta\in X_x^d$ an element $\sigma(\theta)$ of X_x. Thus, X_x^d is the dual vector space to the tangent vector space X_x. $\theta_1(\sigma(\theta_2))$, as it

appears in relation (2.9), is the value that the 1-covector θ_1 takes on the tangent vector $\sigma(\theta_2)$. The algebraic identity (2.10) now follows readily from (2.6).

Theorem 2.3: Let $\{\ ,\ \}$ be an R-bilinear map: $\mathcal{F}(X)\times\mathcal{F}(X)\to\mathcal{F}(X)$ satisfying (2.6)–(2.7). Then, there is a unique bivector field $\omega\in\Gamma(T(X)\wedge T(X))$ such that $\{\ ,\ \}$ is *associated with* ω *in the sense of formula (2.5).*

Proof: Set

$$\omega(\theta_1,\theta_2)=\theta_1(\sigma(\theta_2))\,, \tag{2.12}$$

where σ is as in (2.9), the symbol of the assignment $f\to V_f$. Formula (2.5) is now readily verified by tracing backwards from the definition.

3. THE SCHOUTEN–NIJENHUIS TENSOR ASSOCIATED WITH A BIVECTOR FIELD

Let $\{\ ,\ \}$ be a fixed bilinear differential operator: $\mathcal{F}(X)\times\mathcal{F}(X)\to\mathcal{F}(X)$ satisfying (2.6)–(2.7). For $f_1,f_2,f_3\in\mathcal{F}(X)$, set

$$\begin{aligned}\Omega(f_1,f_2,f_3)=&\{f_1\{f_2,f_3\}\}-\{\{f_1,f_2\},f_3\}\\&-\{f_2\{f_1,f_3\}\}\,.\end{aligned}\tag{3.1}$$

Notice that Ω is a trilinear, skew-symmetric differential operator:

$$\mathcal{F}(X)\times\mathcal{F}(X)\times\mathcal{F}(X)\to\mathcal{F}(X)\,.$$

It is identically zero if and only if $\{\ ,\ \}$ defines a Lie algebra operation on $\mathcal{F}(X)$.

For $f\in\mathcal{F}(X)$, let V_f be the vector field on X defined by

$$V_f(f_1)=\{f,f_1\}\quad\text{for } f_1\in\mathcal{F}(X)\,. \tag{3.2}$$

$f\to V_f$ is then a linear mapping of the vector space $\mathcal{F}(X)$ into the Lie algebra $\mathcal{V}(X)$. We want to find the conditions that it is a homomorphism of the algebra defined by $\{\ ,\ \}$ into the Lie algebra structure. To do this, consider $f_1,f_2,f_3\in\mathcal{F}(X)$:

$$\begin{aligned}[V_{f_1},V_{f_2}](f_3)&=V_{f_1}(V_{f_2}(f_3))-V_{f_2}(V_{f_1}(f_3))\\&=V_f\{f_2,f_3\}-V_{f_2}\{f_1,f_3\}\\&=\{f_1,\{f_2,f_3\}\}-\{f_2,\{f_1,f_3\}\}\,,\end{aligned}\tag{3.3}$$

$$V_{\{f_1,f_2\}}(f_3)=\{\{f_1,f_2\},f_3\}\,. \tag{3.4}$$

Theorem 3.1: Ω is identically zero if and only if the map $f\to V_f$ is a homomorphism of the algebra structure $\{\ ,\ \}$ defined on $\mathcal{F}(X)$ into the Lie algebra structure on $\mathcal{V}(X)$.

Proof: Notice that (3.3) and (3.4) give the formula

$$\Omega(f_1,f_2,f_3)=[V_{f_1},V_{f_2}](f_3)-V_{\{f_1,f_2\}}(f_3)\,. \tag{3.5}$$

This formula makes Theorem 3.1 evident.

Theorem 3.2: Ω is a skew-symmetric, first-order, homogeneous, trilinear differential operator. For fixed f_1,f_2 the map $f_3\to\Omega(f_1,f_2,f_3)$ of $\mathcal{F}(X)\to\mathcal{F}(X)$ is a derivation of the associative (i.e., pointwise-product) algebra structure on $\mathcal{F}(X)$.

Proof: This is also obvious from formula (3.5).

From general principles of the algebra of multilinear differential operators (again, refer to Chap. 1 of Ref. 20) one can now define the *symbol* of Ω, $\sigma(\Omega)$. For $x\in X$, it is a trilinear, skew-symmetric map

$$\sigma(\Omega)_x:\quad X_x^d\times X_x^d\times X_x^d\to R\,. \tag{3.6}$$

$\sigma(\Omega)_x$ is defined as follows. For $\theta_1,\theta_2,\theta_3\in X_x^d$, choose $f_1,f_2,f_3\in\mathcal{F}(X)$ such that

$$f_1(x)=f_2(x)=f_3(x)=0\,, \tag{3.7}$$

$$df_1(x)=\theta_1,\quad df_2(x)=\theta_2,\quad df_3(x)=\theta_3\,. \tag{3.8}$$

Then,

$$\sigma(\Omega)_x(\theta_1,\theta_2,\theta_3)=\Omega(f_1,f_2,f_3)(x)\,. \tag{3.9}$$

As x varies, $x\to\sigma(x)_x$ defines a tensor field on X, i.e., a cross section of the vector bundle

$$T(X)\wedge T(X)\wedge T(X)\,. \tag{3.10}$$

This tensor field is called the *Schouten tensor* of the bivector field ω.

We can sum up what we have proved as follows:

Theorem 3.3: Let $\omega\in\Gamma(T(X)\wedge T(X))$ be a bivector field on the manifold X. Then, its Schouten tensor $\Omega\in\Gamma(T(X)\wedge T(X)\wedge T(X))$ is zero if and only if the associated Poisson operator $\{\ ,\ \}_\omega$ on $\mathcal{F}(X)$ satisfies the Jacobi identity, i.e., defines a Lie algebra structure on $\mathcal{F}(X)$.

Theorem 3.4: Let ω be a bivector field. If each point x of X is contained in a coordinate system such that the components of ω in this coordinate system are *constant*, then the Schouten tensor Ω vanishes. In particular, the associated Poisson structure satisfies the Jacobi identity.

Proof: This follows from the "tensorial" property of Ω.

4. EHRESMANN PSEUDOGROUPS ON A FIXED MANIFOLD AND THEIR LINEAR ISOTROPY GROUPS

In the 1950s, Ehresmann developed[10] a major foundational concept for differential geometry, the *pseudogroup*. Unfortunately, he presented his ideas (which were abstracted from the classical work of Lie, Vessiot, and Cartan) only in short notes and conference proceedings; hence his ideas and theories have not penetrated directly to the physics and applied mathematics world. In this section I will recapitulate some of this material in a form that will be useful in the broad areas of physics and system theory, partially following ideas of Plante.[11] A key concept in Ehremann's work is the generalization of the "linear isotropy subgroup of a transformation group" concept to a pseudogroup.

Definition: Let X be a C^∞, paracompact manifold. A *local* C^∞ *map* is a triple

$$\delta=(D,\phi,R)$$

consisting of open subsets D, R of X, and a (C^∞) map ϕ: $D\to R$ mapping D *onto* R. D is the *domain* of δ, R the *range*. If ϕ is a *diffeomorphism* between D and R, we say that δ is a *local diffeomorphism* for X.

If $\delta=(D,\phi,R)$ is such a local map, with ϕ a diffeomorphism, define

$$\delta^{-1}=(R,\phi^{-1},D)\,, \tag{4.1}$$

where ϕ^{-1} is the inverse map to ϕ, a diffeomorphism from R to D.

If

$$\delta=(D,\phi,R),\quad \delta'=(D',\phi',R')\,,$$

are local maps, we say that

$$\delta\subset\delta' \tag{4.2}$$

if the following conditions are satisfied:

$$R \subset R', \quad D \subset D', \quad \delta = \delta' \text{ restricted to } D. \tag{4.3}$$

If $\delta = (D, \phi, R)$ and $\delta' = (D', \phi', R')$ are local maps, we say that

$$\delta \sim \delta' \tag{4.4}$$

if the following condition is satisfied:

$$\phi(x) = \phi'(x) \quad \text{for all } x \in D \cap D'. \tag{4.5}$$

If δ and δ' satisfy (4.4), we will define

$$\delta \cup \delta' = (D'', \phi'', R'') \tag{4.6}$$

as follows:

$$\begin{aligned} &D'' = D \cup D', \quad R'' = R \cup R'', \\ &\phi''(x) = \phi(x) = \phi'(x) \quad \text{for } x \in D \cap D', \\ &\phi''(x) = \phi(x) \quad \text{for } x \in D, \\ &\phi''(x) = \phi'(x) \quad \text{for } x \in D'. \end{aligned} \tag{4.7}$$

It is readily seen that these formulas define ϕ'' as a C^∞ map. However, of course, it may not be a diffeomorphism, even if ϕ and ϕ' are diffeomorphisms.

Suppose $\delta = (D, \phi, R)$ and $\delta'(D', \phi', R')$ are local diffeomorphisms. Define another local diffeomorphism

$$\delta \circ \delta' = (D'', \phi'', R'') \tag{4.8}$$

as follows:

$$D'' = \phi'^{-1}(D \cap R'), \tag{4.9}$$

$$R'' = \phi(D \cap R'), \tag{4.10}$$

$$\phi''(x) = \phi(\phi'(x)) \quad \text{for } x \in D''. \tag{4.11}$$

Definition: A collection Δ of local diffeomorphisms of the manifold X is said to be an *Ehresmann pseudogroup acting on X* if the following conditions are satisfied:

(1) The identity map

$$1 = (X, \mathrm{id}, X)$$

is an element of Δ.

(2) If $\delta \in \Delta$, then $\delta^{-1} \in \Delta$.

(3) If $\delta, \delta' \in \Delta$, then

$$\delta \circ \delta' \in \Delta.$$

(4) If $\delta' \in \Delta$ and $\delta \subset \delta'$, then $\delta \in \Delta$.

(5) If $\delta, \delta' \in \Delta$ with $\delta \sim \delta'$ and if $\delta \cup \delta'$ is a local diffeomorphism, then $\delta \cup \delta' \in \Delta$.

The most important pseudogroups in both geometry and physics are those defined via tensor fields. A *tensor field* τ on a manifold X is a smooth cross section of a tensor product bundle of the tangent bundle $T(X)$ and its dual $T^d(X)$. [Alternately, it may be considered as an associated bundle, in the sense of Steenrod,[22] of the principal tangent bundle with structure group GL(n, R), defined by a linear action of GL(n, R) on a real vector space.]

A local diffeomorphism $\delta = (D, \phi, R)$ is then said to be a *symmetry* (or automorphism) of τ if, for each $x \in D$, the linear map $\phi_*: X_x \to X_{\phi(x)}$, acting on tensors, sends the value of τ at x to the value of τ at $\phi(x)$. It is readily seen that the set of such symmetries of τ defines an Ehresmann pseudogroup, as defined above. We will call such a pseudogroup a *tensorial pseudogroup*.

The pseudogroups encountered in physics and geometry often have a special property, that one calls "flatness."

Definition: A tensor field τ on a manifold X is said to be *locally flat* if the following condition is satisfied:

Each point x has a neighborhood U and a coordinate system (x^i) defined on U such that the components of τ in U with respect to the bases of tensor spaces constructed from $\partial/\partial x^i$ and dx^i [which are cross sections of $T(U)$ and $T^d(U)$] are *constant* in U.

We can also consider relations between Ehresmann pseudogroups and sets of vector fields on X. Let V be such a vector field. Let $t \to \exp(tV)$ be the one-parameter pseudogroup it generates, i.e., the collection of local diffeomorphism: $D \to R$ obtained by finding the orbit curves of V, starting at points $x \in D$, then going out t time units.

Definition: V is associated with the Ehresmann pseudogroup if each local diffeomorphism generated by the orbit curves of V belong to D.

Standard results prove the following:

Theorem 4.1: Let τ be a tensor field on a manifold X and let $\mathcal{G}(\tau)$ be the Lie algebra of vector fields $V \in \mathcal{V}(X)$ such that

$$\mathcal{L}_V(\tau) = 0 \tag{4.12}$$

($\mathcal{L}$ denotes "Lie derivative.") Then, the pseudogroup of symmetries of τ contains the one-parameter pseudogroup generated by the vector fields in $\mathcal{G}(\tau)$.

Now, for the "linear isotropy subgroup" concept.

Definition: Let $\mathcal{P} = (D, \phi, R)$ be a pseudogroup on a manifold X. Let x be a point of X. Let $(D, \phi, R) \in \mathcal{P}$ be the set of elements of pseudogroups such that the following conditions are satisfied:

$$x \in D \cap R, \quad \phi(x) = x. \tag{4.13}$$

Associate with (D, ϕ, R) the linear map

$$\phi_*: X_x \to X_x.$$

The collection of linear maps on the tangent space X_x, which, it is readily seen, forms a group, is called the *linear isotropy group of the pseudogroup* $\mathcal{P}$ *at* x.

This concept unifies many special situations encountered in classical and modern differential geometry and differential equation theory.

5. PSEUDOGROUPS AND POISSON TENSORS

We can see the immediate relevance of these concepts to the study of Poisson tensors ω, and the associated Poisson operators $\{\ ,\ \}$. Let ω be a bivector field, an element of $\Gamma(T(X) \wedge T(X))$; $\{\ ,\ \}$ is an R-bilinear map $\mathcal{F}(X) \times \mathcal{F}(X) \to \mathcal{F}(X)$. The *Schouten tensor* of ω, Ω, is an element of $\Gamma(T(X) \wedge T(X) \wedge T(X))$. Let $\mathcal{G}(\omega)$ be the Lie algebra of vector fields $V \in \mathcal{V}(X)$ such that

$$\mathcal{L}_V(\omega) = 0, \tag{5.1}$$

i.e.,

$$\begin{aligned} V(\omega(\theta_1, \theta_2)) &= \omega(\mathcal{L}_V(\theta_1), \theta_2) \\ &\quad + \omega(\theta_1, \mathcal{L}_V(\theta_2)) \quad \text{for all } \theta_1, \theta_2 \in \mathcal{D}^1(X). \end{aligned} \tag{5.2}$$

We will state a few typical results, which are readily proved using the general geometric principles proved above.

Theorem 5.1: If ω is flat, then $\Theta = 0$, and $\{\ ,\ \}$ satisfies the Jacobi identity, and makes $\mathscr{F}(X)$ into a Lie algebra.

Theorem 5.2: A vector field V is a symmetry of the bivector field if and only if it satisfies the following condition:

$$V(f_1, f_2) = V(f_1), f_2 + f_1, V(f_2) \quad \text{for } f_1, f_2 \in \mathscr{F}(X), \tag{5.3}$$

i.e., Lie derivation by V is a *derivation* of the algebraic operation on $\mathscr{F}(X)$.

Proof: That (5.3) follows from (5.2) is a routine derivation, left to the reader.

For the converse, suppose that (5.3) is satisfied. Note that this proves that

$$\mathscr{L}_V(\omega)(df_1, df_2) = 0 \quad \text{for all } f_1, f_2 \in \mathscr{F}(X).$$

But that $\mathscr{L}_V(\omega)$ vanishes follows from the *tensorial* property of $\mathscr{L}_V(\omega)$.

Theorem 5.3: If $\Omega = 0$, if $f \in \mathscr{F}(X)$, and if V_f is the vector field defined by $V_f(f') = \{f, f'\}$, then V_f satisfies

$$\mathscr{L}_{V_f}(\omega) = 0.$$

In particular, the one-parameter pseudogroup $t \to \exp(tV_f)$ belongs to the pseudogroup of all symmetries of ω.

6. SINGULAR FOLIATIONS AND THE FROBENIUS INTEGRABILITY THEOREM: IMPLICATIONS FOR THE POISSON STRUCTURE

I will now recapitulate work done in singular foliation theory. Let X be a manifold, and let $\mathscr{W}$ be a linear subspace of $\mathscr{V}(X)$. For $x \in X$, let $\mathscr{W}(X)$ be the linear subspace $\{V(x): V \in \mathscr{W}\}$ of values of $\mathscr{W}$ at x. A continuous, piecewise C^∞ curve $t \to x(t)$, $a \leqslant t \leqslant b$, is said to be an *orbit curve* of $\mathscr{W}$ if the following conditions are satisfied:

$$\frac{dx}{dt}(t) \in \mathscr{W}(x(t)) \quad \text{for } a \leqslant t \leqslant b. \tag{6.1}$$

(dx/dt denotes the tangent vector field to the curve.) For $x_0 \in X$, let $C(\mathscr{W}, x_0)$ be the *accessible set* from x_0 along orbit curves of $\mathscr{W}$, the set of points of X which can be joined to x_0 by an orbit curve. (The letter "C" is in honor of Caratheodory[23] and Chow,[24] who started this development.)

Theorem 6.1: The subsets $C(\mathscr{W}, x_0)$ of X, as x runs through X, define an equivalence relation on X for which $C(\mathscr{W}, x)$ are the equivalence classes. Each such set can be given the structure of an immersed submanifold.

For proof and background, see Refs. 25–27.

Suppose now that $\mathscr{W}$ is a Lie subalgebra of $\mathscr{V}(X)$, i.e.,

$$[\mathscr{W}, \mathscr{W}] \subset \mathscr{W}. \tag{6.2}$$

In general, one does not know that the accessible submanifolds $C(\mathscr{W}, x)$ are integral submanifolds of the tangent vector distribution $x \to \mathscr{W}(x)$. The following result was proved in Ref. 25.

Theorem 6.2: Suppose that (6.1) and an additional following condition is satisfied:

> For each orbit curve $t \to x(t)$ of $\mathscr{W}$, the dimension of the tangent vector space $\mathscr{W}(x(t))$ is constant as t varies. (6.3)

Then, the submanifolds $C(\mathscr{W}, x_0)$ are maximal integral submanifolds of the singular foliation $\mathscr{W}$ in the sense that the tangent space of $C(\mathscr{W}, x_0)$ at each point $x \in C(\mathscr{W}, x_0)$ is equal to $\mathscr{W}(x)$. Further, (6.3) is satisfied if either X and $\mathscr{W}$ are *real analytic*, or $\mathscr{W}$ is *locally finitely generated* in the sense defined in Ref. 25.

There are evident implications for this basic theorem to the study of curvature-zero Poisson structures and the Ehresmann pseudogroups they generate. Let $\{\ ,\ \}$ be such an operation on $\mathscr{F}(X)$. For $f \in \mathscr{F}(X)$ let V_f be the vector field on X such that

$$V_f(f') = \{f, f'\} \quad \text{for } f \in \mathscr{F}(X). \tag{6.4}$$

Then, $f \to V_f$ is a Lie algebra homorphism and

$$\mathscr{G} = \{V_f: f \in \mathscr{F}(X)\}$$

is a Lie subalgebra of $\mathscr{V}(X)$. Hence, Theorems 5.1 and 5.2 apply. $\mathscr{G}$ defines a *singular foliation* of X, i.e., a decomposition into submanifolds. The Lie algebra of vector fields $\mathscr{G}$ (a the pseudogroup it generates) are tangent to the leaves of this foliation.

7. HOMOMORPHISMS OF BIVECTOR FIELDS AND FUNCTION GROUPS IN THE SENSE OF SOPHUS LIE

The Poisson bracket operation of analytical mechanics is usually defined by means of a closed 2-differential form, i.e., a twice-*covariant* tensor field, a cross section of the vector bundle $T^d(X) \wedge T^d(X)$. The case where the manifold X is even-dimensional and the form is of maximal rank is the traditional one, treated in all the modern treatises on mechanics. The case where the form does not have maximal rank is also interesting, and was first treated systematically in Ref. 13. In this case, the Poisson bracket cannot be defined on *all* functions on X, but only on a subalgebra. This "covariant" formalism can also be extended[13] to higher degree differential forms, and thereby to field theories.

Now, covariant tensor fields have certain properties relative to C^∞ mappings between the manifolds on which they live: They "pull back" dually to the mapping. (This is the meaning in differential geometry of "covariant"!) The Poisson structures arise from contravariant tensor fields, which "push forward." However, the push-forward map cannot be defined for an arbitrary *tensor field*.

To elaborate algebraically, let X and X' be C^∞ paracompact manifolds (possibly of different dimension) and ϕ: $X \to X'$ a C^∞ map. At each point $x \in X$, the differential ϕ_* maps the tangent vector space X_x linearly to the tangent space $X'_{\phi(x)}$. This leads to a linear bundle map ϕ_*: $T(X) \to T(X')$ and the following commutative mapping diagram:

$$\begin{array}{ccc} \phi_*: T(X) & \to & T(X') \\ \downarrow & & \downarrow \\ \phi: X & \to & X' \end{array}$$

Consider the cross sections $\Gamma(T(X))$ and $\Gamma(T(X'))$. They do not map naturally under ϕ_*: Given a $V \in \Gamma(T(X))$, $\phi_*(V)$: $X \to \phi_*(V(X))$ cannot be defined naturally as a cross section of $T(X')$ because a fiber $X'_{x'}$ may arise as the image under ϕ_* of *two* fibers X_{x_1} and X_{x_2}. However, one can impose an extra condition that $\phi_*(V)$ be well defined as a cross section of $T(X')$. Let us say that V and V' are ϕ-*related* if

$$\phi_*(V(x)) = V'(\phi(x)) \quad \text{for all } x \in X, \tag{7.1}$$

i.e., if the following diagram of maps is commutative:

$$\begin{array}{ccc} T(X) & \xrightarrow{\phi_*} & T(X') \\ \uparrow V & & \uparrow V' \\ X & \xrightarrow{\phi} & X' \end{array} \tag{7.2}$$

If V and V' satisfy (7.1), then it is readily seen that *ϕ maps orbit curves of V into orbit curves of V', i.e., ϕ is an intertwining map for the one-parameter pseudogroup of diffeomorphisms generated by V and V'.* In this geometric form, the ϕ-related vector fields play a basic role in the Lie–Cartan geometric theory of differential equations. (They are "prolongation maps" for the underlying differential equations.)

Thus, the geometric relation between contravariant tensor fields on X and X' can only be considered as involving special pairs (τ, τ') which are ϕ-related via (7.1)–(7.2). We can do the same for bivector fields.

Definition: Let ϕ: $X \to X'$ be a C^∞ map between manifolds, and let $\omega \in \Gamma(T(X) \wedge T(X))$ and $\omega' \in \Gamma(T(X') \wedge T(X'))$ be bivector fields on X and X'. They are said to be *ϕ-related*, and we write

$$\phi_*(\omega) = \omega', \tag{7.3}$$

if the following condition is satisfied:

$$\omega(\phi^*(\theta'_1), \phi^*(\theta'_2)) = \omega'(\theta'_1, \theta'_2)$$
$$\text{for all } x \in X, \text{ all } \theta'_1, \theta'_2 \in X'_{\phi(x)}. \tag{7.4}$$

The following results are easily proved.

Theorem 7.1: Let $\{\ ,\ \}_\omega$ and $\{\ ,\ \}'_{\omega'}$ be the operations on $\mathcal{F}(X)$ and $\mathcal{F}(X')$ defined by the bivector fields ω and ω'. Then (7.4) is satisfied, i.e., $\phi_*(\omega) = \omega'$, if and only if the following condition is satisfied:

$$\phi^*(\{f'_1, f'_2\}_{\omega'}) = \{\phi^*(f'_1), \phi^*(f'_2)\}_\omega$$
$$\text{for all } f'_1, f'_2 \in \mathcal{F}(X'). \tag{7.5}$$

Theorem 7.2: If ω and ω' are ϕ-related, so are their curvature tensors Ω and Ω'. (This says that the differential operator $\omega \to \Omega$ is a "natural" operation.)

8. HOMOMORPHISMS OF POISSON STRUCTURES AND FUNCTION GROUPS IN THE SENSE OF SOPHUS LIE

Let X continue as a manifold with ω a bivector field defining a Poisson operation

$$(f_1, f_2) \to \{f_1, f_2\}$$

on $\mathcal{F}(X)$.

Definition: A set $f_1, \ldots, f_m$ of functions on X is said to form a *function group*, in the sense of Lie, relative to the Poisson tensor ω, if and only if there are C^∞ functions

$$F_{ij}: R^m \to R, \quad 1 \leqslant i, j \leqslant m,$$

such that

$$\{f_i, f_j\}_\omega(x) = F_{ij}(f_1(x), \ldots, f_m(x)) \quad \text{for all } x \in X. \tag{8.1}$$

Remark: The terminology "function group" is obviously archaic—they are not *groups* in the sense we use the term (except trivially, in the additive structure). However, what Lie usually called a "group" we would call a "Lie algebra." Hence, an appropriate modern name might be *Lie function algebras.*

Theorem 8.1: Suppose that ϕ: $x \to (f_1(x), \ldots, f_m(x))$ defines submersion maps from X to R^m. Suppose D is the image in R^m of this submersion; it is, of course, an open subset of R^m. Then, $(f_1, \ldots, f_m)$ form a function group if and only if there is a bivector field ω' on D' such that

$$\phi_*(\omega) = \omega'.$$

The *proof* is given in Ref. 7.

9. LIE "FUNCTION GROUPS" GENERATED BY LIE ALGEBRAS OF VECTOR FIELDS

Let us first recall the definition of a "symplectic structure."

Definition: Let X be a manifold. A *symplectic structure* on X is defined by a 2-differential form η which satisfies the following conditions:

$$d\eta = 0, \tag{9.1}$$

$$v \rfloor \eta = 0 \quad \text{for } v \in T(X) \Rightarrow v = 0, \tag{9.2}$$

i.e., η has no nonzero (Cauchy) characteristic vectors.

A symplectic form η defines a Poisson bracket operation on $\mathcal{F}(X)$. For $f \in \mathcal{F}(X)$, define $V_f \in \mathcal{V}(X)$ as follows:

$$df = V_f \rfloor \eta.$$

Then set

$$\{f, f_1\} = -V_f(f_1). \tag{9.3}$$

Theorem 9.1: There is a bivector field ω on X with zero Schouten tensor, which gives rise to the Poisson bracket operation (9.2). Algebraically, ω is the dual tensor to the (nonsingular) η. In local coordinates (x^i), $1 \leqslant i, j \leqslant n$, if $\eta = \eta_{ij}\, dx^i \wedge dx^j$, then

$$\omega = \eta^{ij} \frac{\partial}{\partial x^i} \wedge \frac{\partial}{\partial x^j}, \tag{9.4}$$

where (η^{ij}) is the inverse matrix to (η_{ij}) [which exists because of condition (9.2)].

Now, let Q be a manifold and let

$$X = T^d(Q) \tag{9.5}$$

be its cotangent bundle. Let $V \to f_V$, $\mathcal{V}(Q) \to \mathcal{F}(X)$ be the map defined as follows:

$$f_V(\theta) = \theta(V(q)) \quad \text{for } V \in \mathcal{V}(Q), \quad \theta \in Q_q, \ q \in Q. \tag{9.6}$$

Let η be the canonical symplectic form on $T^d(Q) \equiv X$, and let $\{\ ,\ \}$ be the corresponding Poisson bracket on $\mathcal{F}(X)$. (It is the standard Poisson bracket used in analytical mechanics when Q is the configuration space manifold of the mechanical system.)

Theorem 9.2: For $V_1, V_2 \in \mathcal{V}(Q)$, then

$$\{f_{V_1}, f_{V_2}\} = f_{[V_1, V_2]}, \tag{9.7}$$

i.e., the mapping $V \to f_V$ is a Lie algebra homomorphism from $\mathcal{V}(Q)$ to the Lie algebra $\mathcal{F}(T^d(Q))$.

Proof: Well known.

Suppose now that $\mathcal{G}$ is a finite-dimensional Lie algebra

of vector fields on Q. Let

$$(V^a), \quad 1 \leqslant a, b, c \leqslant m ,$$

be a basis for $\mathscr{G}$. It satisfies relations of the following form:

$$[V^a, V^b] = \lambda_c^{ab} V^c , \tag{9.8}$$

where (λ_c^{ab}) are the *structure constants* of the Lie algebra.

By Theorem 9.2, the functions $f^a \equiv f_{V^a}$ on $T^d(Q) = X$ satisfy the same Poisson-bracket relations:

$$\{ f^a, f^b \} = \lambda_c^{ab} f^c . \tag{9.9}$$

Theorem 9.3: Let D be an open subset of X on which the differentials

$$df^1, \ldots, df^m$$

have constant rank. Let ϕ: $D \to R^m$ be the following map:

$$x \to (f^1(x), \ldots, f^m(x)) .$$

Then, there is a Poisson structure on the submanifold $\phi(D)$ of R^m. If the indices are relabelled so that the $df^1, \ldots, df^m$ are a maximal linearly independent set among the (df^a) on D, then the $(f^1, \ldots, f^n)$ form a function group in the sense of Lie.

Proof: The $f^{m+1}, \ldots, f^m$ can be written locally as functions of the $f^1, \ldots, f^n$.

We temporarily leave the general theory of function group–Poisson–cosymplectic structures in order to make differential-geometrically precise one version of what is meant (in the mathematical-physics–applied mathematics literature) by a "Lax representation" of a system of differential equations.

10. THE LAX REPRESENTATION OF VECTOR FIELDS

Let us begin with one of the many possible definitions (in the context of differential geometry) of what is meant by the "Lax representation."

Let X be a C^∞, paracompact finite-dimensional manifold. Let $V \in \mathscr{V}(X)$ be a C^∞ vector field on X. $\mathbf{x}$: $t \to x(t)$, $a \leqslant t \leqslant b$, is a (continuous piecewise C^∞) curve in X. $\mathrm{x} \to (\mathrm{d}\mathbf{x}/dt)(t) \in X_{x(t)}$ denotes its tangent vector field. $\mathbf{x}$ is an *orbit curve* (or "integral curve") of V if

$$\frac{d\mathbf{x}}{dt}(t) = V(\mathbf{x}(t)) \quad \text{for } a \leqslant t \leqslant b .$$

Definition: Let W be a real vector space, and let $L(W)$ denote the vector space of linear maps: $W \to W$. A *Lax representation* for the vector field $V \in \mathscr{V}(X)$ is a C^∞ map α: $X \to L(W) \times L(W), \alpha(x) = (A(x), B(x))$ such that the following condition is satisfied:

For every orbit curve $t \to x(t)$ of the vector field V, the curve

$$t \to \alpha(x(t)) = (A(x(t)), B(x(t)))$$

in $L(W) \times L(W)$ satisfies the following differential equations:

$$\frac{d}{dt} A(x(t)) = [B(x(t)), A(x(t))] , \tag{10.1}$$

where $[\ ,\ ]$ denotes the commutator bracket in $L(W)$.

There are now many explicit examples of such representations in the physics and applied mathematics literature, and suggestions that it might be possible to extend the ideas to infinite-dimensional manifolds and vector spaces. (Of course, the original Lax work,[16] based on the inverse scattering method of Gardner *et al.*[28] was in such an infinite-dimensional context, but Lax did not attempt to fit it into a precise framework of infinite-dimensional manifolds.)

Restricting attention to the finite-dimensional case, I will now describe how the basic geometry of Lie algebras of vector fields on manifolds may be used to construct Lax representations.

11. LAX REPRESENTATIONS CONSTRUCTED FROM LIE ALGEBRAS OF VECTOR FIELDS

Let X continue as a manifold with $\mathscr{G}$ a real Lie subalgebra of $\mathscr{V}(X)$, the Lie algebra of all vector fields on X.

For $x \in X$, let

$$\mathscr{G}(x) = \{ V(x) : V \in \mathscr{G} \} \subset X_x . \tag{11.1}$$

A curve $\mathbf{x}$: $t \to x(t)$, $a \leqslant t \leqslant b$, is an *orbit curve* of $\mathscr{G}$ if

$$\frac{d\mathbf{x}}{dt} \in \mathscr{G}(\mathbf{x}(t)) \quad \text{for } a \leqslant t \leqslant b . \tag{11.2}$$

Let

$$C(\mathscr{G}, x) = \text{set of points which can be reached starting at } x \text{ along a continuous, piecewise } C^\infty \text{ orbit cuve of } \mathscr{G} .$$

As we recalled in Sec. 6, $C(\mathscr{G}, x)$ can be given the structure of an immersed submanifold of X, and

$$C(\mathscr{G}, x)_y \supset \mathscr{G}(y) \quad \text{for } y \in C(\mathscr{G}, x) . \tag{11.3}$$

[$C(\mathscr{G}, x)_y$ denotes the tangent space to the manifold $C(\mathscr{G}, x)$ at the point y, identified with a linear subspace of X_y.] The pseudogroup generated by $\mathscr{G}$ maps each $C(\mathscr{G}, x)$ into itself.

Now, let V be a vector field on X such that

$$V(x) \in \mathscr{G}(x) \quad \text{for all } x \in X . \tag{11.4}$$

Let W be a real finite-dimensional vector space with $L(W)$ the vector space of linear maps: $W \to W$. Let

$$\phi: X \to L(W), \quad \beta: \mathscr{G} \to L(W)$$

be maps, with β *linear*.

$L(W)$ is, of course, a vector space. For $\alpha \in L(W)$, identify then the tangent space $L(W)_\alpha$ with $L(W)$. With this understanding, suppose the following condition is satisfied: There is a Lie algebra homomorphism β: $\mathscr{G} \to L(W)$ such that, for each $x \in X$, each $B \in \mathscr{G}$,

$$\phi_*(B(x)) = \beta(B) \in L(W) . \tag{11.5}$$

Theorem 11.1: With the conditions stated above, (ϕ, β) defines a Lax representation for the differential equations of the orbit curves of the vector field V.

Proof: Our hypotheses imply that there is a curve $t \to B(t)$ in $\mathscr{G}$ such that

$$\frac{dx}{dt} = B(t)(x(t)) \quad \text{for } a \leqslant t \leqslant b . \tag{11.6}$$

Then, (11.4)–(11.5) imply that

$$\frac{d}{dt}(\phi(x(t))) = [\beta(B(t)), \phi(x(t))] , \tag{11.7}$$

which is precisely what is meant by a "Lax representation."

Having perceived a geometric mechanism for generating Lax representations of dynamical systems, we now turn to investigation of the geometric-physical structure for mechanical systems that will generate the conditions required for a Lax representation.

12. GENERAL REMARKS ABOUT THE GEOMETRIC MEANING OF "QUANTIZATION" OF PHYSICAL SYSTEMS

Let us summarize what the mathematical physicists of the period 1925–1930[29–31] meant by "quantization," and the crucial contributions made later by Moyal[32] and van Hove.[33]

Let X be a manifold with a Poisson bracket structure $\{\ ,\ \}$ on the space $\mathcal{F}(X)$ of C^∞, real-valued functions on X. Then endow $\mathcal{F}(X)$ with both a Lie algebra structure (the $\{\ ,\ \}$ bracket) and a commutative, associative algebra structure (pointwise product). Dirac's original idea of what "quantization" means can be put into the following terms: Suppose a physical system prescribes a linear subset $\mathcal{O}$ of $\mathcal{F}(X)$, the distinguished "observables." Form the subset of $(f_1, f_2) \in \mathcal{O} \times \mathcal{O}$ such that

$$\{f_1, f_2\} \in \mathcal{O}\ . \tag{12.1}$$

Find a vector space V and a linear mapping

$$Q\colon\ \mathcal{O} \to L(V)$$

such that the following condition is satisfied:

$$Q(f)^2 = Q(f)^2 \quad \text{for all } f \in \mathcal{O}\ , \tag{12.2}$$

$$Q(\{f_1, f_2\}) = Q(f_1)Q(f_2) - Q(f_2)Q(f_1)$$

$$\text{for all } (f_1, f_2) \text{ satisfying (12.1)}\ . \tag{12.3}$$

Dirac's work did not treat definitively the mathematical question of whether "quantization" could be regarded as a Lie algebra representation of the Lie algebra defined by the Poisson bracket structure on the real-valued C^∞ functions of the classical positions and momenta, although that, of course, was an obvious question after his work. The monograph by van Hove[33] was a historically crucial clarification of this question: For the case of $X = R^{2n}$, with $\{\ ,\ \}$ the usual Poisson bracket structure of analytical mechanics, he showed that the usual "quantization" rules for Newtonian particle systems cannot be extended on a Lie algebra representation of all polynomial functions on X. As an alternative, he constructed a representation of this infinite-dimensional Lie algebra using the *contact* structure on R^{2n+1}. In a geometrically related work, Boothby and Wang[34] classified globally certain of these contact structures. Putting these two basic papers together led to the extensive "geometric quantization" literature of today.[35–37]

If a "quantization" cannot mean (in general) a Lie algebra representation of the Poisson bracket Lie algebra, what does it mean? Weyl and Moyal gave what seems the reasonable partial answer. In its general form[38] it goes as follows:

Let $\mathcal{A}$ be a linear subspace of $\mathcal{F}(X)$, satisfying the following condition:

> For $f_1, \ldots, f_m \in \mathcal{A}$ which are linearly independent, there is a nonzero *polynomial* $P(\ ,\ldots,\ )$ in m variables such that $P(f_1, \ldots, f_n) = 0$. (In other words, the elements of $\mathcal{A}$ are algebraically independent as functions on X.) (12.4)

Let $\mathcal{P}(\mathcal{A}) \subset \mathcal{F}(X)$ be the algebra of functions on X which are polynomials in the element of $\mathcal{A}$. Let

$$\rho\colon\ \mathcal{A} \to L(V)$$

be a linear mapping of $\mathcal{A}$ to the space of linear maps on a vector space V. Extend ρ to a mapping

$$\rho\colon\ \mathcal{P}(\mathcal{A}) \to L(V)$$

as follows:

$$\rho(f_1, \ldots, f_m) = \frac{1}{m!} \sum_{(i_1 \cdots i_m)} \rho(f_{i_1}) \cdots \rho(f_{i_m})$$

$$\text{for } m = 1, 2, \cdots, \tag{12.5}$$

where the sum in (12.5) extends over all permutations of $(1, \ldots, m)$. [Thus, (12.5) represents "quantization by complete symmetrization."] One can now carry back the Lie algebra structure on $L(V)$ (i.e., operator commutator) to define a Lie algebra structure on $\mathcal{P}(\mathcal{A})$. If $X = R^{2n}$, with $\mathcal{A}$ the linear functions on X, and ρ (depending on a parameter $\hbar$, Planck's constant), the usual Schrödinger operators ($V = C^\infty$, rapidly decreasing functions on R^n), then the operator bracket pulled back to $\mathcal{P}(\mathcal{A})$ is the Weyl–Moyal bracket.[30,32] It depends on the parameter $\hbar$ and reduces, for $\hbar = 0$, to a classical Poisson bracket. It defines a *deformation* of the Poisson bracket Lie algebra.

Thus, we see that in one formulation "quantization" is algebraically a process of extension of linear maps by some process of *symmetrization*. Thus, it is not surprising that it has a close relation to *invariant theory*. (I pointed out in Ref. 38 that operations occurring in the Weyl–Moyal theory were the *transvections* of classical invariant theory.) It is this algebraic aspect of "quantization" that I want to implicate in the search for *Lax representations* of physical systems.

13. PHYSICAL SYSTEMS GENERATED BY FINITE-DIMENSIONAL LIE SUBALGEBRAS OF A POISSON STRUCTURE

Suppose the manifold X has a Poisson structure $\{\ ,\ \}$. Let n be an integer, and let $i, j, k, \cdots$ be indices running over 0 to n with the summation convention. Suppose that

$$(f^i) \in \mathcal{F}(X)$$

are a set of functions on X such that conditions (13.1)–(13.3) are satisfied:

$$f^0(x) = 1 \quad \text{for all } x \in X\ . \tag{13.1}$$

The $f^1, \ldots, f^n$ are algebraically independent, i.e., there is no nonconstant polynomial function $P\colon\ R^n \to \mathbb{C}$ such that

$$P(f^1, \ldots, f^n) \equiv 0\ , \tag{13.2}$$

$$\{f^i, f^j\} = \lambda^{ij}_k f^k\ , \tag{13.3}$$

where (λ^{ij}_k) are real numbers. (They satisfy the Jacobi identity, i.e., are the structure constants of a finite-dimensional Lie algebra.)

Let $\mathcal{A}$ be the smallest associative subalgebra of $\mathcal{F}(X)$ containing the f^i. $\mathcal{A}$ is then also a Lie subalgebra of $\mathcal{F}(X)$

[because of condition (13.3)].

Let h be an element of $\mathscr{A}$. Construct the vector field V_h, with

$$V_h(f) = \{h, f\} \quad \text{for all } f \in \mathscr{F}(X). \tag{13.4}$$

Consider h as the Hamiltonian of a physical system (examples will be given below). Let $\mathscr{G}$ be the finite-dimensional Lie subalgebra of $\mathscr{F}(X)$ spanned by the f^i.

Theorem 13.1: V_h is tangent to the orbit submanifolds of $\mathscr{G}$. Suppose

$$\phi\colon X \to L(W), \quad \beta\colon \mathscr{G} \to L(W)$$

are a pair of mappings such that

$$\beta \text{ is a Lie algebra homomorphism,} \tag{13.5}$$

$$\text{for all } x \in X,\ V \in \mathscr{G},\ \phi_*(V(x)) \text{ is the tangent vector at } t = 0 \text{ of the curve } t \to \phi(x) + t\beta(V). \tag{13.6}$$

Then, V_h has a Lax representation.

Proof: This follows from Theorem 11.1.

We can now use "quantization" ideas to construct the map $\phi\colon X \to L(W)$. Start off with β given as a linear representation of the Lie algebra $\mathscr{G}$.

Let $\mathscr{A}$ be a linear subspace of $\mathscr{F}(X)$ containing $\mathscr{G}$, such that β can be extended to a linear mapping

$$\beta\colon \mathscr{A} \to L(W).$$

(For example, $\mathscr{A}$ might be the space of polynomial functions on the generators of $\mathscr{G}$, with β extended by "totally symmetric quantization," as explained in Sec. 12.)

Let $\phi\colon X \to L(W)$ be a map such that

$$f(x) = \operatorname{tr}(\beta(f)\phi(x)) \quad \text{for all } x \in X,\ f \in \mathscr{A}. \tag{13.7}$$

Suppose now that $h \in \mathscr{A}$, and V_h is the vector field on X defined by the Poisson structure. Consider h as the Hamiltonian of a physical system, so that V_h is the vector field which generates the *dynamics* of the system. Let

$$t \to x(t)$$

be an orbit curve of the vector field V_h,

$$\frac{d}{dt} f(x(t)) = V_h(f) = \{h, f\}(x(t)) \quad \text{for all } f \in \mathscr{F}(X). \tag{13.8}$$

Combine (13.5) and (13.6). For $f \in \mathscr{A}$,

$$\frac{d}{dt}(f(x(t))) = \frac{d}{dt}(\operatorname{tr}(\beta(f)\phi(x(t)))) = \operatorname{tr}\left(\beta(f)\frac{d}{dt}\phi(x(t))\right) = \operatorname{tr}(\beta(\{h, f\})\phi(x(t))).$$

Hence, we have the following relation:

$$\operatorname{tr}\left(\beta(\{h, f\})\phi(x(t)) - \beta(f)\frac{d}{dt}\phi(x(t))\right) = 0. \tag{13.9}$$

So far, we have made no assumption about how β behaves relative to the Poisson bracket structure on $\mathscr{F}(X)$. Let us now assume that β satisfies the following condition:

$$\beta(\{h, f\}) = [\beta(h), \beta(f)] \quad \text{for all } f \in \mathscr{A}. \tag{13.10}$$

Combining (13.7) and (13.8) and using the following identity for operators in finite-dimensional vector space,

$$\operatorname{tr}([A, B]C) = -\operatorname{tr}(B[A, C]), \tag{13.11}$$

we have

$$\operatorname{tr}\left(\beta(f)\left(\frac{d}{dt}\phi(x(t)) - [\beta(h)(x(t)), \phi(x(t))]\right)\right)$$

$$\text{for all } f \in \mathscr{A}. \tag{13.12}$$

We can now state what we have proved in the following form:

Theorem 13.2: Suppose $\mathscr{A}$ is a vector space of functions on the Poisson manifold X such that the following conditions are satisfied:

There is a linear mapping

$$\beta\colon \mathscr{A} \to L(W) \tag{13.13}$$

which is *onto*. h is an element of $\mathscr{A}$ such that

$$\beta(\{h, f\}) = [\beta(h), \beta(f)] \tag{13.14}$$

$$\text{for all } f \in \mathscr{A}.$$

There is a map $\phi\colon X \to L(W)$ such that:

$$f(x) = \operatorname{tr}(\beta(f)\phi(x)) \quad \text{for all } f \in \mathscr{A}. \tag{13.15}$$

Form the vector field V_h on X by the rule

$$V_h(f) = \{h, f\}. \tag{13.16}$$

Then, for every orbit curve $t \quad x(t)$ of V_h we have

$$\frac{d}{dt}\phi(x(t)) = [\beta(h)(x(t)), \phi(x(t))]. \tag{13.17}$$

In words, (ϕ, β) defines a *Lax representation* for the vector field V_h.

14. THE TODA LATTICE AND GENERALIZATIONS IN TERMS OF FUNCTION GROUP-POISSON STRUCTURES

In Ref. 17, I have briefly described how the Toda lattice[39,40] models are described in terms of one type of function group. I will discuss certain elaborations of this material here, and then use the hydrogen atom to illustrate how more complicated function groups may arise naturally in physical problems.

Let M be a symplectic manifold of dimension n. Choose indices as follows, and the summation convention

$$1 \leqslant i, j \leqslant n.$$

Let (q^i, p_i) be a canonical coordinate system for M, i.e., the symplectic form is

$$\omega = dp_i \wedge dq^i. \tag{14.1}$$

Then,

$$\{p_i, q^j\} = \delta_i^j, \qquad \{p_i, p_j\} = 0 = \{q^i, q^j\}, \tag{14.2}$$

where $\{\ ,\ \}$ is the Poisson bracket on $\mathscr{F}(M)$, the C^∞, real-valued functions on M. Let $a^i(q)$ be functions of the q variables such that

$$\{p_i, a^j\} = c^j_{ik}a^k + d^j_i, \tag{14.3}$$

where (c^j_{ik}), (d^j_i) are *constants*. Set

$$h = \sum_{i=1}^{n} p_i^2 + (a^i)^2. \tag{14.4}$$

Let us now use h as a Hamiltonian to construct a set of Hamilton's equations:

$$\dot{q} = H_p, \quad \dot{p} = -H_p . \tag{14.5}$$

As shown in Ref. 17, one choice of a's, c's, d's (involving variable parameters) leads to the Toda lattice. Let us look at this more systematically, without specializing the choice of a's, as in Ref. 17.

Let

$$\partial^i \equiv \frac{\partial}{\partial p_i}, \quad \partial_i = \frac{\partial}{\partial q^i} \tag{14.6}$$

be the coordinate vector fields on M. Let (A_i) be the vector fields on M such that

$$A_i(f) = \{a^i, f\} \quad \text{for } f \in \mathcal{F}(M) . \tag{14.7}$$

Since Poisson bracket is a Lie algebra homomorphism, (14.3) implies the following relation:

$$[\partial_i, A^j] = \lambda^j_{ik} A^k, \quad [A^i, A^j] = 0 . \tag{14.8}$$

This makes evident the following result:

Theorem 14.1: The vector fields (∂_i, A^j) form a Lie algebra $\mathcal{L}$ of vector fields on M which have the following properties:

$\mathcal{L}$ is the semidirect sum of the Lie algebra ideal generated by the A^i and an abelian subalgebra generated by the ∂_i; (14.9)

each vector field in $\mathcal{L}$ generates a one-parameter pseudogroup of automorphisms of the given symplectic structure on M. (14.10)

Also:

Theorem 14.2: Consider $\mathcal{F}(M)$ as a Lie algebra under the Poisson bracket operation. Let $\mathcal{L}'$ be the linear subspace of $\mathcal{F}(M)$ spanned by the p_i, the a^i, and the constant functions. Then, $\mathcal{L}'$ is a Lie subalgebra of $\mathcal{F}(M)$. It is a solvable Lie algebra. Let $\mathcal{A}$ be the linear subspace spanned by the a^i and the constant function; $\mathcal{P}$ the linear subspace spanned by the p_i. Then,

$$\mathcal{L} = \mathcal{A} + \mathcal{P} , \tag{14.11}$$

$$\{\mathcal{A}, \mathcal{A}\} = 0 = \{\mathcal{P}, \mathcal{P}\} , \tag{14.12}$$

$$\{\mathcal{P}, \mathcal{A}\} \subset \mathcal{A} ,$$

$$\{h, \mathcal{A}\} \subset \mathcal{P}\mathcal{A} , \tag{14.13}$$

$$\{h, \mathcal{P}\} \subset \mathcal{A}\mathcal{A} .$$

One key question for the general theory of "integrable systems" is how to choose the functions

$$q \to a^i(q)$$

to satisfy these conditions. We can now determine all such a's.

The functions $q \to a^j(q)$ satisfy the following differential equations:

$$\frac{\partial a^j}{\partial q^i} = \lambda^j_{ik} a^k + d^j_i , \tag{14.14}$$

whence

$$\frac{d}{dt} a^j(qt) = \lambda^j_{ik} a^k(qt) q^i + d^j_i q^i . \tag{14.15}$$

Of course, this constant coefficient ordinary differential equation (with q appearing as a parameter) can be solved explicitly. Let $\mathbf{a}(q,t)$ be the element

$$(a^1(qt), \ldots, a^n(qt))$$

of $\mathbb{C}$. Let

$$A(q) = (\lambda^j_{ik} q) \equiv (A^j_k(q)) \tag{14.16}$$

be the $n \times n$ matrix which depends linearly on q.

Let

$$\mathbf{d}(q) = (d^j_i q^i) \tag{14.17}$$

be the element of $\mathbb{C}^n$, which depends linearly on the parameter q. Then

$$\mathbf{a}(q,t) = A(q)^{-1}[1 - \exp[A(q)t]\mathbf{d}(q) + \mathbf{a}(0) . \tag{14.18}$$

The solutions $q \to a^j(q)$ of the system (14.17) are sums of exponential polynomial functions in the q's; they are sums of functions of the following form:

$$\exp(\alpha_1 q^1 + \cdots + \alpha_n q^n) q_1^{j_1} \cdots q_n^{j_n} , \tag{14.19}$$

$j_1, \ldots, j_n$ integers, $\alpha_1, \ldots, \alpha_n \in \mathbb{C}$.

This determines the physical system completely: The Toda lattice is one special class. Thus, in this case, we are able to use one type of "function group" as an ansatz to determine definitively a whole class of integrable systems.

15. INTEGRABLE SYSTEMS AND FUNCTION GROUPS OF THE HYDROGEN ATOM-KEPLER PROBLEM TYPE

I will now abstract out of the standard[41] Lie group-theoretic treatment of the hydrogen atom a type of function group that seems to be one step more elaborate than those we have encountered in the study of the Toda lattice.

Let M be a symplectic manifold. Let

$$f_1, f_2 \to \{f_1, f_2\}$$

be the Poisson-bracket structure it determines on $\mathcal{F}(M)$.

Choose indices and the summation convention as follows:

$$1 \leqslant i, j \leqslant n, \quad 1 \leqslant a, b \leqslant m .$$

Suppose

$$(f^i, g^a, h) \tag{15.1}$$

are elements of $\mathcal{F}(M)$ and

$$(\lambda^{ij}_k, \lambda^{ia}_b, \lambda^{ab}_i) \tag{15.2}$$

are numbers such that the following structure equations are satisfied:

$$\{f^i, f^j\} = \lambda^{ij}_k f^k , \tag{15.3}$$

$$\{f^i, g^a\} = \lambda^{ia}_b g^b , \tag{15.4}$$

$$\{g^a, g^b\} = h\lambda^{ab}_i f^i , \tag{15.5}$$

$$0 = \{f^i, h\} = \{g^a, h\} . \tag{15.6}$$

Clearly, the (f^i, g^a) form a function group, in Lie's sense. We are now in a situation where the full force of Ehresmann's theory of pseudogroups applies. In order to study the pseudogroup generated by the vector field generated by h, we define M as follows:

$$M_+ = \{ p \in M:\ h(p) > 0 \}\ , \tag{15.7}$$

$$M_- = \{ p \in M:\ h(p) < 0 \}\ , \tag{15.8}$$

$$M_0 = \{ p \in M:\ h(p) = 0 \}\ , \tag{15.9}$$

$$M_E = \{ p \in M:\ h(p) = E \}\ . \tag{15.10}$$

Then, the pseudogroup generated by the V_{f^α}, V_{g^α} leaves these subsets invariant. Of course, on the "energy surfaces," M_E, the pseudogroup is *locally equivalent to a Lie group action.*

I will use this type of example to pose a basic problem of the theory: *Classify locally and globally a function group of this type and the type of physical systems they model.* I plan further work on this problem as well as development of its physical ramifications in later papers in this series.

ACKNOWLEDGMENTS

I would like to thank Martin Kruskal and Alan Weinstein for brief but stimulating discussions about some of these topics. Marsden, Weinstein,[8] and A. Littlejohn are preparing related papers on the Poisson structures.

[1]E. T. Whittaker, *A Treatise on Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U. P., London and New York, 1959).
[2]E. N. Lorenz, J. Atmos. Sci. **20**, 130 (1963).
[3]M. Feigenbaum, J. Stat. Phys. **19**, 25–52 (1978).
[4]S. Lie, *Transformationsgruppen* (Chelsea, New York, 1970), Vol. II.
[5]S. Kowalewska, "Sur le probleme de la rotation d'un corps solide autour d'un point fixe"; Sur une propriete du systeme d'equations differentielles qui definit la rotation d'un corps solide autour d'un pointe fixe," Acta Mat. **12**, 177–232 (1889); **14**, 81–93 (1891).
[6]P. Painlevé, "Memoire sur les equations differentielles dont l'integrale generale est uniforme," Bull. Soc. Math. **28**, 201–261 (1900) [*Oeuvres*, Vol. 3, pp. 123–183].
[7]R. Hermann, *Interdisciplinary Mathematics* (Math. Sci. Press, Brookline, MA, 1975, 1977), Vol. 10, Chaps. 14 and 15, Vol. 14.
[8]A. Weinstein, "The local structure of Poisson manifolds" (to be published).
[9]A. Lichnerowicz, "Les varieties de Poisson et leurs algebras de Lie associees," J. Diff. Geom. **12**, 253–300 (1977).
[10]C. Ehresmann, "Sur les structures infinitesimals regulieres," Congr. Int. Math. Amsterdam **1**, 479–480 (1954); "Connexions infinitesimales," Coll. Top. Alg. Bruxelles, 29–55 (1950); "Structures infinitesimales et pseudogroupes de Lie," Colloq. Int. C.N.R.S. Geom. Diff. Strasbourg, 97–110 (1953); C. R. Acad. Sci. Paris **240**, 1762 (1954); **241**, 397, 1755 (1955); **246**, 360 (1958); "Connexions d'ordre superieur," *Atti 5 Congresso dell'Unione Matematica Italiana 1955* (Cremonese, Rome, 1956), pp. 326–8; "Categories topologiques et categories differentiables," Colloq. Geom. Diff. Globale Bruxelles, C.B.R.M., 137–150 (1958); "Groupoides differenciales," Rev. Un. Mat. Argentina, Buenos-Aires **19**, 48 (1960).
[11]J. Plante, "Foliations with Measure Preserving Holonomy," Ann. Math. **102**, 327–61 (1975).
[12]V. Arnold, Ann. Inst. Grenoble **16**, 319 (1966).
[13]R. Hermann, *Lie Algebras and Quantum Mechanics* (Benjamin, Reading, MA, 1970).
[14]R. Hermann, "Current algebras, the Sugawara model, and differential geometry," J. Math. Phys. **11**, 1825–9 (1970); "Geometric formula for current-algebra commutation relations," Phys. Rev. **177**, 2449 (1969); "Quantum field theories with degenerate Lagrangians," Phys. Rev. **177**, 2453 (1969); in *AIP Conference Proceedings No. 88*, edited by H. Wolf (Am. Inst. of Physics, New York, 1982), p. 67.
[15]M. Gell-Mann and M. Levy, Nuovo Cimento **16**, 705 (1960); T. D. Lee, S. Weinberg, and B. Zumino, Phys. Rev. Lett. **18**, 1029 (1967); K. Bardacki, Y. Frishman, and M. B. Halpern, Phys. Rev. **170**, 1353 (1968); H. Sugawara, Phys. Rev. **170**, 1659 (1968); K. Bardacki and M. B. Halpern, Phys. Rev. **172**, 1542 (1968); H. Sugawara and M. Yoshimura, Phys. Rev. **173**, 1419 (1968); C. Sommerfield, Phys. Rev. **176**, 2019 (1968); L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981).
[16]P. D. Lax, Comm. Pure Appl. Math. **21**, 487 (1968).
[17]R. Hermann, "The Lax representation as a 'quantization' of the function groups of Sophus Lie," Phys. Rev. D **26**, 1491–2 (1982).
[18]J. A. Schouten, "On the differential operators of first order in tensor calculus," *Convengo Internationale Geometria Differenziale Italia, 1953*, (Cremonese, Rome, 1954), pp. 1–7.
[19]A. Nijenhuis, "Jacobi-type identities for bilinear differential concomitants of certain tensor fields," I. Indag. Math. **17**, 390–403 (1955).
[20]R. Hermann, *Geometry, Physics and Systems* (Marcel Dekker, New York, 1973).
[21]R. Palais, *Seminar on the Atiyah–Singer Index Theorem* (Princeton U. P., Princeton, NJ, 1965).
[22]N. Steenrod, *The Topology of Fiber Bundles* (Princeton U. P., Princeton, NJ, 1950).
[23]C. Caratheodory, "Untersuchungen uber die Grundlagen der Thermodynamik," Math. Ann. **67**, 355–86 (1909).
[24]W. L. Chow, "Uber Systeme von Linearen Partiellen Differentialgleichungen Erster Ordnung," Math. Ann. **117**, 98–105 (1940).
[25]R. Hermann, "On the accessibility problem in control," *International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics* (Academic, New York, 1963), pp. 325–32; "Cartan connections and the equivalence problems for geometric structures," Contr. Diff. Eq. **3**, 199–248 (1964).
[26]P. Stefan, "Two proofs of Chow's theorem," in *Geometric Methods in System Theory*, edited by D. Mayne and R. Brockett (D. Reidel, Boston, 1973).
[27]H. J. Sussman, "Orbits of families of vector fields and integrability of systems with singularities," Bull. Am. Math. Soc. **79**, 197–9 (1973).
[28]C. Gardner, J. Greene, M. Kruskal, and R. Miura, "Methods for solving the Korteweg–de Vries equation," Phys. Rev. Lett. **15**, 240–2 (1965).
[29]J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton, U. P., Princeton, NJ, 1955).
[30]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
[31]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford U. P., Oxford, 1958), 4th ed.
[32]J. E. Moyal, Proc. Camb. Philos. Soc. **45**, 99–124 (1949).
[33]L. van Hove, Proc. R. Acad., Belgium **26**, 1–102 (1951).
[34]W. Boothby and H. C. Wang, "On contact manifolds," Ann. Math. **68**, 721–34 (1958).
[35]I. E. Segal, J. Math. Phys. **2**, 701–7 (1964).
[36]J.-M. Souriau, *Structures des Systems Dynamiques* (Dunod, Paris, 1970).
[37]B. Kostant, *Quantization and Unitary Representations*, Lecture Notes in Mathematics 170 (Springer, Berlin, 1970).
[38]R. Hermann, *Quantum and Fermion Geometry*, Interdisciplinary Mathematics, Vol. 16 (Math. Sci. Press, Brookline, MA, 1977).
[39]M. Toda, *Theory of Nonlinear Lattices* (Springer, Berlin, 1981).
[40]H. Flaschka, Phys. Rev. B **9**, 1924 (1974).
[41]E. Loebl, *Group Theory and Its Applications* (Academic, New York, 1971), Vol. II.

Higher-order special self-adjoint equations and particle dynamics

James L. Reid
Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602

John R. Ray
Department of Physics and Astronomy, Clemson University, Clemson, South Carolina 29631

(Received 29 September 1982; accepted for publication 22 April 1983)

We exhibit a remarkable connection between a hierarchy of higher-order special self-adjoint ordinary differential equations and the description of motion of a cluster of particles in classical mechanics. The cluster is assumed to consist of equal mass particles all moving in one dimension. In a perturbation schema based on the first-order equation of motion of the center of mass point, the time evolution of the moments of order $m-1$ is governed by the solution of a special self-adjoint equation of order m. A similar connection exists for the moments of a wave packet in quantum mechanics.

PACS numbers: 03.20. + i, 02.30.Hq

I. INTRODUCTION

Recently the present authors have discussed[1,2] a class of special self-adjoint ordinary differential equations. Using the notation of Ref. 1, hereinafter called I, the general member of this class may be written symbolically as

$$\tilde{L}_m \rho = 0, \quad m \geqslant 3, \tag{1.1}$$

where $\tilde{L}_m$ denotes a linear mth-order ordinary differential operator in which the next to highest-order derivative is absent. The special feature of this self-adjoint operator is that all its coefficients are expressed in terms of a single function, say $\phi(t)$, and its derivatives. Methods for constructing $\tilde{L}_m$ are given in I, and the first three members of the class are the following:

$$\dddot{\rho} + 4\phi\dot{\rho} + 2\dot{\phi}\rho = 0, \tag{1.2}$$

$$\rho^{(4)} + 10\phi\ddot{\rho} + 10\dot{\phi}\dot{\rho} + 3(\ddot{\phi} + 3\phi^2)\rho = 0, \tag{1.3}$$

$$\rho^{(5)} + 20\phi\dddot{\rho} + 30\dot{\phi}\ddot{\rho} + 2(9\ddot{\phi} + 32\phi^2)\dot{\rho} + 4(\dddot{\phi} + 16\phi\dot{\phi})\rho = 0, \tag{1.4}$$

where overdots mean differentiation with respect to the independent variable t. Šapkarev has determined the form of the sixth-order member[3] of this class, as well as the seventh- and eighth-order members,[4] by starting with assumed solutions and employing a process of differentiation and elimination. Indeed, an important property of the special self-adjoint hierarchy of equations (1.1) is that the general solution, as was shown in I, can be written in the form

$$\rho = b_1 u^{m-1} + b_2 u^{m-2} v + b_2 u^{m-3} v^2 + \cdots + b_{m-1} u v^{m-2} + b_m v^{m-1}, \tag{1.5}$$

where u and v are two linearly independent solutions of the second-order equation

$$\ddot{q} + \phi(t) q = 0 \tag{1.6}$$

and the $b_1,\ldots,b_m$ are arbitrary constants of integration. [In I it was convenient to write solution (1.5) in terms of solutions to the second-order equation

$$\ddot{q} + \frac{3}{m+1} a(t) q = 0,$$

in which a numerical factor dependent on the order m appears; since here we use (1.6), the numerical factors in (1.2)–(1.4), etc., will differ from those in analogous equations in I.] The special self-adjoint equation of order $m \geqslant 3$ can thus be solved in terms of two linearly independent solutions to the time-dependent harmonic oscillator equation (1.6), which may be viewed as the lowest-order member of this class of differential equations.

From our point of view of Sec. VI[2] of I, solution (1.5) is not a surprising result. It follows naturally from a class of nonlinear equations of order m, in ρ, for which we derived a superposition rule for composing the solutions. In the derivation of this superposition rule, which is of the form $\rho = x(t) r(\tau)$, $d\tau = dt/x^2$, we find that $r(\tau)$ must satisfy the nonlinear equation

$$\frac{d^m r}{d\tau^m} = \frac{K}{r^{(m+1)/(m-1)}}, \quad K = \text{const}, \tag{1.7}$$

while the variable $x(t)$ is found to satisfy simultaneously a hierarchy of m differential equations. The highest-order equation $L_m x = 0$ of this hierarchy is linear of order m, with arbitrary coefficients, and the lowest-order equation of the hierarchy is the second-order equation (1.6), the remaining $m-2$ equations being nonlinear. The requirement that $x(t)$ consistently satisfy all members of this hierarchy forces conditions on the initially arbitrary coefficients of $L_m x = 0$ in such a way as to yield the special self-adjoint differential equation $\tilde{L}_m x = 0$. The left-hand side of the nonlinear equation satisfied by ρ likewise takes the special self-adjoint form. Therefore when K in (1.7) is put equal to zero our problem reduces to the linear case $d^m r/d\tau^m = 0$. Since, as shown in I, a particular solution to $\tilde{L}_m x = 0$ is given by $x(t) = u^{m-1}$, where $u(t)$ is a solution to (1.6), we deduce solution (1.5) readily for any order $m \geqslant 3$. See I for details.

We thus arrive at the special self-adjoint differential equation and its general solution entirely within a mathematical context detached from any physical motivation. However, since the appearance of I, we have found that these special self-adjoint equations do occur in very basic physical problems. One such problem is the time evolution of the

 0022-2488/83/102433-05$02.50

average moments of a one-dimensional cluster of classical particles. This problem was first discussed by Andrews,[5] who showed the connection between the mean-square position relative to the center of mass and the third-order self-adjoint equation (1.2). The primary goal of this paper is to extend Andrews' treatment to higher orders in his perturbation schema and to show thereby the connection with the higher-order special self-adjoint equations of I. We shall prove that our perturbation equations for the time evolution of the moments of the particle cluster are solvable in terms of solution (1.5). As a by-product of the present treatment we obtain a set of m first-order differential equations from the analysis of the moments of order $m-1$ for each $m \geqslant 3$. When these m first-order equations are combined appropriately one obtains (1.1). Of course, $m-1$ additional mth-order differential equations can be similarly generated, whose exact solutions can be obtained from operations on solution (1.5).

Instead of studying the time evolution of a cluster of particles we could treat the evolution of a wavepacket in quantum mechanics. The wavepacket problem has been discussed by Andrews in another paper.[6] See likewise the papers[7,8] by Remaud and Hernandez, who also treat the quantum problem. For simplicity, in this paper we shall discuss only the one-dimensional classical cluster problem.

In Sec. II we review the work of Andrews[5] for the second-order moments. We note in Sec. III the results of carrying Andrews' perturbation series to third- and higher-order moments of the cluster. We relate the third- and higher-order moments to fourth- and higher-order self-adjoint equations. We conclude with suggestions for further work.

II. CLUSTER DYNAMICS (SECOND-ORDER MOMENTS)

We follow the work of Andrews[5] and consider N particles of identical mass M undergoing one-dimensional motion in the potential field $V(x)$, the force on each particle being expressed by

$$M\ddot{x}^i = -\left.\frac{\partial V(x)}{\partial x}\right|_{x=x^i}. \tag{2.1}$$

The index $i = 1,2,\ldots,N \geqslant 3$ labels the particles in the cluster. To effect an economy of notation in subsequent equations, we shall write (2.1) in the form

$$M\ddot{x}^i = -\frac{\partial V(x^i)}{\partial x^i}.$$

Here the notation $V(x^i)$ means that the potential energy of particle i of the cluster depends only on the position x^i of that particle in the potential field $V(x)$.

The center of mass of the cluster is here defined by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x^i,$$

and the center of mass point $\bar{x}$ satisfies the equation of motion

$$M\ddot{\bar{x}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial V(x^i)}{\partial x^i}. \tag{2.2}$$

We define the variable X^i as the position of particle i relative to the center of mass point

$$X^i = x^i - \bar{x}. \tag{2.3}$$

The average of any quantity A^i related to the cluster will be defined by

$$\overline{A} = \frac{1}{N}\sum_i A^i,$$

where the summation is understood to be over all particles of the cluster. For example, the square of the position of the ith particle from the center of mass is $(X^i)^2$, and the average of this quantity is thus given by

$$\overline{X^2} = \frac{1}{N}\sum_i (X^i)^2. \tag{2.4}$$

We refer to $\overline{X^2}$ as a second-order moment. Two other second-order moments of interest are

$$\overline{X\dot{X}} = \frac{1}{N}\sum_i X^i\dot{X}^i \tag{2.5}$$

and

$$\overline{\dot{X}^2} = \frac{1}{N}\sum_i (\dot{X}^i)^2. \tag{2.6}$$

With (2.3) and the definition of $\bar{x}$ it is a simple matter to obtain the identity

$$M\ddot{X}^i = -\frac{\partial V(x^i)}{\partial x^i} + \frac{1}{N}\sum_j \frac{\partial V(x^j)}{\partial x^j}. \tag{2.7}$$

We now expand $\partial V(x^i)/\partial x^i$ in a Taylor series about the center of mass point. Thus we obtain a series in powers of $x^i - \bar{x} = X^i$ of the form

$$\begin{aligned}\frac{\partial V(x^i)}{\partial x^i} &= \frac{\partial V(X^i+\bar{x})}{\partial \bar{x}} \\ &= \frac{\partial V(\bar{x})}{\partial \bar{x}} + \frac{\partial^2 V(\bar{x})}{\partial \bar{x}^2}X^i + \frac{1}{2!}\frac{\partial^3 V(\bar{x})}{\partial \bar{x}^3}(X^i)^2 \\ &\quad + \cdots.\end{aligned} \tag{2.8}$$

Now the terms of the right-hand side of (2.7) can be replaced by a Taylor polynomial of any desired order. As a first-order approximation keep the first two terms in expansion (2.8) and obtain from (2.7) the equation

$$\ddot{X}^i = \phi(t)X^i = 0, \tag{2.9}$$

where

$$\phi(t) = \frac{1}{M}\frac{\partial^2 V(\bar{x})}{\partial \bar{x}^2}. \tag{2.10}$$

Thus in this lowest order of approximation, the problem is N uncoupled time-dependent harmonic oscillators for the N particles about their center of mass. For this order, the equation satisfied by $\bar{x}$ follows from (2.2) and expansion (2.8) as

$$M\ddot{\bar{x}} = -\frac{\partial V(\bar{x})}{\partial \bar{x}}, \tag{2.11}$$

subject to the initial conditions $\bar{x}(t_0) = \bar{x}_0$, $\dot{\bar{x}}(t_0) = \dot{\bar{x}}_0$. Notice that this equation is the same differential equation as (2.1). Therefore, in order to determine $\phi(t)$, we must be able to solve the original equation of motion (2.1). Because of this fact the perturbation scheme that we develop in this paper may not be of great practical utility. However, we do not propose a perturbation scheme for solving the equation of motion, but rather a scheme for relating the higher-order moments of the cluster to the lowest-order motion of the

center of mass point. Equation (2.11) defines the center of mass trajectory of the cluster in first order. Consequently, by definition, we shall carry out our perturbation expansions about this first-order trajectory of the center of mass. We thus will determine $\bar{x}(t)$, and subsequently $\phi(t)$, from (2.11), and we use these same functions in all higher orders of the expansion.

We return to the moments $\overline{X^2}$, $\overline{X\dot{X}}$, and $\overline{\dot{X}^2}$ and introduce, following Andrews, the notation

$$\chi = \overline{X^2}, \quad \zeta = \overline{X\dot{X}}, \quad \omega = \overline{\dot{X}^2}. \tag{2.12}$$

We can now use the approximate equation of motion (2.9) and the definitions (2.12) to derive equations satisfied by the second-order moments. These equations were obtained by Andrews,[5] and they have the forms

$$\dot{\chi} = 2\zeta, \tag{2.13a}$$

$$\dot{\zeta} = \omega - \phi\chi, \tag{2.13b}$$

$$\dot{\omega} = -2\phi\zeta. \tag{2.13c}$$

By differentiation one can eliminate ζ and ω from the set (2.13) to obtain a third-order equation for χ, i.e.,

$$\dddot{\chi} + 4\phi\dot{\chi} + 2\dot{\phi}\chi = 0, \tag{2.14}$$

which is identical to (1.2). Hence, from (1.5) we can write the general solution to (2.12) in the form

$$\chi = b_1u^2 + b_2uv + b_3v^2, \tag{2.15a}$$

where u and v are linearly independent solutions of (1.6). One may now use solution (2.15a) and the first-order equations (2.13) to find the solutions for ζ and ω; these take the forms

$$\zeta = b_1u\dot{u} + \tfrac{1}{2}b_2(u\dot{v} + \dot{u}v) + b_3v\dot{v}, \tag{2.15b}$$

$$\omega = b_1\dot{u}^2 + b_2\dot{u}\dot{v} + b_3\dot{v}^2. \tag{2.15c}$$

The constants b_1, b_2, and b_3 can be determined from the initial values of χ, ζ, and ω. The moments ζ and ω given by (2.15b) and (2.15c) satisfy the linear third-order equations

$$\dddot{\zeta} - (\ddot{\phi}/\dot{\phi})\ddot{\zeta} + 4\phi\dot{\zeta} + 2(3\dot{\phi} - 2\ddot{\phi}\phi/\dot{\phi})\zeta = 0, \tag{2.16}$$

$$\dddot{\omega} - 3(\dot{\phi}/\phi)\ddot{\omega} + (4\phi + 3\dot{\phi}^2/\phi^2 - \ddot{\phi}/\phi)\dot{\omega} - 2\dot{\phi}\omega = 0, \tag{2.17}$$

respectively. These last two equations can each be found by eliminating variables in the first-order set (2.13). Unlike (2.14), satisfied by the spatial moment χ, (2.16) and (2.17) are not self-adjoint. A similar remark holds for all higher orders.

III. CLUSTER DYNAMICS (THIRD- AND HIGHER-ORDER MOMENTS)

In the next order of approximation we retain the first three terms of expansion (2.8), and we have from (2.4) the equation of motion

$$\ddot{X}^i = -\phi(t)X^i - \theta(t)(X^i)^2 + \theta(t)\overline{X^2}, \tag{3.1}$$

where $\phi(t)$ is given by (2.10) as before and the new function $\theta(t)$ is defined by

$$\theta(t) = -\frac{1}{2!M}\frac{\partial^3 V(\bar{x})}{\partial \bar{x}^3}, \tag{3.2}$$

where $\bar{x}(t)$ is the first-order center of mass trajectory defined by Eq. (2.11). For this perturbation treatment to be useful, the terms in θ appearing in (3.1) should be small compared to the leading term in ϕ in the sense that

$$|\phi X| \gg |\theta X^2|. \tag{3.3}$$

Assuming the validity of (3.3), which is a condition on how rapidly the potential may vary, the terms in θ in (3.1) give a small correction to the motion.

Consider next the second-order moments defined in (2.11). At our higher level of approximation, the first-order equations satisfied by χ, ζ, and ω are now given by

$$\dot{\chi} = 2\zeta, \tag{3.4a}$$

$$\dot{\zeta} = \omega - \phi\chi - \theta\eta, \tag{3.4b}$$

$$\dot{\omega} = -2\phi\zeta - 2\theta\sigma, \tag{3.4c}$$

where η and σ are third-order moments defined by

$$\eta = \overline{X^3}, \tag{3.5a}$$

$$\sigma = \overline{X^2\dot{X}}. \tag{3.5b}$$

Thus, the corrections to second-order moment equations (2.13) involve the terms in θ in (3.4b) and (3.4c) which in turn introduce the third-order moments η and σ. The other third-order moments are

$$\kappa = \overline{X\dot{X}^2}, \tag{3.5c}$$

$$\tau = \overline{\dot{X}^3}. \tag{3.5d}$$

Next, we must find the equations satisfied by the third-order moments η, σ, κ, and τ. Using the definitions (3.5) along with the equation of motion (3.1) we obtain the system of first-order equations

$$\dot{\eta} = 3\sigma, \tag{3.6a}$$

$$\dot{\sigma} = 2\kappa - \phi\eta - \theta\overline{X^4} + \theta\chi^2, \tag{3.6b}$$

$$\dot{\kappa} = \tau - 2\phi\sigma - 2\theta\overline{X^3\dot{X}} + 2\theta\chi\zeta, \tag{3.6c}$$

$$\dot{\tau} = -3\phi\kappa - 3\theta\overline{\dot{X}^2X^2} + 3\theta\chi\omega. \tag{3.6d}$$

Note that the third-order moment equations contain fourth-order moments $\overline{X^4}$, etc. Clearly, in order to obtain a closed set of equations, we must neglect the terms involving fourth-order moments in (3.6). To the same order of accuracy, we can also neglect the other terms in θ. That those terms in θ are small compared to the terms in ϕ also follows from the perturbation inequality (3.3). It thus suffices to solve the third-order moment equations (3.6) in the linear form

$$\dot{\eta} = 3\sigma, \tag{3.7a}$$

$$\dot{\sigma} = 2\kappa - \phi\eta, \tag{3.7b}$$

$$\dot{\kappa} = \tau - 2\phi\sigma, \tag{3.7c}$$

$$\dot{\tau} = -3\phi\kappa. \tag{3.7d}$$

These equations are to be solved for the third-order moments, and then the solutions are to be used in (3.4) to obtain the perturbed equations for the second-order moments.

To effect the required solution we eliminate σ, κ, and τ from (3.7) and thus arrive at the differential equation satisfied by η, i.e.,

$$\eta^{(4)} + 10\phi\ddot{\eta} + 10\ddot{\phi}\dot{\eta} + 3(\dddot{\phi} + 3\phi^2)\eta = 0, \tag{3.8}$$

which is the fourth-order special self-adjoint equation (1.3). The general solution for η from (1.5) thus takes the form

$$\eta = b_1u^3 + b_2u^2v + b_3uv^2 + b_4v^3. \tag{3.9a}$$

Using solution (3.9a) and the first-order system (3.7) we obtain the solutions for σ, κ, and τ as the forms

$$\sigma = b_1u^2\dot{u} + \tfrac{1}{3}b_2(2u\dot{u}v + u^2\dot{v}) + \tfrac{1}{3}b_3(2uv\dot{v} + v^2\dot{u}) + b_4v^2\dot{v}, \tag{3.9b}$$

$$\kappa = b_1u\dot{u}^2 + \tfrac{1}{3}b_2(2u\dot{u}\dot{v} + \dot{u}^2v) + \tfrac{1}{3}b_3(2\dot{u}v\dot{v} + u\dot{v}^2) + b_4v\dot{v}^2, \tag{3.9c}$$

$$\tau = b_1\dot{u}^3 + \tfrac{1}{3}b_2\dot{u}^2\dot{v} + \tfrac{1}{3}b_3\dot{u}\dot{v}^2 + b_4\dot{v}^3. \tag{3.9d}$$

If the initial values $u(0) = 1$, $\dot{u}(0) = 0$, $v(0) = 0$, and $\dot{v}(0) = 1$ are chosen then the constants b have the values

$$b_1 = \eta_0, \quad b_2 = 3\sigma_0, \quad b_3 = 3\kappa_0, \quad b_4 = \tau_0.$$

Equations (3.9) can then be written in the matrix form

$$\begin{pmatrix}\eta\\ \sigma\\ \kappa\\ \tau\end{pmatrix} = \mathscr{M}(t)\begin{pmatrix}\eta_0\\ \sigma_0\\ \kappa_0\\ \tau_0\end{pmatrix} \quad \text{or} \quad \mathscr{Y}(t) = \mathscr{M}(t)\mathscr{Y}_0, \tag{3.10}$$

where the 4×4 evolution matrix $\mathscr{M}(t)$ can be found easily from (3.9).

The final step at this level of correction to the second order moments is to return with solutions (3.9) and solve (3.4) for the adjusted second-order moments. Note that $\phi(t)$ and $\theta(t)$ in these equations are known functions of time given by (2.10) and (3.2), respectively.

Before leaving this set of equations, we remark that the inverse of (3.10),

$$\mathscr{Y}_0 = \mathscr{M}(t)^{-1}\mathscr{Y}(t), \tag{3.11}$$

leads to four first integrals η_0, σ_0, κ_0, and τ_0 in analogy to those found by Andrews[5] for χ_0, ζ_0, ω_0. Using the latter three quantities Andrews was able to give a new derivation of the well-known Lewis invariant. On the other hand, Hernandez and Remaud[9] have discussed the quantum-mechanical aspect of analogous invariants involving second-order moments. We shall not present the explicit forms of the invariants η_0, σ_0, κ_0, and τ_0 since they are not needed in this paper.

The procedure to extend the analysis to higher approximations is now clear. The inclusion of the next term from (2.8) into an approximation to the equation of motion (2.7) will introduce a new function

$$\gamma(t) = \frac{1}{3!M}\frac{\partial^4 V(\bar{x})}{\partial\bar{x}^4} \tag{3.12}$$

and associated cubic nonlinearities. The equations for the second- and third-order moments will now involve fourth-order moments. There are five of these moments, i.e.,

$$\overline{X^4}, \quad \overline{X^3\dot{X}}, \quad \overline{X^2\dot{X}^2}, \quad \overline{X\dot{X}^3}, \quad \overline{\dot{X}^4}. \tag{3.13}$$

If one denotes these quantities by a new set of symbols then, as before, by differentiation and appropriate substitution for $\ddot{X}^i$ one will obtain a set of five first-order equations in the new variables. One may use the condition that terms involving $\theta(t)$ and $\gamma(t)$ are small compared to those in $\phi(t)$. Keeping only the largest terms in these five moment equations, one thus obtains a set of five linear first-order equations that will yield the fifth-order self-adjoint equation (1.4) for the spatial moment $\rho = \overline{X^4}$. The solution to (1.4) follows of course from (1.5); the general solutions for the four remaining moments now readily follow. These solutions are then used to obtain corrections to the third-order moment equations and then to the second-order moment equations.

The foregoing procedure can be continued to any higher order in the perturbation series. At the $(m-1)$st order of moments we arrive at m first-order equations which become linear upon dropping all terms in $\theta(t)$, $\gamma(t)$,... and keeping those in $\phi(t)$. In a systematic notation, the first-order system under discussion takes the form

$$\begin{aligned}
\dot{a}_1 &= (m-1)\ a_2, \\
\dot{a}_2 &= (m-2)\ a_3 - \phi\ a_1, \\
\dot{a}_3 &= (m-3)\ a_4 - 2\phi\ a_2, \\
&\ \ \vdots \\
\dot{a}_{m-2} &= 2\ a_{m-1} - (m-3)\phi\ a_{m-3}, \\
\dot{a}_{m-1} &= a_m - (m-2)\phi\ a_{m-2}, \\
\dot{a}_m &= -(m-1)\phi\ a_{m-1}.
\end{aligned} \tag{3.14}$$

Eliminating all the other a_i in favor of the spatial moment $a_1 = \overline{X^{m-1}}$ will lead to the mth-order self-adjoint equation (1.1). Solution (1.5) and system (3.14) can now be combined to yield the solutions for the remaining $m-1$ moments. These m known quantities can now be used to correct the lower-order moment equations from the $(m-2)$nd order down to second order $(m \geqslant 4)$.

There is thus a direct relationship between the foregoing perturbative treatment for the moments of a cluster of particles and the special self-adjoint equation introduced in I.

IV. CONCLUSIONS

We have demonstrated in this paper a close connection between the description of motion of a cluster of particles and the hierarchy of special self-adjoint differential equations of I. These equations were arrived at in I from mathematical premises without any hint of physical origin. It is remarkable that the very same hierarchy of differential equations arises in a basic way in Newtonian mechanics. The special self-adjoint differential equation is in fact the key element to effect a perturbative solution for the time evolution of the moments of the particles within a cluster.

We fully expect the analysis of this paper to bear a close analogy to that required for the description of the higher-order moments of a spreading quantum-mechanical wave-packet. Extending the work of Andrews[6] and of Remaud and Hernandez[7–9] to third-order moments and beyond is an interesting topic for further work.

We concur with Andrews[5] that the present study can be extended to include classical particles with unequal mass, motion in three dimensions, mutual interactions between particles, and external forces. Each of these extensions is an important topic for future work.

[1] J. L. Reid and J. R. Ray, J. Math. Phys. **23**, 503 (1982).
[2] Note in Sec. VI of Ref. 1 there are some misprints:

(a) Eq. (6.13) should read $\rho = [b_1 x_2^2 + 2b_2 x_1 x_2 + b_1^{-1}(b_2^2 + K)x_1^2]^{1/2}$, where x_1 and x_2 have unit Wronskian;
(b) Eq. (6.14) should read $\ddot{\rho} + 3a_2\dot{\rho} + \frac{3}{2}\dot{a}_2\rho = K\rho^{-2}$;
(c) Eq. (6.15) should read $\rho = xr = (b_1 X_1 + b_2 X_2)^2 r(\tau)$;
(d) Eq. (6.16) should read $r''' = Kr^{-2}, d\tau = dt/X^2, X = b_1 X_1 + b_2 X_2$; finally, the next to last sentence in this section should read "For the general solution of (6.7) we need at least $m - 2$ constants of integration from (6.9)."

[3] I. A. Šapkarev, Mat. Vesnik **1(1b)**, 95 (1964).
[4] I. A. Šapkarev, Bull. Soc. Math. Phys. Macédoine 14 (1963); 41 (1964).
[5] M. Andrews, Phys. Lett. A **83**, 95 (1981).
[6] M. Andrews, J. Phys. A: Math. Gen. **14**, 1123 (1981).
[7] B. Remaud and E. S. Hernandez, J. Phys. A: Math. Gen. **13**, 2013 (1980).
[8] B. Remaud and E. S. Hernandez, Physica A **103**, 35 (1980).
[9] E. S. Hernandez and B. Remaud, Phys. Lett. A **75**, 269 (1980).

Lagrange equations for a spinning gas cloud

S. Y. Shieh
Physics Department, University of Tennessee, Knoxville, Tennessee 37996-1200

(Received 30 March 1983; accepted for publication 26 May 1983)

Lagrange equations are derived for a spinning gas cloud. The rotational and vortex velocities are treated as independent variables and their defining equations as equations of constraint. Application of the formalism of Lagrange multipliers to this case of 27 variables and 18 equations of constraint yields nine final equations which are simple in form and contain only variables explicitly included in the kinetic and potential energy.

PACS numbers: 03.20. + i, 47.30. + s

I. INTRODUCTION

The internal and rotational motions of a cloud of compressible gas expanding freely in an otherwise empty space was treated by Dyson.[1] He studied the problem under the assumption that the position vector $\gamma(t)$ at any time of each element of the gas is a linear function of its initial position $\mathbf{a} = \gamma(0)$, namely,

$$\gamma(t) = F\mathbf{a}, \tag{1}$$

where nine components F_{ij} are functions of time alone. This ansatz was first introduced by Dirichlet[2] into the problem of equilibrium figures of gravitating ellipsoids. Those who are not familiar with this field often get a misleading impression from Eq. (1) that only trivial results can follow from such a simplifying assumption. However, prior to Dirichlet, only the effect of rotation was considered on the equilibrium figures of gravitating ellipsoids. Now, the assumption of Eq. (1) adds internal vortex motion as well as the pulsation of the semiaxes into the study of the problem. Furthermore, Eq. (1) contains the interactions of these three types of motion. Even with the aid of modern computers, these interactions are not yet fully explored, and Chandrasekhar's book[3] is a convenient source for a modern account of this problem. For those who enjoy the study of classics, Riemann's paper[4] is a *Chef-d'oeuvre* of extracting a great wealth of beautiful results from Dirichlet's assumption. Equation (1) is one of those examples which illustrates in a rather dramatic way that nature is profound in its simplicity. Dyson found that these solutions have a symmetry group $O(4)$ whose six generators are constants of the motion. Utilizing the isomorphism between $[O(3)\times O(3)]$ and $O(4)$, he introduced a symmetric traceless (4×4) matrix in place of the asymmetric (3×3) matrix F_{ij} and gave the nine equations of motion an elegant form [Eqs. (47) and (48) in his paper]. But his equations, even in the symmetric and elegant form, are essentially Newtonian. Here, the Lagrange equations are derived for this problem. The difficulty in applying the Lagrangian formalism to this problem lies in the defining equations of the components of the rotational angular velocity vector in terms of the elements and the derivatives of the elements of an orthogonal matrix. Here, we also have an additional vortex vector related in a similar manner to a second orthogonal matrix. The method of overcoming this hurdle was first employed by Kirchhoff[5] in his derivation of the equations for the translation and rotation of a rigid body in a fluid. Following Kirchhoff, we will initially treat the rotational and vortex velocities as independent variables and regard their defining equations as the equations of constraints. We also have the twelve orthogonality conditions of the two orthogonal matrices. After carrying through the formalism of Lagrange multipliers for 27 variables with 18 equations of constraints, we will obtain, in a simple final form, the equations for the three major axes, the three components of the rotational velocity, and the three components of the vortex velocity. Our derivation and results are applicable to problems more general than Dyson's case.

II. PRELIMINARIES

In this section, we briefly describe those pertinent results from Dyson's work which are needed for our derivation of the equations in the Lagrangian form. For details, we refer the reader to Dyson's paper.

First, we represent the nine variables F_{jk} in Eq. (1) in the form

$$F = O_1 D O_2, \tag{2}$$

where $O_1(\alpha_{jk})$ and $O_2(A_{jk})$ are orthogonal and D is diagonal with (D_1, D_2, D_3) as its diagonal elements. Dyson assumes that the potential energy U is a function of the temperature of the gas and shows that it depends only on $b = \det F$. Now, due to the decomposition Eq. (2) and the orthogonality of O_1 and O_2, we have

$$b = D_1 D_2 D_3, \tag{3}$$

so that the potential energy U is a function of D_1, D_2, D_3 alone.

Next, before we define the rotational velocity ω, we introduce an antisymmetric matrix K, whose components are related to the components of ω by $K_{lm} = \epsilon_{lmn}\omega_n$, where $\epsilon_{lmn} = 0$ if l, m, n are not distinct and $\epsilon_{lmn} = +1\ (-1)$ if lmn is obtained from 123 by an even (odd) number of exchanges. Now, the defining equation for ω from the orthogonal matrix, O_1 reads

$$K = O_1^T \dot{O}_1, \tag{4}$$

where the dot denotes differentiation with respect to time and the superscript T indicates the transposition of the rows and columns. Similarly, we define the vortex velocity ϕ from the second orthogonal matrix O_2

$$H = O_2^T \dot{O}_2, \tag{5}$$

 0022-2488/83/102438-03$02.50

where the antisymmetric matrix H is related to the vector ϕ by $H_{lm} = \epsilon_{lmn}\phi_n$. The kinetic energy T will be a function explicitly of D_i, $\dot{D}_i$, ω, and ϕ alone. But because the velocities ω and ϕ are related to α_{ij}, $\dot{\alpha}_{ij}$, A_{ij}, and $\dot{A}_{ij}$ through Eqs. (4) and (5), we cannot take advantage of this simple feature of the kinetic energy unless we treat ω and ϕ as independent variables and regard the defining Eqs. (4) and (5) as equations of constraint. By raising the status of ω and ϕ to the independent variables, in addition to D_i, α_{ij}, A_{ij}, we now have 27 variables and 18 equations of constraint, of which the remaining 12 are the following orthogonality relations.

$$O_1^T O_1 - I = 0, \tag{6}$$

and

$$O_2^T O_2 - I = 0, \tag{7}$$

where I is the unit matrix.

III. DERIVATION OF LAGRANGE EQUATIONS

When there exist equations of constraint, the Lagrangian $L = T - V$ must be extended to include the sum $\Sigma_{i=1}^{18}\mu_i f_i$, where f_i represents the left-hand side of the equations of constraint when written in the form $f_i = 0$, and μ_i are the Lagrange multipliers.

$$\mathscr{L} = L + \sum_{i=1}^{18}\mu_i f_i, \tag{8}$$

$$\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{q}_i} - \frac{\partial \mathscr{L}}{\partial q_i} = 0. \tag{9}$$

Both the number of variables and the number of equations of constraint may seem unusually large, but due to the simplifying fact that different sets of variables enter separate parts of the extended Lagrangian, the computation is relatively easy in spite of the number of variables involved.

First, we use the three defining equations (4) for ω_i as equations of constraint and multiply each equation by multiplier Ω_i, respectively, for the inclusion in Eq. (8). Then, because there are no derivatives $\dot{\omega}_i$ of ω_i in Eq. (8), we immediately obtain

$$\frac{\partial T}{\partial \omega_i} = \Omega_i \tag{10}$$

from Eq. (9). Similarly, denoting the multipliers for Eqs. (5) as Φ_i, we get

$$\frac{\partial T}{\partial \phi_i} = \Phi_i. \tag{11}$$

Next, the set of nine variables α_{jk} enter only through Eqs. (4) and (6), and for these variables, Eqs. (9) becomes

$$G_{lm} = \frac{d}{dt}(\alpha_{ij}\Omega_{jm}) - (\alpha\Lambda)_{lm} = 0. \tag{12}$$

There are nine equations $G_{lm} = 0$ for all combinations of $l, m = 1, 2, 3$, and the repeated index of j implies a summation. Λ is a symmetric matrix whose components λ_{lm} are the multipliers $\lambda_{11}, \lambda_{22}, \lambda_{12} = \lambda_{21}$ etc. for Eqs. (6). For convenience, the diagonal elements are divided by two before they are used as multipliers for Eqs. (6). $\Omega_{lm} = \epsilon_{lmn}\Omega_n$.

In order to eliminate λ_{lm} from Eq. (12), we multiply each of the nine equations $G_{lm} = 0$ by $\epsilon_{ljn}\alpha_{jm}$ (for fixed $n = 1, 2, 3$) and add the nine products to obtain

$$\frac{d}{dt}\left(\alpha_{ni}\frac{\partial T}{\partial \omega_i}\right) = 0, \quad n = 1,2,3. \tag{13}$$

We need the relations $\alpha_{11} = \alpha_{22}\alpha_{33} - \alpha_{23}\alpha_{32}$ etc. between the elements of an orthogonal matrix and its cofactors in the above derivation. We also replaced the multipliers Ω_i by $\partial T/\partial\omega_i$ obtained from Eq. (10). $\partial T/\partial\omega_i$ represents the components of the angular momentum in the moving coordinate system of O_1, and from Eq. (13), its components in the fixed system are constant.

$$\alpha_{ni}\frac{\partial T}{\partial \omega_i} = l_n. \tag{14}$$

Again, if we multiply each of Eqs. (12) by $\epsilon_{nmj}\alpha_{lj}$ (for fixed $n = 1, 2, 3,$) and add the nine products, we get

$$\frac{d}{dt}\frac{\partial T}{\partial \omega_n} = \epsilon_{nlm}\omega_l\frac{\partial T}{\partial \omega_m}. \tag{15}$$

Equations (14) represent the conservation of the angular momentum of the system. Equations (15) describe the variation with time of the angular velocity in the O_1 system.

In a similar manner, we obtain the corresponding equations for the vortex part

$$\alpha_{ni}\frac{\partial T}{\partial \phi_i} = \eta_n \tag{16}$$

and

$$\frac{d}{dt}\frac{\partial T}{\partial \phi_n} = \epsilon_{nlm}\phi_l\frac{\partial T}{\partial \phi_m}. \tag{17}$$

We note that the coefficients of $\partial T/\partial\phi_i$ in Eq. (17) are the vortex components ϕ_i, not the rotational components ω_i. This formal separation ω and ϕ in the final Eqs. (15) and (17) is not real, because of the simultaneous entry of both ϕ and ω in the kinetic energy T, and these two motions do interact despite the fact that they appear to be separated in the outlook of Lagrange equations (15) and (17).

Finally, for the three variables D_i which do not enter into any of the equations of constraint, we need only the original Lagrangian and obtain

$$\frac{d}{dt}\frac{dL}{d\dot{D}_i} - \frac{\partial L}{\partial D_i} = 0, \quad i = 1,2,3. \tag{18}$$

Here, we note that D_i enters in the kinetic energy term T as well as the potential term U so that we cannot separate the Lagrangian into $L = T - U$ and use T and U separately in the first and second terms of Eqs. (18). Physically speaking, $D_i^2 + D_j^2$ etc. represent the moments of inertia and, thus, enter in the kinetic energy term. In Dyson's case, the potential U depends only on the product $D_1D_2D_3$.

IV. DISCUSSION

The final apparent form of the equations we obtained in Eqs. (15), (17), and (18) is indeed that of the familiar Lagrange equations. However, ω and ϕ are not generalized velocities of some hypothetical generalized coordinates. In fact, in the rigorous derivation given above, both ω and ϕ initially were introduced as independent variables, together with their de-

rivatives $\dot{\omega}$ and $\dot{\phi}$, which did not occur, but in the final Lagrangian form, Eqs. (15) and (17), both ω and ϕ enter with the appearance of generalized velocities. Kirchhoff 's powerful method here brings out, in the form of the basic equations of motion, the deeply lying consistency of Dyson's formulation of the problem of a spinning gas cloud.

Finally, the derivation given is applicable to problems more general than Dyson's case. If a dynamical system has a kinetic energy which is a function of D_i, $\dot{D}_i$, ω, and ϕ, and a potential energy depending only on D_i, the equations derived here can be applied *mutatis mutandis*. The separation of variables in Eqs. (15), (17), and (18) and the relatively wide choice of T and U which satisfy the conditions in this derivation of the equations should be useful in a search for analytically soluble cases.

[1]F. J. Dyson, J. Math. Mech. **18**, 91 (1968).
[2]G. Dirichlet, J. Reine Angew. Math. **58**, 181 (1860), also in *Werke* (Reimer, Berlin, 1897), Vol. 2, p. 265.
[3]S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale U. P., New Haven, CT, 1969).
[4]B. Riemann, Abh. d. Konigl. Gesell. der Wis zu Gottingen **9**, 3 (1860), also in *Gesammelte Math. Werke* (B. G. Teubner, Leipzig, 1892), p. 182.
[5]G. Kirchhoff, J. Reine Angew. Math. **71**, 237 (1869), also in *Gesammelte Abhandlungen* (J. A. Barth, Leipzig, 1882), p. 376.

Classical theories and nonclassical theories as special cases of a more general theory

Dirk Aerts
Theoretische Natuurkunde, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

(Received 3 August 1982; accepted for publication 5 November 1982)

We analyze the difference between classical mechanics and quantum mechanics. We come to the conclusion that this difference can be found in the nature of the observables that are considered for the physical system under consideration. Classical mechanics can only describe a certain kind of what we called "classical observable." Quantum mechanics can only describe another kind of observable; it cannot describe, however, classical observables. To perform this analysis, we use a theory where every kind of observable can be treated and which is in a natural way a generalization of both classical and quantum mechanics. If in a study of a physical system in this theory we restrict ourselves to the classical observables, we rediscover classical mechanics as a kind of first study of the physical system, where all the nonclassical properties are hidden. If we find that this first study is too rough we can also study the nonclassical part of the physical system by a theory which is eventually quantum mechanics.

PACS numbers: 03.65.Bz

INTRODUCTION

What is the relation between classical mechanics and quantum mechanics and in which aspects are they different physical theories? This is the question that we should like to investigate. Many different interpretations of quantum mechanics have been put forward during the years. It is indeed not straightforward to interpret the complicated mathematical formalism on which quantum mechanics is based.

For classical mechanics there have never been many discussions about the interpretation of the theory. Probably this is so because the interpretation of classical mechanics seems to be straightforward. We shall show, however, that there is not so much difference between the two theories and that a lot of the mystery of quantum mechanics is already present in classical mechanics.

Often one tries to see classical mechanics as a kind of limit of quantum mechanics (e.g., for $\hbar \to 0$). We think that this is not a correct way to see the relation between the two theories, because it starts from the idea that quantum mechanics is a more general theory than classical mechanics. It becomes more and more clear that this is not the case. First of all, it seems to be rather impossible to give a satisfactory description of a macroscopical system that is well described by classical mechanics by using quantum mechanics. This should, however, in principle be possible if quantum mechanics were a more general theory than classical mechanics. But even for microscopical systems already a long time ago a shortcoming of quantum mechanics was noted. It became clear that some superpositions of states of microscopical systems never do occur, although they are contained in the description of the system by quantum mechanics. To take this fact into account, one introduced the concept of "superselection rule." A superselection rule is a rule that forbids certain superpositions. It is not very satisfactory that one has to introduce this concept *a posteriori* in the theory. There exist, however, more general theories than quantum mechanics where the possibility of describing superselection rules is present from the start. This is, for example, the case in the algebraic approach to quantum mechanics and also in the quantum logic approach.

That it is possible to have also continuous superselection rules was shown by Piron[1] and used by Piron to give a description of what he calls a Galilean particle[2]. In this description time is considered to be a continuous superselection variable. We shall analyze how these superselection rules are described in this quantum logic approach. Although superselection rules must not be introduced *a posteriori* in this theory anymore, we shall not be satisfied with this description. Indeed, we should try to "understand" why and when these superselection rules are present. This shall follow immediately out of the analysis that we will make in the following. There is another reason why we are not satisfied with the state of affairs as it is now. In quantum logic a physical system is described by the lattice of its properties (yes–no experiments); often the properties of a physical system are also called propositions; this, however, makes it possible to confuse with the term proposition of logic). This lattice plays the role of the complex Hilbert space of ordinary quantum mechanics. To be able to show that the lattice can be decomposed in the direct union of irreducible lattices [see Ref. 2, Theorem (2.45)] and in this way introduce superselection rules, several axioms have to be satisfied in this lattice of propositions. We showed that some of these axioms, namely the weak modularity and the covering law, are axioms, that cannot be satisfied in nature if one wants to be able to describe separated physical systems.[3–5] As we shall show, it is possible to find a decomposition of the lattice of propositions as a direct union of irreducible lattices without the weak modularity to be satisfied (neither the covering law and neither the atomicity of the lattice has to be satisfied). To be able to do this we have to use a different approach to the problem. In this way we really understand the nature of these superselection rules and also the relation between classical mechanics and quantum mechanics.

What is now the most important difference between a classical theory and a quantum theory? In both theories the

 0022-2488/83/102441-13$02.50

concept of "state" of the physical system and the concept of "observable" is defined. In classical mechanics the state of the physical system is represented by a point in the state space of the system and an observable is represented by a function of the state space to an outcome set. If the system is in a state p and f is the function corresponding to a certain observable, then $f(p)$ is the value that this observable "has." In classical mechanics we do not specify what we mean by this word "has." In quantum mechanics the state of the physical system is represented by a ray in the Hilbert space of the system and an observable is represented by a self-adjoint operator on the Hilbert space. There is a collection of states, namely the eigenstates of the self-adjoint operator, such that when the system is in one of these eigenstates, we can predict that if we should perform the experiment corresponding to the observable in question, we would find the eigenvalue corresponding to the eigenstate is question for the observable. If the system is not in an eigenstate, we cannot make a prediction about the value of the observable for an experiment. What is now the difference between these two theories? First of all, we can remark that quantum mechanics is much more specific and detailed while classical mechanics is rather vague. To analyze the real difference, we shall be obliged to specify more in detail what classical mechanics means with the word "has." Often it is claimed that classical mechanics is a theory that neglects the effect of the measurement, or less strong classical mechanics is a theory where the effect of the measurement can be taken into account while in quantum mechanics this is not the case anymore. This is then asserted to be the difference between the two theories. This difference does not follow out of an analysis of the two theories, but comes from the intuitive idea that a microscopical system is more easily perturbed by a macroscopical measuring apparatus than a macroscopical system. It reduces classical mechanics to a very idealized theory, which, as we shall see, is not necessary at all. We should like to propose a concrete example of a physical system that we shall use to make our analysis. Let us consider a physical system which is a piece of wood. We should like to test whether the piece of wood burns well or not. A possible test consists of taking the piece of wood and setting it on fire. In general, when we perform the test on a piece of dry wood, the piece of wood will be destroyed by the test. However, for a piece of dry wood we can "predict" that if we should perform the test, the piece of wood would burn. This is the reason why we give the property of "burning well" to such a piece of dry wood. What we want to point out is that for a test, in general, there are two different aspects that need to be analyzed.

First of all, there is the aspect concentrated on the result of the test. The aim of a physical theory of the physical system is to "*predict*" the result of a certain test, and this prediction is done before the test is carried out and no matter whether the test will be carried out. For the case of the piece of wood and the test that we proposed, the property of "burning well" must be contained in a physical theory of the piece of wood. There is another aspect concentrated on the performance of the test and the changing of the state of the physical system by the performance of the test. In the first place it is not the aim of a physical theory of the physical system to describe this changing of the state. This must be done by a physical theory of the measuring process which is, in fact, a physical theory of the joint system of the measuring apparatus and the physical system. At first sight it seems to be artificial to make a distinction between these two aspects. It is, however, interesting to make this distinction because in general the joint system of the measuring apparatus and the physical system is much more complicated than the physical system itself. As a consequence, it will be much easier to make a physical theory where in the description of the physical system the first aspect is treated. It is also interesting to make this distinction, because the distinction is also made in the two existing physical theories, classical mechanics and quantum mechanics. In classical mechanics only the first aspect of the measurement is considered. The theory does not describe the changing of the state of the system during a measurement. The theory, however, makes predictions about the results of the measurements independently whether they change the state of the system or not. Hence it is not correct to say that in classical mechanics only measurements that do not change the state of the system are considered. Also in quantum mechanics practically only the first aspect of the measurement is also considered. Often it is pretended that the changing of the state is also described by quantum mechanics, in the sense that the state of the system after the measurement is taken to be the eigenstate corresponding to the value of the observable that has been obtained after the measurement. This is, however, only true for a special kind of measurement, which were named by Pauli[6] measurements of the first kind. A spin measurement by the Stern–Gerlach method is such a measurement of the first kind. It would be very easy to define also measurements of the first kind in classical mechanics. For these measurements also in classical mechanics the theory would then describe the changing of the state of the system for such a measurement. The measurement to test whether the piece of wood burns or not is not a measurement of the first kind since if the test is performed and we have gotten the answer yes, the wood does not burn anymore.

Often it is also claimed that the fundamental difference between classical mechanics and quantum mechanics is the fact that in quantum mechanics certain observables cannot be measured together, while in classical mechanics every two observables can be measured together. Usually one refers to measurements of the position of a physical system and the momentum of a physical system. Again in this statement no distinction is made between the two aspects of the measurement. If we consider the second aspect, namely the possibility of performing the measurement and the possibility of describing the changing of the state by the performance of the measurement, then almost never can two measurements be performed together, neither in classical mechanics nor in quantum mechanics. Since classical mechanics does not treat this second aspect of the measurement, no distinction is made in classical mechanics between observables that can be measured together and observables that cannot be measured together; but clearly both sorts of observables do exist. In quantum mechanics the concept of compatibility of observables is introduced. Two observables are compatible iff their

corresponding operators commute. What does this mean? This means that there exist a complete set of states such that every state of this set of states is an eigenstate of the two observables. Hence for such a state we can predict the value of both observables. However, from the theory does not follow at all that such compatible observables can be measured together, as is often vaguely stated. This can only be deduced if we assume that we allow only measurements of the first kind. So we see that compatibility of observables is not so easy to interpret if we think of the second aspect of the measurement. For the first aspects of the measurements the meaning is very clear; namely, it means that there exists a set of states such that when the system is in one of the states of this set the value of both observables can be predicted.

So we can conclude from this analysis that there is no difference between classical mechanics and quantum mechanics if we regard the second aspect of the measurement. In both cases the state of the system is very often changed by the performance of the experiment. There is, however, a fundamental difference between the two theories concerning the first aspect of the measurement, namely the possibility in being able to predict the outcome of a measurement. In classical mechanics, if we consider an arbitrary observable, then for an arbitrary state of the system we can predict the outcome of an eventual measurement of the observable. In quantum mechanics this is only the case for states that are eigenstates of the operator corresponding to this observable. It seems to be so that when the system is not in such a state, the observable is not an "element of reality" for the system in question.

We use here the word "element of reality" as it was introduced by Einstein, Podolsky, and Rosen.[7] It is this difference that gives rise to the uncertainty relations of Heisenberg for two observables that cannot be predicted together (that are not an "element of reality" at the same time.) Again, it is not the fact that they cannot be measured together that gives rise to the uncertainty relations. Since in classical mechanics an arbitrary observable is always an element of reality for the system (we can always predict with certainty the result of a measurement for a certain observable), for every couple of observables, each observable is, of course, also, at the same time as the other one, an element of reality for the system, this independently of whether the observables can be measured together or not.

We want again to call attention to the fact that, to be able to make a prediction for two observables at the same time, it is not at all necessary to be able to measure the observables at the same time. Let us consider again the example of the physical system which is a piece of wood. We are going to consider two observables. The observable γ which consists of testing whether the piece of wood burns well and the observable δ which consists of testing whether the piece of wood floats on water. Both experiments change in general the state of the piece of wood and it is clear that they cannot be performed together. We can, however, for every state of the piece of wood predict whether it will burn well and whether it will float on water, and we can make this prediction at the same time. This fact is analyzed in greater detail on the same example in Refs. 3 and 4 and also can be retraced in the definition of the product of two questions as defined by Piron in Ref. 2. After this analysis it becomes clear that we can distinguish very well between the observables of classical mechanics and the observables of quantum mechanics.

Suppose now that we have a physical system S and suppose that we know what an observable is for such a system S; then it is very easy to define the concept of "classical observable" for such a system S, and this by using only experimentally verifiable statements.

Definition: An observable of a physical system is a classical observable iff for every state of the physical system we can predict the value of the observable in this state.

The observables used in classical mechanics are classical observables. In quantum mechanics none of the observables used is a classical observable. This fact is already an indication of the fact that quantum mechanics is perhaps not a more general theory than classical mechanics, because, as we shall explain in the following, also microscopical systems have in general classical observables. There are just these classical observables that give rise to superselection rules. The idea of characterizing a classical system by the fact that for an arbitrary experiment on such a system for any state of the system the outcome of the experiment is certain can be found in Ref. 8. There is also emphasized that such an hypothesis is not equivalent to the hypothesis of determinism of the outer world.

I. THE DESCRIPTION OF SUPERSELECTION RULES IN THE QUANTUM LOGIC APPROACH

As we remarked already, if one wants to describe superselection rules in quantum mechanics, one has to do this *a posteriori* by decomposing the Hilbert space into coherent subspaces. For the case of continuous superselection variables one Hilbert space is not appropriate anymore and a set of Hilbert spaces $\mathscr{H}_i$ is needed. This set of Hilbert spaces find their "natural" representation in the direct union of the irreducible Hilbert space lattices $P(\mathscr{H}_i)$, where $P(\mathscr{H}_i)$ is the lattice of all closed subspaces of the Hilbert space $\mathscr{H}_i$. This is the way of describing superselection rules as introduced by Piron,[1] and this is also the way in which superselection rules appear in quantum logic. We shall shortly explain this structure of direct union because we shall find a similar structure if we try to entangle the classical part and the nonclassical part of the description of a physical system.

In quantum logic a physical system is described by the collection of all the properties of this physical system. A property is in fact a yes–no observable (an observable having two possible outcomes yes and no). In classical mechanics the yes–no observables are represented by functions on the phase space to the set $\{0,1\}$. The set of all these functions forms a complete Boolean algebra. In quantum mechanics the yes–no observables are represented by the projection operators of the Hilbert space. The set of all these projection operators forms a lattice that is complete, orthocomplemented, weakly modular, and atomic. It is never a Boolean algebra if the dimension of the Hilbert space is greater than 1. Now a complete Boolean algebra is also a complete orthocomplemented weakly modular lattice. In quantum logic one postulates therefore often that the set of properties con-

cerning a physical system is a complete orthocomplemented weakly modular lattice. Let us shortly define what is a complete orthocomplemented weakly modular lattice.

Definitions 1.1: (i) A set $\mathscr{T}$ is a "partially ordered" set, if there exists a relation which has the properties

$$a < a,$$

$$a < b \text{ and } b < c \Rightarrow a < c,$$

$$a < b \text{ and } b < a \Rightarrow a = b.$$

If we have a family a_i of elements of I we will denote the infimum of this family by $\wedge_i a_i$ provided that this infimum exists. We will denote the supremum of this family by $\vee_i a_i$ provided that this supremum exists. So we must have

$$x < a_i \quad \forall\ i \Leftrightarrow x < \bigwedge_i a_i,$$

$$a_i < y \quad \forall\ i \Leftrightarrow \bigvee_i a_i < y.$$

(ii) A partially ordered set $\mathscr{L}$ is called a "complete lattice" if, for each family $a_i \in \mathscr{L}$, $\wedge_i a_i$ and $\vee_i a_i$ exist.

(iii) If $\mathscr{T}$ is a partially ordered set, we will say that $\mathscr{T}$ is orthocomplemented iff $\mathscr{T}$ has a least element O and if there exists a map $a \to a'$ of $\mathscr{T}$ onto itself which satisfies

$$a < b, \quad \text{then} \quad b' < a',$$

$$a'' = a \quad \text{and} \quad a \wedge a' = 0,$$

the mapping $a \to a'$ is called an orthocomplementation and a' is called the orthocomplement of a.

(iv) If $\mathscr{T}$ is a partially ordered set that is orthocomplemented, then $\mathscr{T}$ is said to be weakly modular iff

$$\text{for } a < b \quad \text{we have} \quad (a \vee b') \wedge b = a.$$

Often in quantum logic one postulates only a weaker structure for the set of propositions of a physical system, this depending on what one wants to do; but one can say that the structure of a complete orthocomplemented lattice is the one used when one wants to do physics with quantum logic (see Ref. 2). A complete orthocomplemented weakly modular lattice where the distributive law between the infimum and the supremum is valid is called a Boolean lattice or a Boolean algebra. One can define the concept of compatible properties in quantum logic.

Definition 1.2: Two properties a, b of a complete orthocomplemented lattice are compatible iff the lattice generated by $\{a, a', b, b'\}$ is a Boolean lattice.

For quantum mechanics we have that two properties represented by projection operators are compatible iff the projection operators commute. In classical mechanics every pair of properties is a compatible pair. On the other hand, if every pair of properties of a complete orthocomplemented weakly modular lattice is compatible, then this lattice is a Boolean lattice. This is the reason why in quantum logic one sees the difference between classical mechanics and quantum mechanics as a difference in structure of the set of properties of both theories. Although a great step forward has been done, in the sense that in quantum logic both theories can be described by the same formalism, we are not totally satisfied with this approach. Indeed expressing the classicality of a yes–no experiment by means of this relation of compatibility is not very satisfactory because this relation of compatibility is an algebraic relation that is, just as the relation of commutativity of operators in ordinary quantum mechanics, not physically interpretable.

Let us give now the construction of the direct union of lattices and then explain how superselection rules appear in quantum logic.

Definition 1.3: Suppose that $\mathscr{L}_i$ is a collection of complete lattices. We shall denote the direct union of the $\mathscr{L}_i$ by $\bigotimes_i \mathscr{L}_i$. An element $b \in \bigotimes_i \mathscr{L}_i$ will be written $\bigotimes_i b_i$, where $b_i \in \mathscr{L}_i$. We define a partial order relation as follows:

$$\bigotimes_i b_i < \bigotimes_i c_i \Leftrightarrow b_i < c_i \quad \forall i.$$

It is then very easy to show that $\bigotimes_i \mathscr{L}_i$ is a complete lattice iff every $\mathscr{L}_i$ is a complete lattice. If $\mathscr{L}_i$ are orthocomplemented we define an orthocomplementation on $\bigotimes_i \mathscr{L}_i$ as follows:

$$\left(\bigotimes_i b_i\right)' = \bigotimes_i b_i'.$$

It is then again easy to check that $\bigotimes_i \mathscr{L}_i$ is weakly modular iff every $\mathscr{L}_i$ is weakly modular.

Suppose now that we have an arbitrary complete orthocomplemented weakly modular lattice, then we can prove the following.

Theorem 1.4: The center of a complete orthocomplemented weakly modular lattice is a complete Boolean lattice.

Proof: See Ref. 2, p. 29.

It is by means of this center that one can distinguish between the classical case and the quantum case, and mixed cases. If the center of the lattice is the whole lattice, then the system is described by a classical theory. If the center of the lattice contains only O and I, we have the pure nonclassical case. Quantum mechanics without superselection rules can only describe such a pure nonclassical case, because the center of the lattice $P(\mathscr{H})$ of all the projection operators of the Hilbert space $\mathscr{H}$ does contain only O and I. Such a lattice is called irreducible. To be able to write down the main theorem that can be proved in quantum logic, we have to give some more definitions.

Definition 1.5: An element p of a lattice is called an atom iff whenever a is an element of the lattice such that $0 < a < p$, then $a = 0$ or $a = p$.

Definition 1.6: A lattice is said to be atomic if for every element a there is at least one atom $p < a$.

The structure theorem that shows that every lattice of properties is the direct union of irreducible lattices can only be proved for atomic lattices that are weakly modular.

Theorem 1.7 (Piron): Every atomic complete orthocomplemented weakly modular lattice is the direct union of irreducible lattices.

Proof: See Ref. 2, p. 35.

We shall show that an analog decomposition exists for lattices of properties that are not necessarily weakly modular and atomic. Our aim is, however, as we told in the introduction to make such a construction starting with the concept of classical observable and not with the concept of compatibility which is not clear at all physically.

II. THE FORMALISM

We want to introduce classical mechanics in the following way. We have a physical system. In general this physical system can have classical observables and nonclassical observables. A theory that studies only the classical observables of the physical system in question will be classical theory. A theory that wants to study the physical system in more detail must often allow also nonclassical observables. Quantum mechanics is a theory that allows the study of nonclassical observables. There are, however, two things that go wrong with quantum mechanics. First of all, it allows only nonclassical observables of a very specific nature, which is due to the very specific mathematical structure of quantum mechanics. As a consequence quantum mechanics cannot, for example, describe separated systems (see Refs. 3 and 4). Secondly, it does not allow the description of classical observables. Hence it can in a certain sense only describe the nonclassical part of the physical system. Due to these two shortcomings of quantum mechanics we certainly cannot formulate our problem in a theory as quantum mechanics. Quantum logic does not have the second shortcoming; it allows the description of classical observables and nonclassical observables. However, again due to its specific mathematical structure it still can only describe nonclassical observables of a very specific nature. We want to be able to formulate our problem in a theory without these shortcomings. Moreover, in this theory it has to be possible to define the concept of classical observable as we put forward in the Introduction. Piron introduces the concept of "question" to give a physical meaning to the concept of proposition (yes–no observable) that is used in quantum logic.[2] He then introduces the lattice of properties of a physical system from this concept of question. He also defines a set of axioms on this lattice, such that when these axioms are satisfied, the theory becomes a theory equivalent with quantum mechanics, but allows the description of superselection rules as explained in Sec. I. Of course, his aim was to clarify quantum mechanics, and therefore he was looking for a set of axioms that would reduce the *a priori* more general theory to a theory as quantum mechanics, with superselection rules. Two of the axioms, namely the weak modularity and the covering law (axiom P and axiom A_2 in Ref. 2) do not allow the description of separated physical systems (see Refs. 3 and 4). Also axiom C of Ref. 2 has to be weakened in a certain sense if one wants to avoid paradoxical situations for the description of separated systems. Therefore, we will not retain the axioms of Piron but only the structure of his theory without the additional structure implied by the axioms. We will also not only use this concept of question to justify the structure of the set of properties, but we will use the concept of questions as a basic work object in the theory. We shall shortly recall some definitions, but the theory as we will use it is explained in Refs. 3 and 4. In Refs. 3 and 4 another set of axioms is put forward, enabling us to just drop the two wrong axioms, weak modularity and the covering law. We want to mention that no claim of truth is implied in the term axiom as it is used here. The axioms must merely be seen as physical hypothesis. First of all, we introduce the concept of *entity* to make clear what we mean by a physical system. An entity corresponds to a *phenomenon* that we can experience without being forced to experience also the rest of the world. It is an idealization in the sense that we decide to study a well-defined set of properties of the phenomenon. It is possible to make *statements* about the state of the phenomenon. Such a statement only defines a *property* of the phenomenon if it is *testable*. A proposal of such an experiment to test a statement is called a *question*. Hence to define a question one has to define:

—the measuring apparatus used to perform the experiment

—the manual of operation of the apparatus

—a rule that allows us to interpret the result in terms of "yes" and "no"

A property of a phenomenon can be *actual*, the entity has the property "*in acto*" or *potential*, the entity has the possibility of obtaining the property.

A. Testing of properties and the concept of truth

A question α of an entity S is said to "*true*" (and the corresponding property is said to be "actual") iff when we should decide to perform the test proposed by α, the expected answer "yes" would come out with certainty.

B. Inverse questions

If α is a question of the entity S, we can consider the question that consists of proposing the same test as the one proposed by α, but changing the role of yes and no. We will denote this new question by $\alpha^{\sim}$, and call it the *inverse* question.

C. Testing several properties at once

If we have a family of properties a_i and questions α_i testing a_i, a question that tests the actuality of all the properties a_i, and which we will denote by $\pi_i\alpha_i$ and which we will call the *product* of the α_i is the following:

We choose as we want, at random or not one of the α_i and accord to $\pi_i\alpha_i$ the answer obtained by performing the test of this chosen question. Clearly, $\pi_i\alpha_i$ is true iff α_i is true for every i.

D. A generating set of questions

We shall denote by Q the set of questions of the entity S. We will consider Q to be closed for the "product" operation and for the "inverse" operation. Hence, if $\alpha_i \in Q$, then $\pi_i\alpha_i \in Q$, and, if $\alpha \in Q$, then $\alpha^{\sim} \in Q$. We can see very easily that $(\pi_i\alpha_i)^{\sim} = \pi_i\alpha_i^{\sim}$. A subset $G \subset Q$, such that if $\alpha \in G$ we have $\alpha^{\sim} \in G$ and such that

$$Q = \{\pi_i\alpha_i \,|\, \alpha_i \in G\}$$

is called a "*generating set*" of questions.

E. A physical law on the questions of an entity

If we have the situation that whenever a question α is true, then also the question β is true, we shall denote this as

$$\alpha < \beta$$

and we shall say "α is *stronger* than β." This physical law has

the following properties:

(1) $\alpha < \alpha$,

(2) if $\alpha < \beta$ and $\beta < \gamma$, then $\alpha < \gamma$,

where α, β, and γ are questions. Hence $<$ is a preorder relation on Q.

F. Properties of an entity

If α and β are questions of an entity S such that

$$\alpha < \beta \quad \text{and} \quad \beta < \alpha$$

then we will say that α is *equivalent* to β and we will denote $\alpha \approx \beta$. If $\alpha \approx \beta$, then α and β test the same property of the entity. This is why we shall identify the properties of the entity with the classes of equivalence of questions. The collection of properties of the entity we will denote by $\mathscr{L}$. The collection of questions that are never true we will denote by O. For an arbitrary question α we have $\alpha \cdot \alpha^{\sim} \in O$. It is easy to see that $O \in \mathscr{L}$.

A *trivial question* is a question that is always true. If τ_1 is a trivial question and τ_2 is a trivial question, then $\tau_1 \approx \tau_2$. Hence all the trivial questions define a property that we will denote by I. The preorder relation $<$ on the set of questions induces a relation on the set of properties, if $a, b \in \mathscr{L}$

$$a < b \quad \text{iff} \quad \alpha < \beta \quad \text{for} \quad \alpha \in a \text{ and } \beta \in b.$$

It is easy to see that $<$ is a "partial order relation." Hence $\mathscr{L}$ is a "*partially ordered set.*"

If α_i is a family of properties and $\alpha_i \in a_i$, let us then denote the property tested by $\pi_i \alpha_i$ by $\wedge_i a_i$.
It is easy to see that $\wedge_i a_i$ is an infimum of the family a_i. Let us define now for an arbitrary family a_i

$$\bigvee_i a_i = \bigwedge_{a_j < b \ \forall j,\ b \in \mathscr{L}} b.$$

Then $\vee_i a_i$ is a supremum for the family a_i. This shows that $\mathscr{L}$ is a "*complete lattice.*"

G. The set of states of an entity

The state of an entity is the set ϵ of all actual properties. We can remark that this state is totally determined by the infimum of this set ϵ. Indeed if

$$p = \bigwedge_{a \in \epsilon} a$$

then $\epsilon = \{a | p < a,\ a \in \mathscr{L}\}$. In the following we will represent the state ϵ of the entity by the property p. We will denote by Σ the set of all states.

We can see that a is actual iff the entity is in a state p such the $p < a$. This shows that for every $a \in \mathscr{L}$ we have

$$a = \bigvee_{p < a} p.$$

Therefore, we will say that Σ is a "*full set of states*" for $\mathscr{L}$.

H. An orthogonality relation

If p and q are two states of S, we will say that p is *orthogonal* to q, iff there is a question γ such that γ is true if S is in the state p and $\gamma^{\sim}$ is true if S is in the state q. We will then denote $p \perp q$. If $p, q, r, s \in \Sigma$, then

(1) $p \perp q \Rightarrow q \perp p$,

(2) $p \perp q$ and $r < p$ and $s < q$, then $r \perp s$,

(3) $p \perp q \Rightarrow p \wedge q = 0$.

We shall say that two properties $a, b \in L$ are orthogonal iff for every $p, q \in \Sigma$ such that $p < a$ and $q < b$ we have $p \perp q$. We shall also denote $a \perp b$.

In Refs. 3 and 4 is shown in which way this formalism can be found in classical mechanics and in quantum mechanics. As we remarked already in the Introduction, in classical mechanics every property of the entity corresponds with a subset of the state space, namely the subset of all those states that make the property actual. In quantum mechanics every property corresponds to a projection operator, because these are indeed the self-adjoint operators with two possible outcomes, yes and no, or we can also say that every property corresponds to a closed subspace of the Hilbert space, namely the closed subspace of all the eigenstates of the projection operator with eigenvalue 1.

I. Elements of reality and completeness of the theory

Let us recall the definition of an element of reality given by Einstein, Podolsky, and Rosen (EPR):

"If without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then there exists an element of reality corresponding to this physical quantity."

If we know that the proposal of a test has an answer that is certain, then we know that one of the questions α or $\alpha^{\sim}$ corresponding to this test is true. So we see that the "*true questions*" that we defined are just the elements in our theory that correspond with the elements of reality of the entity. Now our theory examines a set of questions of a phenomenon. This set of questions defines an entity. The condition of completeness put forward by EPR is the following: "A theory is complete if every element of reality has a counterpart in the theory." Certainly, EPR did not mean that a theory should describe all the possible elements of reality of the phenomenon. A theory never describes exactly the phenomenon, but always an entity corresponding to this phenomenon. Therefore, we shall say: "A theory is complete if it can describe every possible element of reality of the phenomenon, without leading to contradictions."

In Refs. 3 and 4 we show that this is not the case for quantum mechanics. By construction this is the case for the theory that we put forward. If we add elements of reality to the entity, we just have to add the corresponding questions, and we will never find contradictions since the structure of the theory does not change by adding questions or taking questions away.

III. THE CLASSICAL PART OF THE DESCRIPTION OF AN ENTITY

A. Classical questions and classical properties

After the analysis that we made about the difference between classical mechanics and quantum mechanics, we can very easily invent the definition of a "*classical question.*"

Definition 3.1: A classical question is a question for which we can predict the answer for every state of the entity.

It is very easy to see that α is a classical question iff α is true or $\alpha^{\sim}$ is true for any state of the entity.
The two questions γ and δ defined in the Introduction for the piece of wood are both classical questions. The question $\gamma\cdot\delta$ is not a classical question. Indeed it is very easy to put the piece of wood in such a state that neither $\gamma\cdot\delta$ nor $(\gamma\cdot\delta)^{\sim}$ are true: for example, if the piece of wood is wet and floats on water. Then the answer that we get for the question $\gamma\cdot\delta$ can be "yes" or "no" depending on whether during the test corresponding to $\gamma\cdot\delta$ we choose to perform the question δ or the question γ. Let us explain why this is the case. Suppose that we denote by c the property of the piece of wood tested by γ and by d the property of the piece of wood tested by δ. Hence c is the property "the piece of wood burns well" and d is the property "the piece of wood floats on water."

Let us denote by c' the property tested by $\gamma^{\sim}$ and by d' the property tested by $\delta^{\sim}$. Hence c' is the property "the piece of wood does not burn well" and d' is the property "the piece of wood does not float on water."

The property tested by $\gamma\cdot\delta$ is $c\wedge d$ "the piece of wood burns well and floats on water" and the property tested by $(\gamma\cdot\delta)^{\sim}=\gamma^{\sim}\cdot\delta^{\sim}$ is $c'\wedge d'$ "the piece of wood does not burn well and does not float on water." For a piece of wet wood that floats on water both properties $c\wedge d$ and $c'\wedge d'$ are potential. This example shows that, first of all, there is no "logical" necessity for a question to be classical and, secondly, it is very easy to find an example of a nonclassical question. We can even show that every nontrivial product question is a nonclassical question.

Theorem 3.2: If α_i are questions of an entity S, then $\pi_i\alpha_i$ is a classical question iff for every i, j we have $\alpha_i\approx\alpha_j$ and α_i are classical questions.

Proof: Suppose that $\pi_i\alpha_i$ is a classical question; then $\pi_i\alpha_i$ is true or $(\pi_i\alpha_i)^{\sim}=\pi_i\alpha_i^{\sim}$ is true. Suppose that α_j is true, then $\pi_i\alpha_j^{\sim}$ is not true. As a consequence, $\pi_i\alpha_i$ is true. Hence α_i is true $\forall$ i. Clearly also α_i are classical questions. ■

Hence a classical question can only be a product question of equivalent classical questions. We could have expected this result since from the definition of a classical question immediately follows that a classical question is a primitive question as defined in Ref. 4. Already a primitive question can only be a product question of equivalent primitive questions as is shown in Theorem 4.2 of Ref. 4. However, every primitive question is not necessarily a classical question. It is this fact which for the first time appeared in quantum mechanics. Quantum mechanics treats primitive questions that are not classical questions.

B. The classical property lattice

The problem that we want to consider is the following: Which kind of theory do we find if we decide for a certain phenomenon to consider only the classical questions of the phenomenon and to study the set of properties generated by these classical questions. Let us introduce the necessary symbols to be able to treat this problem. Let us call K the set of all classical questions of the entity S, and let us call C the set of questions generated by K. Hence

$$\mathrm{C}=\{\pi_i\alpha_i\,|\alpha_i \text{ is a classical question}\}.$$

Definition 3.3: A property of the entity S that can be tested by a product of classical questions will be called a classical property.

Let us denote by $\mathscr{C}$ the set of all classical properties. We shall call $\mathscr{C}$ the "*classical property lattice*" of S. We shall show that the study of the classical properties of the entity S can always be done by a theory as classical mechanics. First of all, we remark that everything that can be shown for the structure of an arbitrary property lattice is, of course, also true for $\mathscr{C}$. Hence $\mathscr{C}$ is a complete lattice. We have to be careful now. Indeed, if $a\in\mathscr{C}$, this means that there exists a question $\pi_i\alpha_i$ of S testing a, where α_i are classical questions of S. Now $\pi_i\alpha_i$ determines uniquely an element of $\mathscr{L}$. Let us denote this element by $\mathrm{f}(a)$. Mathematically a and $\mathrm{f}(a)$ are different objects. Indeed a is an equivalence class of questions of $\mathscr{C}$ while $\mathrm{f}(a)$ is an equivalence class of questions of Q. Hence we have that if $\alpha\in a$ then $\alpha\in\mathrm{f}(a)$, but not necessarily the inverse. Physically, of course, a and $\mathrm{f}(a)$ indicate the same property of the phenomenon under consideration. Let us study more in detail this classical property lattice of the entity S.

C. Classical mixtures and the classical state space

If ϵ is the collection of all properties that are actual for the entity S, as we explained in Sec. II G we represent the state of S by means of the minimal element of ϵ. We can consider now η, the collection of all classical properties that are actual. Then $\mathrm{f}(\eta)\subset\epsilon$. This collection η we will call the "*classical mixture*" of the entity S. Again as we did for the state of S we shall represent this classical mixture of S by the minimum of this collection η. So $w=\wedge_{a\in\eta}a$.

Note that for every state p we find a unique classical mixture w_p such that $p<f(w_p)$. But, if w is actual, the entity can be in different states. This is the reason why we call w a mixture. The collection of all classical mixtures of the entity we will denote by Ω, and we will call Ω the "*classical state space*" of S. From Sec. II G follows that Ω is a full set for $\mathscr{C}$ and from II H we have an orthogonality relation that is defined on Ω and on $\mathscr{C}$. Namely, two classical mixtures w and v are orthogonal iff there exists a classical question γ such that γ is true if S is in the classical mixture w and $\gamma^{\sim}$ is true if S is in the classical mixture v. We shall denote $w\perp v$.

Theorem 3.4: Two different classical mixtures of S are always orthogonal and the classical mixtures of S are atoms of the classical property lattice C of S.

Proof: Suppose that w and v are two different classical mixtures of S. There exist questions $\pi_i\alpha_i\in w$ and $\pi_j\beta_j\in v$ such that α_i and β_j are classical questions. Since $w\neq v$, we must have $w\not<v$ or $v\not<w$. Suppose $w\not<v$. Suppose that S is in the classical mixture w. Then v is not actual. So there is at least one j such that β_j is not true. But since β_j is a classical question, it follows that $\beta_j^{\sim}$ is true. If S is in the classical mixture v, then β_j is true. This shows that $w\perp v$.

Let us consider now an arbitrary classical mixture w of the entity S. Suppose b is a classical property of S such that $0<b<w$ and $b\neq 0$. If b is actual, the entity is in a classical mixture $v<b$. But then $v<w$. From this follows that $v\wedge w=v$. But then v cannot be orthogonal to w. Hence

$v = w$. As a consequence, $b = w$, which shows that w is an atom. ■

This theorem shows that the orthogonality relation on the classical states space becomes trivial as is indeed the case in classical mechanics, where one does not use the notion of orthogonal states. This theorem also shows that the classical mixtures of the entity become atoms of the classical property lattice $\mathscr{C}$. As a consequence, we can say that points $\{w\}$, where $w \in \Omega$, really represent the classical mixtures of the entity as one has in classical mechanics. To be able to see that the part of the entity represented by the classical properties can really be described by a theory as classical mechanics, we shall introduce the state space description of an entity.

D. The state space description of an entity

If Σ is the set of states of the entity S, we can consider the lattice $P(\Sigma)$ of all subsets of Σ. We can consider then the map that makes correspond with every property a the set $\mu(a)$ of all the states that make a actual. Hence

$$\mu: \mathscr{L} \to P(\Sigma)$$
$$a \to \{p \mid p < a, p \in \Sigma\}.$$

It is easy to see that μ has the following properties:

Theorem 3.5: If a, b, $a_i \in \mathscr{L}$, then:
(i) $a < b$ iff $\mu(a) \subset \mu(b)$,
(ii) μ is injective,
(iii) $\mu(\wedge_i a_i) = \cap_i \mu(a_i)$,
(iv) $\mu(\mathsf{O}) = \Phi$ and $\mu(\mathsf{I}) = \Sigma$,
(v) $a \perp b \Rightarrow \mu(a) \perp \mu(b)$.

Proof: See Ref. 3, Theorem 3.1. ■

The reason why it is impossible to describe an entity by just considering the set of states of the entity is because, first of all, one loses the orthogonality relation, but, even when we should think of a state space with an orthogonality relation, it would in general not work. This is so because the points $\{p\}$ of $P(\Sigma)$ do not necessarily correspond to states of the entity. Indeed, if $p \in \Sigma$ and p is not an atom of $\mathscr{L}$, then there exists at least one $q \in \Sigma$ such that $q \neq p$ and $q < p$. But then $\{q,p\} \subset \mu(p)$ such that $\mu(p) \neq \{p\}$ and $\{p\}$ does not correspond to a state of the entity. In this case, of course, it makes no sense to try to describe the entity by means of Σ alone without considering $\mathscr{L}$.

Let us define $\overline{\Omega} = \{\{v\} \mid v \in \Omega\}$, then $\overline{\Omega}$ is a full set for $P(\Omega)$, and in the usual state space description of classical mechanics it is $\overline{\Omega}$ that represents the set of states. We shall also define the trivial orthogonality relation in $P(\Omega)$, which is the following:

$$\{v\} \perp \{w\} \Leftrightarrow \{v\} \neq \{w\}.$$

Then we can show that, for the classical state space of the entity, problems of the kind mentioned above do not occur such that we can describe the classical part of the entity by using $\overline{\Omega}$ and classical mechanics. Let us therefore consider the map

$$\mu_C: \mathscr{C} \to P(\Omega),$$
$$a \to \{v \mid v < a; v \in \Omega\}.$$

Theorem 3.6: The map μ_C is an isomorphism of Ω onto $\overline{\Omega}$.

Proof: Since the classical mixtures are atoms of C we have that for every $v \in \Omega$, $\mu_C(v) = \{v\}$. We also have

$$v \perp w \Leftrightarrow v \neq w \Leftrightarrow \{v\} \neq \{w\} \Leftrightarrow \{v\} \perp \{w\} \Leftrightarrow \mu_C(v) \perp \mu_C(w). \quad ■$$

Often it is claimed that the lattice of properties of an entity described by classical mechanics in a state space Ω must be isomorphic to $P(\Omega)$. The classical property lattice $\mathscr{C}$ is, in general, however, not isomorphic to $P(\Omega)$, because the map μ_C need not be surjective. The reason that one claims that $\mathscr{C}$ should be isomorphic to $P(\Omega)$ is because once again does not make a distinction between the "statements" that can be made about the state of an entity and the "properties" of an entity. Of course, the set of all statements that can be made about the state of an entity must be isomorphic to $P(\Omega)$, because we can always put an arbitrary statement in the following form: "Is the state v of the entity contained in the subset A of Ω" and in this way make correspond to this statement the element A of $P(\Omega)$. This defines an isomorphism between the set of statements and $P(\Omega)$.

A statement does, however, only define a property if it is testable as we explained in II A. This is the reason why in general the map μ_C is not surjective. Let us try to see this with an example. We consider again the phenomenon which is a piece of wood and we suppose that we want to study only the two questions γ and δ.

Let use construct the property lattice of this entity. A generating set of questions is the following:

$$\mathrm{G} = \{\tau, \tau^-, \gamma, \gamma^-, \delta, \delta^-\},$$

where τ is a trivial question. The set of properties corresponding to this generating set is the following:

$$\mathscr{G} = \{\mathsf{I}, \mathsf{O}, c, c', d, d'\}.$$

We have

$$\gamma \cdot \delta \in c \wedge d, \quad \gamma^- \cdot \delta \in c' \wedge d, \quad \gamma \cdot \delta^- \in c \wedge d', \quad \gamma^- \cdot \delta^- \in c' \wedge d'.$$

These are the only new properties defined by the product questions. They are also the states of the piece of wood. As a consequence,

$$\mathscr{L} = \{\mathsf{O}, \mathsf{I}, c, c', d, d', c \wedge d, c' \wedge d, c \wedge d', c' \wedge d'\}$$

and

$$\Sigma = \{c \wedge d, c \wedge d', c' \wedge d, c' \wedge d'\}.$$

Since both questions γ and δ are classical questions, we have, of course, $\mathscr{C} = \mathscr{L}$, $\Omega = \Sigma$. (See Fig. 1.)

Let us consider now the map

$$\mu = \mathscr{L} \to P(\Sigma);$$

then it is easy to see that μ is not a surjective map. For example, the element $\{c \wedge d, c \wedge d', c' \wedge d\}$ is not an image of μ.

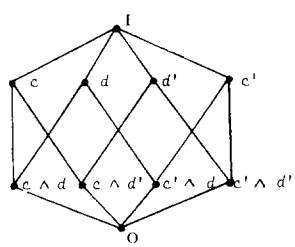


FIG. 1.

This element corresponds to the statement "The piece of wood burns well or floats on water."

There is *a priori* no question to test this statement. This comes from the fact that the performance of the test γ and δ corresponds to different experimental setups that cannot be realized together. Indeed, it is possible in this case to introduce an experiment that makes this statement testable. The experiment is, for example, the following:

We take the piece of wood and break it into two pieces, and we perform the test γ on one of the pieces and the test δ on the other piece. This experiment has four possible outcomes {yes, yes}, {yes, no}, {no, yes}, and {no, no}. We can define new questions by means of this experiment. We define the questions:

$\gamma \Delta \delta$ that consists of performing the experiment and giving the answer yes if we have the outcome {yes, yes}; otherwise, we give the answer no.

$\gamma \nabla \delta$ that consists of performing the experiment and giving the answer yes if we have one of the outcomes {yes, yes}, {yes, no}, or {no, yes}. We give the answer no if we have the outcome {no, no}.

$\gamma \Theta \delta$ that consists of performing the experiment and giving the answer yes if we have one of the outcomes {yes, yes} or {no, no}. We give the answer no if we have one of the outcomes {yes, no} or {no, yes}.

$\gamma \Delta \delta$, $\gamma \nabla \delta$, and $\gamma \Theta \delta$ are classical questions. $\gamma \Delta \delta$ is a question that tests whether the piece of wood burns and floats, $\gamma \nabla \delta$ is a question that tests whether the piece of wood burns or floats, and $\gamma \Theta \delta$ is a question that tests whether the piece of wood burns and floats or whether it does not burn and does not float. Of course, we suppose here that the breaking of the piece of wood into two pieces does not change the properties c and d of the piece of wood. So we suppose that we can attribute properties to the original piece of wood by making tests on pieces of this original piece. We feel very well that this procedure will not hold for an arbitrary entity.

A set of generating questions is now the following:

$$\begin{aligned} G = \{&\tau,\ \tau^{\sim},\ \gamma,\ \gamma^{\sim},\ \delta,\ \delta^{\sim},\ \gamma \Delta \delta,\ \gamma \Delta \delta^{\sim},\\ &\gamma^{\sim} \Delta \delta,\ \gamma^{\sim} \Delta \delta^{\sim},\ \gamma \nabla \delta,\\ &\gamma^{\sim} \nabla \delta,\ \gamma \nabla \delta^{\sim},\ \gamma^{\sim} \nabla \delta^{\sim},\ \gamma \Theta \delta, \gamma^{\sim} \Theta \delta\}. \end{aligned}$$

The set of properties corresponding to G is the following:

$$\begin{aligned} \mathscr{G} = \{&I, O, c, c', d, d',\ c \wedge d,\ c \wedge d',\ c' \wedge d,\\ &c' \wedge d',\ c \vee d,\ c \vee d',\ c' \vee d,\ c' \vee d',\\ &(c \wedge d) \vee (c' \wedge d'),\ (c \wedge d') \vee (c' \wedge d)\}. \end{aligned}$$

We can see very easily that no new properties are defined by the product questions. Hence $\mathscr{L} = \mathscr{G}$.

The set of states did not change by adding these new questions, which shows that they are not so important for the theory. So

$$\Sigma = \{c \wedge d, c \wedge d', c' \wedge d, c' \wedge d'\} \quad \text{(see Fig. 2)}.$$

We have

$$(c \vee d) \wedge (c' \vee d') = (c \wedge d') \vee (d \wedge c'),$$
$$(c \vee d') \wedge (c' \vee d) = (c \wedge d) \vee (c' \wedge d').$$

If we now consider the map μ: $\mathscr{L} \to P(\Sigma)$, we see that it is an isomorphism. For example,

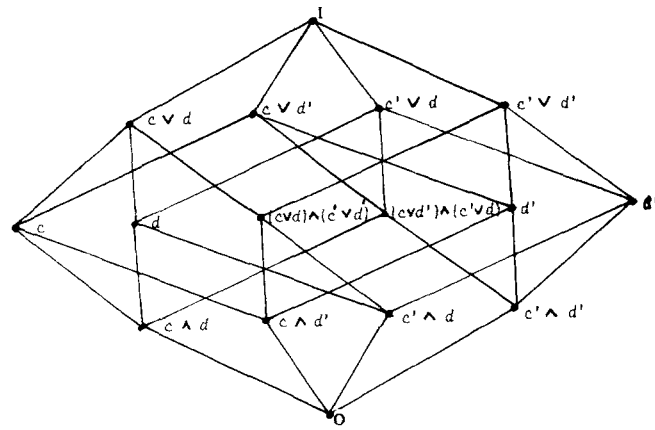


FIG. 2.

$\mu(c \vee d) = \{c \wedge d,\ c \wedge d',\ c' \wedge d\}$, which was the missing statement.

We can now wonder what would be the weakest axiom that we can formulate, such that we do have an isomorphism between $\mathscr{C}$ and $P(\Omega)$.

Theorem 3.7: The

$$\begin{aligned} &\mu_C : \mathscr{C} \to P(\Omega)\\ &\quad a \to \{w \mid w < a, w \in \Omega\} \end{aligned}$$

is an isomorphism iff for every classical mixture w the statement "the entity is in a classical mixture different from w" is a classical property.

Proof: Suppose that μ_C is an isomorphism. Consider then the property $a = \mu_C^{-1}(\Omega \setminus \mu_C(w))$. Then we have: a is actual $\Leftrightarrow$ the entity is in a classical mixture $v < a$, $\Leftrightarrow$ the entity is in a classical mixture v such that $\mu_C(v) \subset \Omega \setminus \mu_C(w)$, $\Leftrightarrow$ the entity is in a classical mixture $v \neq w$.
Suppose now that for an arbitrary classical mixture w the statement "the entity is in a classical mixture different from w" is a classical property, and let us denote this classical property by w'. To show that μ_C is an isomorphism, we only have to show the surjectivity of μ_C. Suppose $A \in P(\Omega)$. Consider the classical property

$$\begin{aligned} &a = \wedge w'\\ &\quad \mu_C(w) \subset (\Omega \setminus A). \end{aligned}$$

Then

$$\begin{aligned} \mu_C(a) &= \bigcap_{\mu_C(w) \subset (\Omega \setminus A)} \mu_C(w')\\ &= \bigcap_{\mu_C(w) \subset (\Omega \setminus A)} (\Omega \setminus \mu_C(w))\\ &= \Omega \setminus (\Omega \setminus A) = A. \end{aligned}$$

■

E. The classical part of the description of an entity and classical entities

As we showed in the foregoing, we can study the classical part of an entity by means of the classical property lattice which leads to a theory as classical mechanics. Let us now see in which way this classical property lattice is a sublattice of the property lattice of the entity.

Theorem 3.8: The map $f: \mathscr{C} \to \mathscr{L}$ has the following properties

(i) $f(\mathsf{O}) = \mathsf{O}$, $f(\mathsf{I}) = \mathsf{I}$,
(ii) $a < b$ iff $f(a) < f(b)$,
(iii) $f(\wedge_i a_i) = \wedge_i f(a_i)$,
(iv) if p is a state such that $p < f(a)$, then there exists a classical mixture $w < a$ such that $p < f(w)$,
(v) $a \perp b$ iff $f(a) \perp f(b)$ for $a, b, a_i \in \mathscr{C}$.

Proof: If $w \perp v$, then clearly $f(w) \perp f(v)$. Suppose now that $f(w) \perp f(v)$; then $f(w) \wedge f(v) = 0$. Hence $f(w \wedge v) = 0$. As a consequence, $w \wedge v = 0$. There exits a question $\pi_i \alpha_i \in w$, where α_i are classical questions. If v is actual, then w is potential. As a consequence, there is at least one i such that α_i is true. Clearly, if w is actual, then α_i is true. This proves that $w \perp v$. If $a \perp b$, then for $w < a$ and $v < b$ we have $w \perp v$. Consider now two states p and q of the entity such that $p < f(a)$ and $q < f(b)$. Then there exists classical mixtures $w_p < a$ and $w_q < b$ such that $p < f(w_p)$ and $q < f(w_q)$. Then $f(w_p) \perp f(w_q)$ and, as a consequence, $p \perp q$. Hence $f(a) \perp f(b)$. Suppose now that $f(a) \perp f(b)$. Consider $w < a$ and $v < b$. Then $f(w) < f(a)$ and $f(v) < f(b)$. Hence $f(w) \perp f(v)$. From this follows that $w \perp v$. As a consequence, $a \perp b$. ■

If an entity S has only classical properties, we will say that S is a "*classical entity*." From Theorems 3.6 and 3.7 follows that such a classical entity can always be described by a theory as classical mechanics.

IV. THE NONCLASSICAL PART OF THE DESCRIPTION OF AN ENTITY

A. The nonclassical components of the property lattice of an entity

Now that we have studied the classical properties of an entity S, let us try to see what we can say about the nonclassical properties of S. For $w \in \Omega$ we can consider

$$\mathscr{L}_w = \{a \mid a \in \mathscr{L} \text{ and } a < f(w)\}.$$

$\mathscr{L}_w$ is the collection of all properties of S that are stronger than the classical mixture w. Let us remark that none of the properties of $\mathscr{L}_w$ except 0 and w are classical properties. Hence $\mathscr{L}_w$ is a collection of nonclassical properties. If the entity S is a classical entity, then, for each w, $\mathscr{L}_w$ is the trivial lattice consisting of 0 and $f(w)$. Let us now define

$$\underset{w}{\Sigma} = \{p \mid p \in \Sigma \text{ and } p < w\}.$$

Σ_w is the collection of all states of S that make the classical mixture w an actual property.

Theorem 4.1: If $a_i \in \mathscr{L}_w$, then $\wedge_i a_i \in \mathscr{L}_w$ and $\vee_i a_i \in \mathscr{L}_w$ and Σ_w is a full set for $\mathscr{L}_w$. The orthogonality relation on $\mathscr{L}$ defines an orthogonality relation on $\mathscr{L}_w$.

For an entity S that we describe by its classical property lattice, $\mathscr{L}_w$ describes the hidden properties of S if S is in the classical mixture w. We shall call $\mathscr{L}_w$ the "*nonclassical component*" corresponding to w.

We remarked already that, although a and $f(a)$ represent physically the same classical property, mathematically they are different objects. To enlighten the notation we will often for both objects $a \in \mathscr{C}$ and $f(a) \in \mathscr{L}$ use the notation a. This will not lead to any confusion.

Theorem 4.2: Suppose that $\mathscr{L}$ is the property lattice of the entity S and Ω is the classical state space of S and $\mathscr{L}_w$, $w \in \Omega$, are the nonclassical components of S. If $a \in \mathscr{L}$, we have

$$a = \bigvee_{w \in \Omega} (a \wedge w).$$

If $a, b \in \mathscr{L}$, we have

$$a < b \Leftrightarrow a \wedge w < b \wedge w \quad \forall\ w \in \Omega,$$

$$a \perp b \Leftrightarrow a \wedge w \perp b \wedge w \quad \forall\ w \in \Omega.$$

If $a_i \in \mathscr{L}$, we have

$$\bigvee_w \left(\bigwedge_i a_i \wedge w \right) = \bigwedge_i \left(\bigvee_w (a_i \wedge w) \right).$$

Proof: Suppose that a is actual. Then the entity is in a state $p < a$. There is, however, also a classical mixture w such that $p < w$. Hence $p < w \wedge a$. But then $p < \vee_{w \in \Omega} (a \wedge w)$. This shows that $a < \vee_{w \in \Omega} (a \wedge w)$. Since $a \wedge w < a$ for every w, we also have $\vee_{w \in \Omega} (a \wedge w) < a$. If $a < b$, then $a \wedge w < b \wedge w$. If $a \wedge w < b \wedge w$ for every w, then $\vee_w (a \wedge w) < \vee_w (b \wedge w)$; hence $a < b$. If $a \perp b$, then $a \wedge w \perp b \wedge w$. Suppose now that $a \wedge w \perp b \wedge w \ \ \forall\ w$. Take $p < a$ and $q < b$, where $p, q \in \Sigma$. Then $p < a \wedge w$ and $q < b \wedge v$ for $w, v \in \Omega$. If $w \neq v$, then $w \perp v$ and so $p \perp q$. If $w = v$, then $a \wedge w \perp b \wedge v$, and as a consequence $p \perp q$. This proves that $a \perp b$. If $a_i \in \mathscr{L}$, we have

$$\bigwedge_i \left(\bigvee_w (a_i \wedge w) \right) \text{ actual} \Leftrightarrow \forall\ i, \quad \vee_w (a_i \wedge w) \text{ actual}$$

$$\Leftrightarrow \forall\ i, \quad \exists\ w \text{ such that } a_i \wedge w \text{ is actual.}$$

Hence for i we have a w_i such that $a_i \wedge w_i$ is actual. Take $j \neq i$; then we have a w_j such that $a_j \wedge w_j$ is actual. Then $a_i \wedge w_i \wedge a_j \wedge w_j$ is actual. This shows that $w_i = w_j$; otherwise, $w_i \wedge w_j = 0$. Hence

$$\bigwedge_i \left(\bigvee_w (a_i \wedge w) \right) \text{ actual}$$

$$\Leftrightarrow \exists\ w \text{ such that } \forall\ i\ a_i \wedge w \text{ actual}$$

$$\Leftrightarrow \exists\ w \text{ such that } \bigwedge_i a_i \wedge w \text{ actual}$$

$$\Leftrightarrow \bigwedge_w \left(\bigwedge_i a_i \wedge w \right) \text{ actual.}$$ ■

B. Decomposition of the property lattice in its nonclassical components

Theorem 4.2 shows that we can replace every property a of S by its component properties $\{a \wedge w\}$. This means in a certain sense that if we know the classical property lattice of the entity and all the nonclassical components of the entity, then we know the property lattice of the entity. And this is reflected by the fact that the property lattice is the direct union of the nonclassical components. Hence the direct union of the nonclassical components. Hence the direct union that we want to consider is

$$\bigotimes_{w \in \Omega} \mathscr{L}_w.$$

As we remarked in Sec. I, this direct union is a complete lattice. It has a natural orthogonality relation

$$\bigotimes_w a_w \perp \bigotimes_w b_w \quad \text{iff} \quad a_w \perp b_w \quad \forall\ w,$$

and it has a natural set of "states" which is a full set

$$\overline{\Sigma} = \left\{ \bigotimes_{v \neq w} O_v \otimes p_w, \text{ where } p_w \in \Sigma_w \right\}.$$

Let us show that $\overline{\Sigma}$ is a full set for $\otimes_w \mathscr{L}_w$. The elements of $\overline{\Sigma}$

are of the form $\bigotimes_{v\neq w} O_v \bigotimes p_w$, where $p_w \in \Sigma_w$. We will denote such an element by $\bar{p}_w$. We shall prove that for $a \in \bigotimes_w \mathscr{L}_w$ we have

$$a = \bigvee_{p<a,\ p\in\bar{\Sigma}} p.$$

If $p \in \bar{\Sigma}$, then $p = \bar{p}_v$ for some v. Now $p < \bigotimes_w a_w$ iff $p_v < a_v$. Suppose now that $\bigotimes_w a_w < \bigotimes_w b_w$

$$\Leftrightarrow a_w < b_w \quad \forall\ w$$

$$\Leftrightarrow \{p_w < b_w \text{ for every } p_w < a_w\} \quad \forall\ w$$

$$\Leftrightarrow \{\bar{p}_w < \bigotimes_v b_v \text{ for every } \bar{p}_w < \bigotimes_v a_v\} \quad \forall\ w.$$

This shows that

$$a = \bigotimes_v a_v = \bigvee_{\bar{p}_w < \bigotimes_v a_v} \bar{p}_w = \bigvee_{p<a} p.$$

Theorem 4.3: Suppose that $\mathscr{L}$ is the property lattice of an entity S and Σ is the state space of S and Ω is the classical state space of S and $\mathscr{L}_w$, $w \in \Omega$ are the nonclassical components of S. Let us define the following map:

$$\mu: \mathscr{L} \to \bigotimes_w \mathscr{L}_w$$

$$a \to \bigotimes_w (a \wedge w);$$

then μ satisfies the following properties:

(i) $\mu(\mathsf{O}) = \mathsf{O}$ and $\mu(\mathsf{I}) = \mathsf{I}$,
(ii) $a < b$ iff $\mu(a) < \mu(b)$,
(iii) $\mu(\wedge_i a_i) = \wedge_i \mu(a_i)$,
(iv) $a \perp b$ iff $\mu(a) \perp \mu(b)$,
(v) $\mu: \Sigma \to \bar{\Sigma}$ is an isomorphism.

Proof: If $a < b$, then

$$a < b \Leftrightarrow a \wedge w < b \wedge w$$

$$\Leftrightarrow \bigotimes_w (a \wedge w) < \bigotimes_w (b \wedge w)$$

$$\Leftrightarrow \mu(a) < \mu(b),$$

$$\mu\left(\bigwedge_i a_i\right) = \bigotimes_w \left(\bigwedge_i a_i \wedge w\right)$$

$$= \bigwedge_i \left(\bigotimes_w a_i \wedge w\right)$$

$$= \bigwedge_i \mu(a_i).$$

If p is a state of S, there is only one classical mixture w such that $p < w$. From this follows that $\mu(p) = \bar{p}_w$. Take $\bar{p}_w \in \bar{\Sigma}$; then $\bar{p}_w = \bigotimes_{v\neq w} O_v \bigotimes p_w$, where $p_w \in \Sigma_w$. But then $p_w \in \Sigma$ and $\mu(p_w) = p_w$. If $p, q \in \Sigma$, then

$$p \perp q \Leftrightarrow p \wedge w \perp q \wedge w \quad \forall\ w$$

$$\Leftrightarrow \mu(p) \perp \mu(q). \qquad \blacksquare$$

So the direct union $\bigotimes_w \mathscr{L}_w$ plays the same role for a general entity as $P(\Omega)$ plays for the classical part of the entity or for a classical entity. Indeed, Theorem 4.3 reduces to Theorem 3.5 and Theorem 3.6 for a classical entity. Again, the map μ is not an isomorphism because $\bigotimes_w \mathscr{L}_w$ represents the set of statements about the entity, and some of these statements are perhaps not testable and in this case do not correspond with properties of the entity. What is important, however, is that we have an isomorphism between the state space Σ and $\bar{\Sigma}$.

The statements that are contained in $\bigotimes_w \mathscr{L}_w$ and that are not testable would in any case not lead to new states of the entity if they would have been testable (or if we would enlarge the entity by inventing questions to test these statements).

C. Pure nonclassical entities

We introduced already the concept of a classical entity. This is an entity of which every property is a classical property. The other extreme situation is the one where the property lattice of the entity has no classical properties except O and I which are always classical properties. Such an entity we will call a pure nonclassical entity. This classification agrees very well with the structure Theorem 4.3. Indeed for the property lattice of a classical entity all the nonclassical components $\mathscr{L}_w$ are trivial lattices $\{0,w\}$, and the $\bigotimes_w \mathscr{L}_w$ is isomorphic to $P(\Omega)$.

For the property lattice of a pure nonclassical entity the classical property lattice $\mathscr{C}$ is a trivial lattice $\{\mathsf{O}, \mathsf{I}\}$ and then $\bigotimes_w \mathscr{L}_w = \mathscr{L}$. In general, an entity will have both classical and nonclassical properties.

D. An entity described by quantum mechanics cannot have classical properties except O and I

One of the great shortcomings of quantum mechanics in one Hilbert space $\mathscr{H}$ is that while it is capable of describing nonclassical properties of an entity, it is uncapable of describing classical properties of an entity.

Theorem 4.4: If an entity S is described by quantum mechanics in a complex Hilbert space $\mathscr{H}$, then S has no classical properties except O and I, and so the entity is necessarily a pure nonclassical entity.

Proof: Consider an arbitrary property a of the entity in question, tested by a question α. In quantum mechanics this property is represented by the projection operator P_a on the closed subspace of the states that make a actual. The question $\alpha^{\sim}$ defines a property $b \perp a$ and hence is represented by a projection operator $\mathrm{P}_b < 1 - \mathrm{P}_a$. Consider two nonzero vectors $\mathrm{x}, \mathrm{y} \in \mathscr{H}$ such that $\mathrm{P}_a(\mathrm{x}) = \mathrm{x}$ and $\mathrm{P}_b(\mathrm{y}) = \mathrm{y}$. Then clearly $\mathrm{x} + \mathrm{y}$ is no eigenvector of P_a and no eigenvector of P_b. Hence, if the entity is in the state represented by $\mathrm{x} + \mathrm{y}$, then neither α nor $\alpha^{\sim}$ is true. This shows that α is not a classical question. ■

If it is clear that quantum mechanics properly used cannot describe nontrivial classical questions, it is not amazing at all that every time that one tries to describe a physical system that manifestly has classical properties (e.g., the part of the measuring apparatus that we use during a measurement) with the formalism of quantum mechanics, one encounters the greatest difficulties. It is also clear then that quantum mechanics is not a more general theory than classical mechanics. Both of them are special cases of a more general theory, and this explains why it is so hard to draw them together.

E. Example that shows that not every entity is a classical or a pure nonclassical entity

If we consider an entity S that exists in space and time, then we can in principle construct the following experiments

for the entity. We put the entity between two oppositely charged parallel plates. If the entity feels a force in the direction of the positively charged plate, we say that it has a negative charge. If it feels a force in the other direction, we shall say that it has a positive charge. If it does not feel a force at all, it is uncharged.

We define the question α that consists of performing the experiment giving the answer "yes" if we find a negative charge. Otherwise, we give the answer "no." Experimentally, one verifies that, for all the entities in nature for which it makes sense to define the question α, the question α is a classical question. Indeed for every entity its charge is negative or nonnegative. There does not exist a state of the entity such that, in measuring the charge, the entity would sometimes have a negative charge and sometimes no negative charge. This example shows that every entity as elementary as one wants for which it makes sense to define the question α has at least one classical property.

V. THE CLASSICAL PART AND THE NONCLASSICAL PART AND THE AXIOMS

The study of the classical part of the description of an entity and of the nonclassical part of the description of the entity is done till now without any axioms to be satisfied in the formalism. In Refs. 3 and 4 we propose some axioms that reduce the formalism in such a way that the nonclassical components of the entity are irreducible complete orthocomplemented weakly modular lattices that satisfy the covering law. By using Piron's representation theorem (see Ref. 2), we have that energy nonclassical component $\mathscr{L}_w$ becomes isomorphic to the lattice of closed subspaces $P(\mathscr{H}_w)$ of a generalized Hilbert space $\mathscr{H}_w$. For the property lattice we find again the structure explained in Sec. I of a direct union $\bigotimes_w P(\mathscr{H}_w)$ of Hilbert space lattices. As we remarked already, only Axioms 1, 2, and 3 do not lead to contradictions in the case of an entity consisting of two separated entities. Axioms 4 and 5 are wrong axioms that make it impossible for the theory to describe separated entities (see Refs. 3–5). Axiom 4 is the axiom that leads to the weak modularity of the nonclassical components. Hence, if we drop this axiom we cannot proceed as explained in Sec. I to find the direct union, nor can we apply Piron's representation theorem to find a Hilbert space representation for the nonclassical components. To define axiom 1 as proposed in Secs. III and IV, we introduced the concept of primitive questions. Let us recall the definition of a primitive question.

Definition 5.1: If α is a question testing the property a such that $\alpha^{\sim}$ tests the property b, then α is a primitive question iff whenever the entity is in a state orthogonal to a, then $\alpha^{\sim}$ is true, and, whenever the entity is in a state orthogonal to b, then α is true.

Let us recall Axioms 1 and 2.

Axiom 1: If S is an entity, then the primitive questions of S form a generating set of questions for the property lattice.

Axiom 2: If S is an entity and p is a state of S, then the statement "the entity S is in a state orthogonal to p" is a property of S.

Axioms 1 and 2 have a consequence that the property lattice $\mathscr{L}$ becomes an orthocomplemented lattice. If $a\in\mathscr{L}$, the interpretation of the orthocomplement is the following:

a' is actual iff the entity is in a state $p\perp a$.

Let us remark that the classical property lattice $\mathscr{C}$ always satisfies Axiom 1, since every classical question is evidently a primitive question. It is now interesting to remark that if Axioms 1 and 2 are satisfied for $\mathscr{L}$ and C, the map

$$\mu:\mathscr{L}\to\bigotimes_{w\in\Omega}\mathscr{L}_w$$

becomes an isomorphism.

Theorem 5.2: Suppose that we have an entity S with a property lattice $\mathscr{L}$ that satisfies Axioms 1 and 2 and a classical property lattice $\mathscr{C}$ that satisfies Axiom 2; then for $a\in\mathscr{C}$ we have $f(a')=f(a)'$ and $f(a)'$ is actual iff $f(a)$ is potential. We will denote in the following $f(a)'$ by a'. For $a_i\in\mathscr{C}$ we have $f(\bigotimes_i a_i)=\bigvee_i f(a_i)$ and $\bigvee_i f(a_i)$ is actual iff there is at least one i such that $f(a_i)$ is actual. We will denote in the following $f(\bigotimes_i a_i)$ by $\bigvee_i a_i$.

Proof: If $\mathscr{C}$ satisfies Axiom 2, we have the following: If $w\in\mathscr{C}$, there exists a question $\pi_i\alpha_i$, where α_i are classical questions such that $\pi_i\alpha_i$ is true iff the entity is in a classical mixture v different from w. So $\pi_i\alpha_i$ is true iff the entity is in a state q orthogonal to $f(w)$. This shows that $\pi_i\alpha_i\in f(w)'$. As a consequence, $f(w')=f(w)'$. If $a\in\mathscr{C}$, then $a'=\bigwedge_{w<a}w'$; hence

$$f(a')=\bigwedge_{w<a}f(w')=\Big(\bigvee_{w<a}f(w)\Big)'=f(a)'.$$

If $a_i\in\mathscr{C}$, then

$$f\Big(\bigvee_i a_i\Big)=f\Big(\bigwedge_i a_i'\Big)'=\Big[\bigwedge_i f(a_i')\Big]'=\bigvee_i f(a_i).$$

If $\bigvee_i f(a_i)$ is actual, then $(\bigvee_i f(a_i))'$ is potential. So $\bigwedge_i f(a_i)'$ is potential. But then there is at least one i such that $f(a_i)'$ is potential. For this follows that $f(a_i)$ is actual. ■

Theorem 5.3: If $\mathscr{L}$ is the property lattice of an entity S satisfying Axioms 1 and 2 and $\mathscr{C}$ is the classical property lattice of an entity satisfying Axiom 2 then the classical properties of S satisfy the following properties:

(i) If $b\in\mathscr{L}$ and $a\in\mathscr{C}$, then

$$b=(b\wedge a)\vee(b\wedge a')$$
$$=(b\vee a)\wedge(b\vee a').$$

(ii) If $b_i\in\mathscr{L}$ and $a\in\mathscr{C}$, then

$$a\wedge(\bigvee_i b_i)=\bigvee_i(a\wedge b_i).$$

Proof: (i) We have clearly $(b\wedge a)\vee(b\wedge a')<b$. Suppose now that b is actual. Since $a\in\mathscr{C}$, we know that a is actual or a' is actual. This shows that $b\wedge a$ is actual or $b\wedge a'$ is actual. In both cases $(b\wedge a)\vee(b\wedge a')$ is actual. As a consequence,

$$b<(b\wedge a)\vee(b\wedge a').$$

We have also $b'=(b'\wedge a)\vee(b'\wedge a')$. Hence $b=(b\vee a')\wedge(b\vee a)$.

(ii) $$a\wedge(\bigvee_i b_i)=a\wedge(\bigvee_i[(b_i\wedge a)\vee_i(b_i\wedge a')])$$
$$=a\wedge(\bigvee_i(b_i\wedge a)\vee_i(b_i\wedge a')).$$

Let us put $b=\bigvee_i(b_i\wedge a)$ and $c=\bigvee_i(b_i\wedge a')$, then $b<a$ and $c<a'$:

$$a \wedge (\vee_i b_i) = a \wedge (b \vee c)$$
$$= a \wedge (b \vee c \vee a) \wedge (b \vee c \vee a')$$
$$= a \wedge (c \vee a) \wedge (b \vee a')$$
$$= a \wedge (b \vee a')$$
$$= (b \vee a) \wedge (b \vee a') = b = \bigvee_i (b_i \wedge a). \quad \blacksquare$$

Theorem 5.2 shows that the orthocomplementation introduced by Axioms 1 and 2 in $\mathscr{L}$ and the orthocomplementation introduced by Axiom 2 in $\mathscr{C}$ are the same. Theorem 5.3 shows that the classical properties satisfy compatibility relations. Every nonclassical component is also an orthocomplemented lattice.

Theorem 5.4: If $\mathscr{L}$ is the property lattice of an entity S satisfying Axioms 1 and 2 and $\mathscr{C}$ is the classical property lattice of S satisfying Axiom 2. We define for $a \in \mathscr{L}_w$

$$a^w = a' \wedge w,$$

Then the map that makes correspond with every $a \in \mathscr{L}_w$ the property a^w is an orthocomplementation of $\mathscr{L}_w$, and a^w is actual iff the entity is in a state orthogonal to a such that w is actual.

Proof: If $a, b \in \mathscr{L}$ and $a < b$, then $b' < a'$. So $b' \wedge w < a' \wedge w$ or $b^w < a^w$. If $a \in \mathscr{L}_w$, then $(a^w)^w = (a' \wedge w)' \wedge w = (a \vee w') \wedge w = a \wedge w = a$. Clearly, $a^w \wedge a = a' \wedge w \wedge a = 0$. $\blacksquare$

If $\mathscr{L}_i$ is a family of complete lattices, we gave a construction of the direct union $\bigotimes_i \mathscr{L}_i$ of the lattices $\mathscr{L}_i$ in 1.3. If $\mathscr{L}_i$ are orthocomplemented lattices and we define for an arbitrary element $\bigotimes_i a_i$ of the direct union

$$\left(\bigotimes_i a_i\right)' = \bigotimes_i a_i',$$

then $': \bigotimes_i \mathscr{L}_i \to \bigotimes_i \mathscr{L}_i$ is an orthocomplementation.

Theorem 5.5: Suppose that $\mathscr{L}$ is the property lattice of an entity S and Ω is the classical state space of S, $\mathscr{L}_w$ for $w \in \Omega$ are the nonclassical components, and $\mathscr{C}$ is the classical property lattice. Suppose that $\mathscr{L}$ satisfies Axioms 1 and 2 and $\mathscr{C}$ satisfies Axiom 2; then

$$\mu: \mathscr{L} \to \bigotimes_w \mathscr{L}_w$$
$$a \to \bigotimes_w (a \wedge w)$$

is an isomorphism.

Proof: Take $b \in \bigotimes_w \mathscr{L}_w$, then $b = \bigotimes_w b_w$, where $b_w \in \mathscr{L}_w$. Consider the property $c = \vee_w b_w$ of $\mathscr{L}$. Then

$$c \wedge w_1 = \left(\bigvee_w b_w\right) \wedge w_1 = \bigvee_w (b_w \wedge w_1) = b_{w_1}.$$

This shows that

$$\mu(c) = \bigotimes_w [c \wedge w] = \bigotimes_w b_w = b.$$

Hence μ is a surjective map. From Theorem 4.3 it follows that μ is an isomorphism. $\blacksquare$

This theorem shows that when Axioms 1 and 2 are satisfied, the property lattice of an entity gets the very nice structure of the direct union of its nonclassical components.

VI. CONCLUSION

Theorems 4.3 and 5.5 show that we can indeed, for every entity, study its classical properties apart by means of a theory as classical mechanics. The changing of actual classical properties in potential and potential classical properties in actual is described by the changing of the classical mixture of the entity, which is in a certain sense the classical state of the entity.

If we want to be able to describe also nonclassical properties of the entity, a theory as classical mechanics does not work anymore for the description of these properties.

Quantum mechanics is a theory that describes nonclassical properties. It cannot, however, describe classical properties. This shows that both classical mechanics and quantum mechanics are special cases of the theory that can describe an arbitrary entity having both classical and nonclassical properties and clarifies in a certain sense the very old question: How many atoms do we have to put together to have a macroscopical entity that has to be described by classical mechanics? Indeed, from our analysis it follows that the degree of classicality of an entity is not defined by the number of atoms that it contains but by the nature of the properties that we take to characterize the entity.

[1] C. Piron, Helv. Phys. Acta **42**, 330 (1969).
[2] C. Piron, *Foundations of Quantum Physics* (Addison-Wesley, Reading, Mass., 1976).
[3] Dirk Aerts, "Description of many physical entities without the paradoxes encountered in quantum mechanics," Found. Phys. **12** (12), 1131 (1982).
[4] Dirk Aerts, "Description of many physical entities without the paradoxes encountered in quantum mechanics," Found. Phys. **12** (12), 1131 (1982).
[5] Dirk Aerts, "The missing elements of reality in the description of quantum mechanics of the E.P.R. paradox situation," Preprint VUB/TF/81/05, Vrije Universiteit Brussel, 1981.
[6] W. Pauli, "Die allgemeinen Prinzipien der Wellenmechanik," *Handbuch der Physik* (Springer-Verlag, Berlin, 1933), 2nd ed., Vol. 24.
[7] A. Einstein, B. Podolsky, and M. Rosen, Phys. Rev. **47**, 777–80 (1935).
[8] C. Piron, Ann. Fond. Louis de Broglie **3**, 131 (1978).

Value preserving quantum measurements: Impossibility theorems and lower bounds for the distortion[a)]

G. C. Ghirardi
International Centre for Theoretical Physics, Trieste, Italy
Istituto di Fisica Teorica dell'Università, Trieste, Italy

A. Rimini
Istituto di Fisica dell'Università, Salerno, Italy

T. Weber
Istituto di Fisica Teorica dell'Università, Trieste, Italy

(Received 30 November 1982; accepted for publication 6 May 1983)

Extending previous works on the subject we consider the problem of the limitations to ideal quantum measurements arising from the presence of additive conservation laws and we discuss impossibility theorems and derive lower bounds for the deviations from the ideal schemes, with particular reference to the distorting case.

PACS numbers: 03.65.Bz

I. INTRODUCTION

As is well known, the existence of additive quantities which are conserved during the system-apparatus interaction in a measurement process leads to limitations on the possibility of an ideal quantum-mechanical measurement.[1–4] In this connection, two kinds of problems have been discussed. The first one concerns the proof of impossibility theorems for some given measurement schemes.[5] The second deals with the modifications which have then to be introduced in such schemes in order to make them compatible with the existence of the additive conservation laws. In particular, it becomes important to evaluate how small the modifications can be kept. This is usually obtained by deriving lower bounds for proper combinations of the norms of the unwanted states in the evolution equation.[6–8]

The ideal measurement schemes which are usually considered are value preserving in the sense that the state of the system after the measurement belongs to the same linear eigenmanifolds of the measured observable to which it belonged before the measurement. Within the framework of the value-preserving schemes, when the eigenmanifolds of the measured quantity are degenerate, one has further to distinguish between nondistorting (or state preserving) and distorting schemes.

As far as the impossibility proofs for such schemes are concerned, in the literature,[3] there exists a very general theorem for the state-preserving case and a less general theorem for the distorting case.

Regarding the derivation of lower bounds for the terms expressing the malfunctioning of the apparatus, there is a quite general treatment[8] which, however, deals with this problem only for the case of a state-preserving scheme. In this paper, after a concise summary of the present situation, we point out (Sec. II) that the impossibility theorem for distorting measurements derived in Ref. 3 holds under less restrictive conditions. The main part of the paper (Sec. III) is devoted to the derivation of lower bounds for the malfunctioning terms in the case of a modification of a value-preserving distorting measurement process. The relevance of this treatment resides in the fact that it is the first time that lower bounds are derived for the distorting case, and in the fact that, to get the result, we make use of less restrictive assumptions than those previously used in the derivation of bounds for the state-preserving case. Some concluding remarks are made in Sec. IV.

II. INTRODUCTORY CONSIDERATIONS

In this paper, we will deal with the problem of the measurement of an observable $\mathscr{M}$ of a quantum system S (associated with a self-adjoint operator M of the system Hilbert space $\mathscr{H}_S$) in the presence of an additive conservation law. Let us state, first of all, a general assumption which we shall suppose to hold throughout the paper.

Assumption 1: Let the self-adjoint operator M have a purely discrete spectrum and E_i be the eigenspaces of M and $\mathscr{E}_i$ the associated projection operators. The system interacts with an apparatus, the interaction being described by the unitary evolution operator U acting in the direct product of the system Hilbert space $\mathscr{H}_S$ and of the apparatus Hilbert space $\mathscr{H}_A$.[9] An additive conservation law for the system-apparatus interaction holds, i.e., there exist two self-adjoint operators N_S and N_A of $\mathscr{H}_S$ and $\mathscr{H}_A$, respectively, such that

$$[U, e^{i(N_S + N_A)r}] = 0, \quad \forall\, r. \tag{2.1}$$

As is well known, when the operators N_S and /or N_A are unbounded, Eq. (2.1) is the mathematically proper way of expressing the conservation law for $N_S + N_A$.

Let us now define the various possible ideal measurement schemes which can be considered in the general case in which M has a degenerate spectrum.

A. Value-preserving nondistorting ideal measurement scheme (VPND)

We say that the system-apparatus interaction corresponds to a VPND measurement of $\mathscr{M}$ if there exists a state

[a)] Work supported in part by the Istituto Nazionale di Fisica Nucleare, Sezioni di Trieste and Pavia.

 0022-2488/83/102454-07$02.50

$\phi_0 \in \mathscr{H}_A$ such that for any $\psi^{(i)} \in E_i$,

$$U\psi^{(i)}\phi_0 = \psi^{(i)}\phi_i, \tag{2.2}$$

with $(\phi_i, \phi_j) = 0$ for $i \neq j$. The linearity of U implies that ϕ_i be the same for all $\psi^{(i)} \in E_i$.

B. Value-preserving distorting ideal measurement scheme (VPD)

We say that we have a VPD measurement process if there exists a state $\phi_0 \in \mathscr{H}_A$, such that for any $\psi^{(i)} \in E_i$,

$$U\psi^{(i)}\phi_0 \in E_i \otimes F_i, \tag{2.3}$$

where F_i are mutually orthogonal closed linear manifolds of $\mathscr{H}_A$, whose associated projection operators will be denoted by $\mathscr{F}_i$. The interest of the scheme (2.3) lies in the fact that it is the most general scheme satisfying the physical requirement that, when the measurement is immediately repeated, the same result is obtained with certainty.

Relation (2.1) forbids, under proper assumptions about the operator N_S, both the VPND and VPD schemes. This situation raises two problems: the first is that of getting impossibility proofs for the ideal measurement schemes holding under the least restrictive assumptions on N_S. Once the impossibility has been proved, it becomes natural to enlarge the measurement scheme allowing the appearance of nonideal terms, and to investigate how small these unwanted terms can be made.

Let us now make precise the modifications which have to be considered. We distinguish again between VPND and VPD schemes.

The basic feature of the VPND scheme is the fact that it is not only value preserving, but also state preserving. To distinguish among different states belonging to the same manifold E_i, we use a further label α. The VPND scheme is expressed then by the evolution equation

$$U\psi_{i\alpha}\phi_0 = \psi_{i\alpha}\phi_i, \tag{2.4}$$

with $(\phi_i, \phi_j) = 0$ for $i \neq j$.

Let us denote by $\mathscr{E}_{i\alpha}$ the projection operator on the state $\psi_{i\alpha}$ and by $\mathscr{f}_i$ the projection operator on the state ϕ_i. To modify scheme (2.4), we consider the completely general evolution

$$\begin{aligned} U\psi_{i\alpha}\phi_0 = {} & \mathscr{E}_{i\alpha}\mathscr{f}_i U\psi_{i\alpha}\phi_0 \\ & + \mathscr{E}_{i\alpha}(\mathbb{I}_A - \mathscr{f}_i)U\psi_{i\alpha}\phi_0 \\ & + (\mathbb{I}_S - \mathscr{E}_{i\alpha})U\psi_{i\alpha}\phi_0. \end{aligned} \tag{2.5}$$

As discussed in Ref. 7, the second term on the rhs of (2.5) represents errors and/or ambiguities in the result of the measurement, while the third term represents the distortion of the state of the system (also associated with possible errors and/or ambiguities). Obviously, in order to have a good apparatus in the VPND spirit, one wants to make as small as possible the norms of such terms.

In the case of the VPD scheme, the evolution equation (2.3) can be rewritten as

$$U\psi^{(i)}\phi_0 = \mathscr{E}_i \mathscr{F}_i U\psi^{(i)}\phi_0, \quad \forall \quad \psi^{(i)} \in E_i. \tag{2.6}$$

We recall that $\mathscr{E}_i$ and $\mathscr{F}_i$ are the projection operators on E_i and F_i, respectively.

To modify this scheme we write, in place of (2.6), the completely general evolution equation

$$\begin{aligned} U\psi^{(i)}\phi_0 = {} & \mathscr{E}_i \mathscr{F}_i U\psi^{(i)}\phi_0 \\ & + \mathscr{E}_i(\mathbb{I}_A - \mathscr{F}_i)U\psi^{(i)}\phi_0 + (\mathbb{I}_S - \mathscr{E}_i)U\psi^{(i)}\phi_0. \end{aligned} \tag{2.7}$$

Again the second term on the rhs represents errors and/or ambiguities while the third term represents the (value-nonpreserving) distortion of the state of the system. To have a good VPD apparatus, we have to assume that all the states at the rhs of (2.7), except the first one, have small norms.

Let us now discuss the situation concerning the impossibility proofs and the derivation of bounds for the unwanted terms.

C. Impossibility proofs

As far as the VPND scheme is concerned, Stein and Shimony[3] have proved that the assumptions listed under Assumption 1 are incompatible with (2.2) unless

$$e^{iN_S r}\psi^{(i)} \in \mathscr{E}_i, \quad \forall \quad r \in \mathbb{R}, \quad \forall \quad \psi^{(i)} \in E_i. \tag{2.8}$$

When N_S is unbounded, this is the proper mathematical way to express the fact that N_S and M "commute."

For the VPD scheme, one can prove that Assumption 1, plus a further requirement, Assumption 2, that we give below, are incompatible with Eq. (2.3), unless, again, Eq. (2.4) holds. Assumption 2 concerns the operator N_S and is formulated as follows.

Assumption 2: All the closed linear manifolds E_i are contained in the domain $\mathscr{D}_{N_S}$ of the self-adjoint operator N_S.

Note (see the Appendix, Theorem 2) that this implies that N_S is bounded on each E_i, in the sense that

$$\|N_S\psi^{(i)}\| \leqslant B_i\|\psi^{(i)}\|, \quad \forall \quad \psi^{(i)} \in E_i. \tag{2.9}$$

Obviously, this does not mean that N_S is a bounded operator on $\mathscr{H}_S$.

To prove that under Assumptions 1 and 2, the evolution Equation (2.3) [or equivalently (2.6)] cannot hold unless, for any i,

$$N_S\psi^{(i)} \in E_i, \quad \forall \quad \psi^{(i)} \in E_i, \tag{2.10}$$

we proceed as follows.

Let us consider the matrix element $(\psi^{(i)}\phi_0, e^{i(N_S + N_A)r}\psi^{(i)}\phi_0)$, with $\psi^{(l)} \in E_l$, $\|\psi^{(l)}\| = \|\phi_0\| = 1$, $(l = ij; i \neq j)$. Due to Eq. (2.1), we have

$$\begin{aligned} & (\psi^{(j)}, e^{iN_S r}\psi^{(i)})(\phi_0, e^{iN_A r}\phi_0) \\ & \quad = (U\psi^{(j)}\phi_0, e^{i(N_S + N_A)r}U\psi^{(i)}\phi_0) \\ & \quad = (\mathscr{E}_j \mathscr{F}_j U\psi^{(j)}\phi_0, e^{iN_S r}e^{iN_A r}\mathscr{E}_i \mathscr{F}_i U\psi^{(i)}\phi_0). \end{aligned} \tag{2.11}$$

For $r \neq 0$, let us define the operator $g(r)$ of $\mathscr{H}_S$ through the relation

$$e^{iN_S r} = \mathbb{I}_S + irg(r), \tag{2.12}$$

and observe that $g(r)$ is everywhere defined in $\mathscr{H}_S$ and bounded. Due to the fact that $(\psi^{(j)}, \psi^{(i)}) = 0$ and $\mathscr{E}_j \mathscr{E}_i = 0$,

we get, from (2.11),

$$(\psi^{(j)},g(r)\psi^{(i)})(\phi_0,e^{iN_A r}\phi_0)$$
$$= (U\psi^{(j)}\phi_0,\mathscr{E}_j g(r)\mathscr{E}_i\mathscr{F}_j e^{iN_A r}\mathscr{F}_i U\psi^{(i)}\phi_0). \tag{2.13}$$

There follows, using the Schwarz inequality,

$$|(\psi^{(j)},g(r)\psi^{(i)})|\ |(\phi_0,e^{iN_A r}\phi_0)|$$
$$\leqslant \|\mathscr{E}_j g(r)\mathscr{E}_i\mathscr{F}_j e^{iN_A r}\mathscr{F}_i U\psi^{(i)}\phi_0\|. \tag{2.14}$$

According to Eq. (A6) of the Appendix, the operator $g(r)\mathscr{E}_i$ is bounded; moreover, its bound $\|g(r)\mathscr{E}_i\|$ is, for any $r\neq 0$, smaller than the bound B_i [defined in Eq. (2.9)] of the restriction of N_S to the closed linear manifold E_i. Therefore we have

$$|(\psi^{(j)},g(r)\psi^{(i)})|\ |(\phi_0,e^{iN_A r}\phi_0)|$$
$$\leqslant B_i\|\mathscr{F}_j e^{iN_A r}\mathscr{F}_i U\psi^{(i)}\phi_0\|. \tag{2.15}$$

Since

$$e^{iN_A r}\xrightarrow[r\to 0]{S}\mathbb{I}_A$$

and

$$g(r)\psi^{(i)}\xrightarrow[r\to 0]{S}N_S\psi^{(i)},$$

recalling that $\mathscr{F}_j\mathscr{F}_i = 0$ and taking the limit of (2.15) for $r\to 0$, we get

$$(\psi^{(j)},N_S\psi^{(i)}) = 0. \tag{2.16}$$

Since (2.16) holds for any i and j, Eq. (2.10) follows for any i.

This theorem constitutes a slight generalization of previous impossibility theorems given in Refs. 2 and 3; its proof follows essentially the same lines used in those papers. The theorem is slightly more general since it makes use of Assumption 2 in place of the more restrictive assumption used previously.

A version of this theorem which is weaker both in the assumptions and in the results can also be given and turns out to be useful (see Sec. IV). Precisely, let us replace Assumption 2 by a similar Assumption 2i concerning only the manifold E_i. By inspecting the proof of the previous theorem, one sees that the conclusions remain valid, the only difference being that now Eq. (2.10) follows for only that i to which Assumption 2i refers. In this form, the theorem can easily be extended also to the case in which a part of the spectrum of the measured quantity M is continuous. In such a case, we assume that the time evolution is of the VPD type (2.6) for any i in the discrete spectrum of M and that

$$U\psi'\phi_0\in\mathscr{F}'U\psi'\phi_0,\quad \forall\quad \psi'\in E', \tag{2.17}$$

where E' is the closed linear manifold in $\mathscr{H}_S$ corresponding to the continuous part of the spectrum of M, and $\mathscr{F}'$ is the projection operator on a closed linear manifold F' in $\mathscr{H}_A$ orthogonal to all manifolds F_i. The property (2.17) means that the apparatus can recognize the states of the continuum. Then, under Assumption 1 (where a purely discrete spectrum is replaced by a partly discrete spectrum), Eq. (2.10) follows for any i in the discrete spectrum for which Assumption 2i is verified.

D. Bounds

When an ideal measurement is forbidden, one is compelled to resort to modified schemes which allow a certain amount of nonideality.

As far as the derivation of bounds for the norms of the terms representing malfunctioning is concerned, up to now the most general result which has been obtained refers only to the case of no degeneracy or to the scheme (2.5) and makes use of the assumptions that N_S is bounded and ϕ_0 belongs to the domain of N_A. In fact, in Ref. 8 we have derived a family of inequalities relating the norms of the distorting terms to the inverse of the mean value of the square of N_A on the state ϕ_0. We now consider a generalization of the previous results yielding a lower bound for the norms of the unwanted terms of (2.5) and (2.7) under less restrictive assumptions about N_S.

III. LOWER BOUNDS FOR DISTORTING MEASUREMENTS

We now want to obtain lower bounds for the two last unwanted terms on the rhs of Eq. (2.7).

We again suppose that Assumptions 1 and 2 hold and, moreover, the following.

Assumption 3: The initial state ϕ_0 of the apparatus belongs to the domain of the operator N_A.

Let us denote by ρ_l and η_l the norms of the unwanted terms,

$$\rho_l = \|\mathscr{E}_l(\mathbb{I}_A - \mathscr{F}_l)U\psi^{(l)}\phi_0\|,$$
$$\eta_l = \|(\mathbb{I}_S - \mathscr{E}_l)U\psi^{(l)}\phi_0\|, \tag{3.1}$$

and by Q_l the projection operator on the manifold orthogonal to E_l,

$$Q_l = \mathbb{I}_S - \mathscr{E}_l. \tag{3.2}$$

Then we can write the identity

$$e^{i(N_S+N_A)r} = Q_j e^{i(N_S+N_A)r} + e^{i(N_S+N_A)r}e^{-iN_S r}\mathscr{E}_j e^{iN_S r}. \tag{3.3}$$

Using Eq. (2.1), we have

$$e^{i(N_S+N_A)r} = U^+Q_jUe^{i(N_S+N_A)r} + e^{i(N_S+N_A)r}U^+e^{-iN_S r}\mathscr{E}_j e^{iN_S r}U. \tag{3.4}$$

Recalling definition (2.12), we put in an analogous way,

$$e^{iN_A r} = \mathbb{I}_A + irG(r). \tag{3.5}$$

Then, sandwiching Eq. (3.4) between the normalized states $\psi^{(i)}\phi_0$ and $\psi^{(i)}\phi_0$, $i\neq j$, we get

$$(\psi^{(j)},g(r)\psi^{(i)}) + ir(\psi^{(j)},g(r)\psi^{(i)})(\phi_0,G(r)\phi_0)$$
$$= (Ug^+(r)\psi^{(j)}\phi_0,\mathscr{E}_j U\psi^{(i)}\phi_0) + (UG^+(r)\psi^{(j)}\phi_0,\mathscr{E}_j U\psi^{(i)}\phi_0) - (U\psi^{(j)}\phi_0,g^+(r)\mathscr{E}_j U\psi^{(i)}\phi_0)$$
$$+ (g^+(r)\mathscr{E}_j U\psi^{(j)}\phi_0,U\psi^{(i)}\phi_0) + (Q_j U\psi^{(j)}\phi_0,Ug(r)\psi^{(i)}\phi_0) + (Q_j U\psi^{(j)}\phi_0,UG(r)\psi^{(i)}\phi_0) + r\Gamma(r), \tag{3.6}$$

where we have denoted by $r\Gamma(r)$ all the other terms arising from the rhs of (3.4). From the explicit expression for $\Gamma(r)$, we have immediately

$$
\begin{aligned}
|\Gamma(r)| \leqslant & |(Ug^+(r)G^+(r)\psi^{(j)}\phi_0, \mathscr{E}_j U\psi^{(i)}\phi_0)| \\
& + |(Ug^+(r)\psi^{(j)}\phi_0, g^+(r)\mathscr{E}_j U\psi^{(i)}\phi_0)| + |(Ug^+(r)\psi^{(j)}\phi_0, \mathscr{E}_j g(r)U\psi^{(i)}\phi_0)| \\
& + |(UG^+(r)\psi^{(j)}\phi_0, g^+(r)\mathscr{E}_j U\psi^{(i)}\phi_0)| + |(UG^+(r)\psi^{(j)}\phi_0, \mathscr{E}_j g(r)U\psi^{(i)}\phi_0)| \\
& + |(\mathscr{E}_j g(r)U\psi^{(j)}\phi_0, \mathscr{E}_j g(r)U\psi^{(i)}\phi_0)| + |(Q_j U\psi^{(j)}\phi_0, Ug(r)G(r)\psi^{(i)}\phi_0)| \\
& + |r|\{|(Ug^+(r)G^+(r)\psi^{(j)}\phi_0, g^+(r)\mathscr{E}_j U\psi^{(i)}\phi_0)| + |(Ug^+(r)G^+(r)\psi^{(j)}\phi_0, \mathscr{E}_j g(r)U\psi^{(i)}\phi_0)| \\
& + |(Ug^+(r)\psi^{(j)}\phi_0, (g^+(r)\mathscr{E}_j)(\mathscr{E}_j g(r))U\psi^{(i)}\phi_0)| + |(UG^+(r)\psi^{(j)}\phi_0, (g^+(r)\mathscr{E}_j)(\mathscr{E}_j g(r))U\psi^{(i)}\phi_0)|\} \\
& + r^2|(Ug^+(r)G^+(r)\psi^{(j)}\phi_0, (g^+(r)\mathscr{E}_j)(\mathscr{E}_j g(r))U\psi^{(i)}\phi_0)|. \qquad (3.7)
\end{aligned}
$$

Using Theorm 3 and the last remarks made in the Appendix, we have that the inequalities

$$
\|g^+(r)\mathscr{E}_j\Psi\| \leqslant B_j\|\Psi\|, \quad \|\mathscr{E}_j g(r)\Psi\| \leqslant B_j\|\Psi\| \qquad (3.8)
$$

hold for any $\Psi \in \mathscr{H}_S \otimes \mathscr{H}_A$ and any $r \neq 0$. We then get, from (3.7), using the Schwarz inequality,

$$
\begin{aligned}
|\Gamma(r)| \leqslant & \|g^+(r)G^+(r)\psi^{(j)}\phi_0\| + 2B_j\|g^+(r)\psi^{(j)}\| + 2B_j\|G^+(r)\phi_0\| + B_j^2 + \|g(r)\psi^{(i)}\|\,\|G(r)\phi_0\| \\
& + r\{2B_j\|g^+(r)\psi^{(j)}\|\,\|G^+(r)\phi_0\| + B_j^2\|g^+(r)\psi^{(j)}\| + B_j^2\|G^+(r)\phi_0\|\} + r^2B_j^2\|g^+(r)\psi^{(j)}\|\,\|G^+(r)\phi_0\|. \qquad (3.9)
\end{aligned}
$$

We now observe that under our assumptions, all the norms appearing on the rhs of (3.9) have finite limits for $r \to 0$. Since B_j is independent of r, it follows that

$$
\lim_{r\to 0} r\Gamma(r) = 0. \qquad (3.10)
$$

We can then take the limit for $r \to 0$ of Eq. (3.6), getting

$$
\begin{aligned}
(\psi^{(j)}, N_S\psi^{(i)}) = & (UN_S\psi^{(j)}\phi_0, \mathscr{E}_j U\psi^{(i)}\phi_0) \\
& + (UN_A\psi^{(j)}\phi_0, \mathscr{E}_j U\psi^{(i)}\phi_0) - (U\psi^{(j)}\phi_0, N_S\mathscr{E}_j U\psi^{(i)}\phi_0) + (N_S\mathscr{E}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0) \\
& + (Q_j U\psi^{(j)}\phi_0, UN_S\psi^{(i)}\phi_0) + (Q_j U\psi^{(j)}\phi_0, UN_A\psi^{(i)}\phi_0). \qquad (3.11)
\end{aligned}
$$

From this equation, using the Schwarz inequality, we have

$$
\begin{aligned}
|(\psi^{(j)}, N_S\psi^{(i)}) \leqslant & 2B_j\|\mathscr{E}_j U\psi^{(i)}\phi_0\| + B_i\|Q_j U\psi^{(j)}\phi_0\| \\
& + \|N_A\phi_0\|\{\|\mathscr{E}_j U\psi^{(i)}\phi_0\| + \|Q_j U\psi^{(j)}\phi_0\|\} + |(N_S\mathscr{E}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0)|. \qquad (3.12)
\end{aligned}
$$

We note that

$$
\|\mathscr{E}_j U\psi^{(i)}\phi_0\| \leqslant \|(\mathbb{I}_s - \mathscr{E}_i)U\psi^{(i)}\phi_0\| = \eta_i, \quad \|Q_j U\psi^{(j)}\phi_0\| = \|(\mathbb{I}_S - \mathscr{E}_j)U\psi^{(j)}\phi_0\| = \eta_j. \qquad (3.13)
$$

Furthermore, we have

$$
\begin{aligned}
& |(N_S\mathscr{E}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0)| \\
& \quad \equiv |(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0) + (N_S\mathscr{E}_j(\mathbb{I}_A - \mathscr{F}_j)U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0)| \\
& \quad \leqslant |(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0)| + \|N_S\mathscr{E}_j(\mathbb{I}_A - \mathscr{F}_j)U\psi^{(j)}\phi_0\|, \qquad (3.14)
\end{aligned}
$$

and

$$
\|N_S\mathscr{E}_j(\mathbb{I}_A - \mathscr{F}_j)U\psi^{(j)}\phi_0\| \leqslant B_j\|\mathscr{E}_j(\mathbb{I}_A - \mathscr{F}_j)U\psi^{(j)}\phi_0\| = B_j\rho_j, \qquad (3.15)
$$

and

$$
\begin{aligned}
|(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0)| \leqslant & |(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, \mathscr{E}_i\mathscr{F}_i U\psi^{(i)}\phi_0)| \\
& + |(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, \mathscr{E}_i(\mathbb{I}_A - \mathscr{F}_i)U\psi^{(i)}\phi_0)| + |(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, (\mathbb{I}_S - \mathscr{E}_i)U\psi^{(i)}\phi_0)|. \qquad (3.16)
\end{aligned}
$$

The first term on the rhs of (3.16) vanishes since $\mathscr{F}_j\mathscr{F}_i = 0$; again using the Schwarz inequality , the bound (2.9) and Definitions (3.1), we get, from (3.16),

$$
|(N_S\mathscr{E}_j\mathscr{F}_j U\psi^{(j)}\phi_0, U\psi^{(i)}\phi_0)| \leqslant B_j(\rho_i + \eta_i). \qquad (3.17)
$$

From (3.12), using (3.13)–(3.15) and (3.17), we finally have

$$
|(\psi^{(j)}, N_S\psi^{(i)})| \leqslant \|N_A\phi_0\|(\eta_i + \eta_j) + B_j(3\eta_i + \rho_j + \rho_i) + B_i\eta_j. \qquad (3.18)
$$

Equation (3.18), obtained using Assumptions 1 and 2 together with the additional hypothesis, Assumption 3, implies the thesis of the theorem of Sec. II, since it shows that the η and ρ terms cannot vanish, unless Eq. (2.10) holds.

We can now use inequality (3.18) to see how well one can approximate a VPD measurement scheme when an additive conservation law holds and N_S couples different eigenmanifolds of the measured quantity. We assume that the hypotheses, Assumptions 2 and 3, are satisfied. The quantities B_l must be considered as given. The only quantity which can be monitored to reduce the malfunctioning of the apparatus is, as is well known,[6–8] the expectation value $(\phi_0, N_A^2\phi_0) \equiv \|N_A\phi_0\|^2$ of N_A^2 on the initial state ϕ_0. Let us then assume that we can make very small the quantities η_l and ρ_l by making $\|N_A\phi_0\|$ very large.

In the limit of very small η_l and ρ_l, Eq. (3.18) yields

$$\eta_i + \eta_j \geqslant |(\psi^{(j)}, N_S \psi^{(i)})| / \|N_A \phi_0\|. \tag{3.19}$$

This relation gives the lower bound for the value-nonpreserving terms in the case of a value-preserving distorting measurement process. It holds under Assumptions 1 and 3 and the condition, Assumption 2, on N_S.

Obviously, the previous derivation of the bounds for VPD measurements also yields bounds for the VPND scheme. In such a case Assumption 2 can be further relaxed. In fact, let us consider a pair of system states $\psi_{i\alpha} \in E_i \cap \mathscr{D}_{N_S}$ and $\psi_{j\beta} \in E_j \cap \mathscr{D}_{N_S}$. Then Assumption 2 is trivially satisfied for the one-dimensional linear manifolds $E_{i\alpha}$ and $E_{j\beta}$ spanned by $\psi_{i\alpha}$ and $\psi_{j\beta}$, respectively. The evolution equation for the state $\psi_{i\alpha}\phi_0$ then takes the form (2.5). Due to the same formal structure of Eqs. (2.5) and (2.7), the derivation given above can then be repeated step by step leading to the result

$$\eta_{i\alpha} + \eta_{j\beta} \geqslant |(\psi_{j\beta}, N_S \psi_{i\alpha})| / \|N_A \phi_0\|, \tag{3.20}$$

with obvious definitions of $\eta_{i\alpha}$, $\eta_{j\beta}$. Equation (3.20) holds under the only Assumptions 1 and 3 for any pair of system states belonging to different eigenmanifolds of the measured observable $\mathscr{M}$ and falling within the domain of N_S.

Before concluding this section, we point out that the results quoted or derived in this and in the previous section can also be relevant for the measurement of an observable $\mathscr{M}$ whose associated self-adjoint operator M does not have a purely discrete spectrum.[2] In such a case, we write $M = \int \lambda \, d\mathscr{E}(\lambda)$, where $\mathscr{E}(\lambda)$ is the spectral family of M. Then each operator $\mathscr{E}(\lambda)$ is associated with the observable corresponding to the yes–no experiment testing whether the value of M is smaller or larger than λ. One says that M can be measured according to an ideal scheme of the VPND or VPD type if all the observables associated with the operators $\mathscr{E}(\lambda)$ for all λ's can be measured according to the considered scheme. This definition seems appropriate since it implies that the state of a system belonging to the linear manifold on which $P_{(\lambda_1,\lambda_2)} = \mathscr{E}(\lambda_2) - \mathscr{E}(\lambda_1)$ projects is not altered (in the VPND scheme), or at least not brought out of this manifold, in a measurement aiming to ascertain whether $\lambda_1 < \lambda < \lambda_2$ or not. These measurements are the physically significant ones for an observable possessing a continuous spectrum. Since the operators $\mathscr{E}(\lambda)$ have a purely discrete spectrum, one can apply the theorems of Sec. II, provided the corresponding assumptions are satisfied for any λ. In particular, if one considers a VPND scheme, the Stein–Shimony theorem of Sec. II implies that

$$e^{iN_S r}\psi \in E(\lambda), \quad \forall \quad \psi \in E(\lambda), \tag{3.21}$$

where $E(\lambda)$ is the closed linear manifold corresponding to the eigenvalue 1 of $\mathscr{E}(\lambda)$. Analogously, in the case of a VPD measurement, the theorm of Sec. II and the bounds of Sec. III hold, provided Assumption 2 is true for all $E(\lambda)$. We note, however, that in this case, Assumption 2 amounts to the requirement that N_S be a bounded operator.

IV. CONCLUSIONS

In this section, we give examples of applications of the theorems of Sec. II about the possibility of ideal measurements of physically significant quantities. All examples, except the last one, were already discussed in Ref. 3. Once again, we have to distinguish the case in which we require that the state of the measured system is not affected by the apparatus (VPND scheme) from that in which we only require that the value of the measured observable is preserved (VPD scheme).

For the case of a VPND scheme, the Stein and Shimony theorem[3] shows that Assumption 1 is incompatible with the scheme (2.2) unless Eq. (2.8) [or (3.21)] holds. As was already stressed, the proof of this theorem does not require any assumption on N_S except its being self-adjoint. Equations (2.8) or (3.21) imply, as is well known,

$$[e^{iN_S r}, e^{iMr'}] = 0, \quad \forall \quad r, r' \in \mathbb{R}. \tag{4.1}$$

Assuming that the system-apparatus interaction is always translationally and rotationally invariant, the total linear momentum $\mathbf{P} = \mathbf{p}_S + \mathbf{p}_A$ and the total angular momentum $\mathbf{J} = \mathbf{J}_S + \mathbf{J}_A$ are additive conserved quantities. Then we get from (4.1) that one can measure, according to a VPND scheme, only the observables of the system satisfying the relations

$$\begin{aligned} e^{i\mathbf{p}_S \cdot \mathbf{r}} e^{iM} e^{-i\mathbf{p}_S \cdot \mathbf{r}} &= e^{iM}, \\ e^{i\mathbf{J}_S \cdot \mathbf{r}} e^{iM} e^{-i\mathbf{J}_S \cdot \mathbf{r}} &= e^{iM}. \end{aligned} \tag{4.2}$$

These equations imply that no quantity of the system which is not translationally and rotationally invariant can be measured according to a VPND scheme, and, in particular, forbid VPND measurements of components of position and angular momentum.

For a VPD measurement, the situation is much less restrictive. However, when the Hilbert space $\mathscr{H}_S$ is finite dimensional, the impossibility theorems always apply. In particular, this means that a value-preserving measurement of a spin component is always forbidden. For angular momentum, we point out[3] that the impossibility theorems forbid a value-preserving simultaneous measurement of the commuting observables L^2 and L_i.

We finally consider a VPD measurement of the internal energy of a quantum system, for example, a system of two particles whose Hamiltonian is

$$H = P^2/2M + p^2/2\mu + V(r), \tag{4.3}$$

where $\mathbf{P}$ and $\mathbf{p}$ are the center of mass and relative momenta, respectively, and r is the relative coordinate of the two particles. Let us assume that the (discrete) eigenvalues of the observable to be measured,

$$h = p^2/2\mu + V(r), \tag{4.4}$$

are only finitely degenerate. Since the system-apparatus interaction is translationally invariant, the total linear momentum

$$\mathscr{P} = \mathbf{P} + \mathbf{p} + \mathbf{p}_A \tag{4.5}$$

(where we have denoted by $\mathbf{p}_A$ the linear momentum of the apparatus) is conserved. The eigenstates of $\mathscr{H}$ are bound states so that they belong to the domain $\mathscr{D}_{p_x}$ of a component, say, p_x, of $\mathbf{p}$. Due to the assumed finite degeneracy, it follows that any eigenmanifold E_i of $\mathscr{H}$ is also contained in $\mathscr{D}_{p_x}$. We can then apply the theorem of Sec. II in the weaker form illustrated in the final part of the section, identifying p_x with N_S. It follows that a VPD measurement is possible only if E_i is an invariant manifold of p_x. This is impossible, since E_i is finite dimensional.

From the preceding considerations, one is led to the conclusion that in quantum mechanics, due to the additive conservation laws of the linear and angular momenta, in many cases and for several physically interesting observables, it is not possible to perform an ideal value-preserving measurement. Obviously, due to the relations we have derived expressing lower bounds for the terms representing a malfunctioning of the apparatus, one sees that from a practical point of view, the unwanted distortions arising in the measurement turn out to be very small due to the macroscopic character of the apparatus (which allows us to make the expectation value of N_A^2 very large). Regarding the practical measurement of an observable, it is then possible to have apparatuses working almost ideally. However, we stress that the impossibility proofs for ideal measurements can have also a conceptual relevance. For instance, arguments of this type have been used to reject some suggestions of instantaneous transfer of information using the reduction of the wave packet.[10]

APPENDIX

We briefly summarize, in this Appendix, some very well-known general theorems and properties of operators in Hilbert space which are used in the derivation of the results of the paper.

We shall always deal with a self-adjoint operator A defined on a domain $\mathscr{D}_A$ of a separable Hilbert space $\mathscr{H}$.

Given such an operator A, we denote by $E(s)$ the spectral family associated with A, and we define the operator function $g(r)$ of A according to

$$g(r) = (e^{iAr} - \mathbb{I})/ir, \quad r \in \mathbb{R}. \tag{A1}$$

We note that, for any given $r \neq 0$, $g(r)$ is everywhere defined and bounded.

We recall some very well-known properties of $g(r)$; for a proof of the quoted theorems, the reader is referred to any good book of functional analysis, e.g., to Ref. 11.

Theorem 1: A is a self-adjoint operator on a Hilbert space $\mathscr{H}$, ψ a vector of its domain $\mathscr{D}_A$, and $g(r)$ the operator defined by (A1). Then the state $g(r)\psi$ coverges strongly for $r \to 0$ to the state $A\psi$:

$$g(r)\psi \xrightarrow[r\to 0]{S} A\psi, \quad \forall \quad \psi \in \mathscr{D}_A. \tag{A2}$$

From the above theorem, it follows trivially that if $\mathscr{P}$ is a projection operator onto a closed linear manifold $E \subset \mathscr{D}_A$, we have

$$g(r)\mathscr{P} \xrightarrow[r\to 0]{S} A\mathscr{P}, \tag{A3}$$

the convergence being in the strong operator topology.

Theorem 2: A is self-adjoint and E a closed linear manifold contained in $\mathscr{D}_A$. The restriction $A \restriction E$ of A to E is a bounded operator on E.

The proof makes use of the fact that the self-adjointness of A implies its closure, and this in turn implies that $A \restriction E$ is closed as an operator on E. Since $E \subset \mathscr{D}_A$, $A \restriction E$ is defined everywhere on E; this, together with its being closed, implies that $A \restriction E$ is bounded. Note that the above means that there exists a constant B such that

$$\|A\psi\| \leqslant B\,\|\psi\|, \quad \forall \quad \psi \in E. \tag{A4}$$

Theorem 3: A is self-adjoint, $E \subset \mathscr{D}_A$ a closed linear manifold, $\mathscr{P}$ the projection operator onto E. Then

$$\|A\mathscr{P}\psi\| \leqslant B\,\|\psi\|, \quad \forall \quad \psi \in \mathscr{H} \tag{A5}$$

and

$$\|g(r)\mathscr{P}\| \leqslant B, \quad \forall \quad r \neq 0, \tag{A6}$$

where B is the bound of $A \restriction E$.

Equation (A5) is a trivial consequence of Theorem 2. Equation (A6) is proved as follows:

$$\|g(r)\mathscr{P}\psi\|^2 = \int \left| \frac{e^{irs} - 1}{r} \right|^2 d\,\|E(s)\mathscr{P}\psi\|^2$$
$$\leqslant \int s^2 d\,\|E(s)\mathscr{P}\psi\|^2 \equiv \|A\mathscr{P}\psi\|^2 \leqslant B^2\|\psi\|^2,$$

where we have used the fact that $|(e^{irs} - 1)/r|^2 \leqslant s^2$, $\forall\, r \neq 0$. Since the adjoint of a bounded operator is bounded with the same bound, we have also

$$\|\mathscr{P}g^+(r)\| \leqslant B, \quad \forall \quad r \neq 0. \tag{A7}$$

From the definition, it follows that $g^+(r) = g(-r)$, so that all the above propositions remain valid by interchanging $g(r)$ and $g^+(r)$.

Finally we recall that when an operator A on $\mathscr{H}$ is considered as an operator on the direct product $\mathscr{H} \otimes \mathscr{R}$ of $\mathscr{H}$ with another separable Hilbert space $\mathscr{R}$, then: (i) A bounded on $\mathscr{H}$ with bound B implies $A \otimes \mathbb{I}_R$ bounded on $\mathscr{H} \otimes \mathscr{R}$ with the same bound. (ii) Let $A(r)$ be a bounded family of operators,

$$\|A(r)\| < C, \ \forall\, r \neq 0 \text{ and let } A(r) \xrightarrow[r\to 0]{S} A, \text{ then}$$
$$A(r) \otimes \mathbb{I}_{\mathscr{R}} \xrightarrow[r\to 0]{S} A \otimes \mathbb{I}_{\mathscr{R}}.$$

The proof is straightforward. For our purposes, we shall use statement (i) to assert that the operator $g^+(r)\mathscr{P}$ and $\mathscr{P}g(r)$ are bounded operators with the bound B of $A \restriction E$, even when they are considered as acting on $\mathscr{H} \otimes \mathscr{R}$. Analogously, statement (ii) implies that

$$g^+(r)\mathscr{P} \otimes \mathbb{I}_{\mathscr{R}} \xrightarrow[r\to 0]{S} A\mathscr{P} \otimes \mathbb{I}_{\mathscr{R}}.$$

[1] E. P. Wigner, Z. Phys. **133**, 101 (1952).
[2] H. Araki and M. M. Yanase, Phys. Rev. **120**, 622 (1960).
[3] H. Stein and A. Shimony, in *Foundations of Quantum Mechanics*, edited by B. d'Espagnat (Academic, New York, 1971), p. 56.

[4]M. M. Yanase, in *Foundations of Quantum Mechanics*, edited by B. d'Espagnat (Academic, New York, 1971), p. 77.
[5]For a detailed discussion on this problem, see Ref. 3.
[6]M. M. Yanase, Phys. Rev. **123**, 666 (1966).
[7]G. C. Ghirardi, F. Miglietta, A. Rimini, and T. Weber, Phys. Rev. D **24**, 347 (1981).
[8]G. C. Ghirardi, A. Rimini, and T. Weber, J. Math. Phys. **23**, 1792 (1982).
[9]For brevity, we shall, throughout the paper, omit (except in the Appendix) the cross symbol and the identities $\mathbf{I}_S$ and $\mathbf{I}_A$ in the operator cross products, and the cross symbol in the vector cross products.
[10]G. C. Ghirardi and T. Weber, Lett. Nuovo Cimento **26**, 599 (1979).
[11]J. Weidmann, *Linear Operators in Hilbert Spaces* (Springer-Verlag, New York, 1980).

An analytic formula for u(3)-boson matrix elements

G. Rosensteel
Department of Physics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118
D. J. Rowe
Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

(Received 25 March 1983; accepted for publication 6 May 1983)

The u(3)-boson Lie algebra is the liquid limit of the symplectic algebra sp(3, R). An analytic formula is given for the u(3)-boson matrix elements in irreducible unitary representations corresponding to the sp(3, R) discrete series. The formula also applies to the generators of the mathematically isomorphic, but physically different, interacting boson algebra.

PACS numbers: 03.65.Fd, 02.20.Sv, 21.60.Fw

I. INTRODUCTION

In this paper, an analytic formula is reported for the generators of the u(3)-boson Lie algebra for those irreducible unitary representations needed in the theory of nuclear collective motion.[1,2] This Lie algebra is the liquid (i.e., many-nucleon) limit of the symplectic Lie algebra sp(3, R).[3–7] Indeed, the ultimate physical justification for the u(3)-boson model depends upon its relationship with the microscopically exact symplectic collective model.

Since the action of sp(3, R) on many-fermion state space decomposes solely into discrete series representations, this series exhausts the physically allowed symplectic collective models. Hence the relevant irreducible representations of the u(3)-boson algebra correspond to sp(3, R) discrete series representations.

The plan of this paper is to first define the u(3)-boson Lie algebra and then define a basis for the discrete series which is symmetry adapted to u(3). The results for the u(3) reduced matrix elements are given in Eqs. (8) and (9).

II. DEFINITION OF THE u(3)-BOSON ALGEBRA

The u(3)-boson Lie algebra is a 22-dimensional semidirect sum consisting of the unitary algebra u(3) plus a 13-dimensional boson (Heisenberg) algebra as the ideal.[1,2]

The unitary subalgebra is just the symmetry algebra of the harmonic oscillator H_0. The remaining eight generators, spanning the special unitary algebra su(3), form a $(\lambda\mu) = (11)$ irreducible tensor operator (via the adjoint representation) which is denoted by $C^{(11)}$.

Since the boson algebra is an ideal, its generators are given conveniently as u(3) irreducible tensor operators too. With respect to su(3), the boson algebra is spanned by (1) a $(\lambda\mu) = (20)$ tensor $A^{(20)}$, (2) a $(\lambda\mu) = (02)$ tensor $B^{(02)}$, the adjoint of $A^{(20)}$, and (3) the identity I, spanning the center of the entire u(3)-boson algebra. Note that both $A^{(20)}$ and $B^{(02)}$ are six-dimensional tensors. Furthermore, their commutators with the harmonic oscillator are

$$[H_0, A^{(20)}] = 2A^{(20)}, \quad [H_0, B^{(02)}] = -2B^{(02)}. \tag{1}$$

Thus $A^{(20)}$ increases the oscillator eigenvalue in any representation by two units, whereas $B^{(02)}$ decreases the eigenvalue by two units.

By coupling the commutator of two tensors to good total symmetry, the commutation relations may be expressed in tensor form. The tensor commutator of $A^{(20)}$ with itself and $B^{(02)}$ with itself always vanishes:

$$[A^{(20)}, A^{(20)}]^{(21)} = 0, \quad [B^{(02)}, B^{(02)}]^{(12)} = 0. \tag{2}$$

However, the commutator of $B^{(02)}$ with $A^{(20)}$ is a multiple of the identity

$$[B^{(02)}, A^{(20)}]^{(\lambda\mu)} = \begin{cases} \sqrt{\frac{8}{3}} N_0 I, & (\lambda\mu) = (00), \\ 0, & (\lambda\mu) = (11) \text{ or } (22), \end{cases} \tag{3}$$

where N_0 is a real constant. This completes the definition of the u(3)-boson algebra.

It is interesting to compare the u(3)-boson algebra with the sp(3, R) algebra.[3] The symplectic algebra is a 21-dimensional simple Lie algebra containing u(3) as its maximal compact subalgebra (spanned by H_0 and $C^{(11)}$). In addition, sp(3, R) contains a (20) tensor $A^{(20)}$ and a (02) tensor $B^{(02)}$; there is no identity operator. Its commutation relations differ from the u(3)-boson algebra only in the commutator of $B^{(02)}$ with $A^{(20)}$, which in sp(3, R) lies in u(3):

$$[B^{(02)}, A^{(20)}]^{(\lambda\mu)} = \begin{cases} \sqrt{\frac{8}{3}} H_0, & (\lambda\mu) = (00), \\ \sqrt{\frac{5}{2}} C^{(11)}, & (\lambda\mu) = (11), \\ 0, & (\lambda\mu) = (22). \end{cases} \tag{4}$$

III. BASIS STATES FOR DISCRETE SERIES REPRESENTATIONS

An orthonormal basis for vectors symmetry adapted to u(3) will be defined now for each discrete series irreducible representation of sp(3, R).[3,8,9] The corresponding u(3)-boson representations are spanned by the same orthonormal basis.

The spectrum of the harmonic oscillator in a discrete series representation is $N_0 + n$, where N_0 is the smallest eigenvalue and n is an even nonnegative integer, $n = 0, 2, 4, \ldots$. The eigenspace belonging to N_0 contains a single u(3) irreducible representation $N_0(\lambda_0\mu_0)$, where $(\lambda_0\mu_0)$ is its su(3) content. This "starting" representation $N_0(\lambda_0\mu_0)$ of u(3) completely determines and, hence, labels a discrete series representation.

A spanning set of vectors for the oscillator eigenspace belonging to $N_0 + n$ is created by applying $n/2$ times the raising tensor $A^{(20)}$ to the starting u(3) irreducible representation space. In order to form states with good u(3) symmetry, it is necessary to first construct all possible tensor couplings of $n/2$ products of $A^{(20)}$. Since the $A^{(20)}$ commute with each

 0022-2488/83/102461-03$02.50

other, only totally symmetric couplings are nonzero. These u(3) symmetric couplings are enumerated by the set[10]

$$\Omega = \{(n_1 n_2 n_3) | n_1 \geqslant n_2 \geqslant n_3 \geqslant 0 \text{ and } n_1, n_2, n_3 \text{ even integers}\}, \quad (5)$$

where $n = n_1 + n_2 + n_3$, and the su(3) content of the symmetric tensor is $(n_1 - n_2, n_2 - n_3)$.

A u(3) symmetry-adapted basis for the discrete series representation $N_0(\lambda_0 \mu_0)$ is given by the tensor products of totally symmetric tensors with the starting u(3) representation. The resulting orthonormal basis is denoted by

$$|(n_1 n_2 n_3) \rho(\lambda \mu) \alpha\rangle, \quad (6)$$

where $(n_1 n_2 n_3)$ ranges over the set Ω, ρ denotes the multiplicity of $(\lambda \mu)$ in the su(3) tensor product $(n_1 - n_2, n_2 - n_3) \otimes (\lambda_0 \mu_0)$, and α indexes a basis for $(\lambda \mu)$. The above vector is an eigenvector of the harmonic oscillator belonging to the eigenvalue $N_0 + n = N_0 + n_1 + n_2 + n_3$. Note that there is an additional possible multiplicity for $(\lambda \mu)$, since two different symmetric tensors $(n_1 n_2 n_3)$ and $(n_1' n_2' n_3')$ with $n = n_1 + n_2 + n_3 = n_1' + n_2' + n_3'$ may produce the same $(\lambda \mu)$ with nonzero multiplicities ρ and ρ'.

IV. MATRIX ELEMENTS OF u(3)-BOSON GENERATORS

Since $A^{(20)}$ and $B^{(02)}$ are irreducible u(3) tensor operators and our basis is symmetry adapted to u(3), the Wigner–Eckart theorem is applicable and, hence, it suffices to ascertain the reduced matrix elements of the u(3)-boson generators. Moreover, since $B^{(02)}$ is the adjoint of $A^{(20)}$, its reduced matrix elements are given by

$$\begin{aligned}\langle (n_1 n_2 n_3) \rho(\lambda \mu) \| B^{(02)} \| (n_1' n_2' n_3') \rho'(\lambda' \mu') \rangle &= (-1)^{\lambda + \mu + \lambda' + \mu'} (\dim(\lambda' \mu')/\dim(\lambda \mu))^{1/2} \\ &\times \langle (n_1' n_2' n_3') \rho'(\lambda' \mu') \| A^{(20)} \| (n_1 n_2 n_3) \rho(\lambda \mu) \rangle^*, \end{aligned} \quad (7)$$

where $\dim(\lambda \mu) = (\lambda + 1)(\mu + 1)(\lambda + \mu + 2)/2$ is the dimension of the su(3) representation $(\lambda \mu)$. Furthermore, from (1), these matrix elements vanish if $n_1' + n_2' + n_3' \neq n_1 + n_2 + n_3 + 2$.

The reduced matrix elements of $A^{(20)}$ are given in two cases.

(1) Closed shell nuclei, $(\lambda_0 \mu_0) = (00)$: If the starting u(3) representation $(\lambda_0 \mu_0) = (00)$, then the basis states are labeled completely by $|(n_1 n_2 n_3) \alpha\rangle$ for $(n_1 n_2 n_3) \in \Omega$, since the coupling to the scalar starting irrep is trivial. The nonzero reduced matrix elements of $A^{(20)}$ are as follows:

$$\langle (n_1' n_2' n_3') \| A^{(20)} \| (n_1 n_2 n_3) \rangle = \begin{cases} \left[\dfrac{N_0 (n_1 + 4)(n_1 - n_2 + 2)(n_1 - n_3 + 3)}{3(n_1' - n_2' + 1)(n_1' - n_3' + 2)}\right]^{1/2}, & n_1' = n_1 + 2, \quad n_2' = n_2, \quad n_3' = n_3, \\ \left[\dfrac{N_0 (n_2 + 3)(n_1 - n_2)(n_2 - n_3 + 2)}{3(n_1' - n_2' + 1)(n_2' - n_3' + 1)}\right]^{1/2}, & n_1' = n_1, \quad n_2' = n_2 + 2, \quad n_3' = n_3, \\ \left[\dfrac{N_0 (n_3 + 2)(n_2 - n_3)(n_1 - n_3 + 1)}{3(n_1' - n_3' + 2)(n_2' - n_3' + 1)}\right]^{1/2}, & n_1' = n_1, \quad n_2' = n_2, \quad n_3' = n_3 + 2. \end{cases} \quad (8)$$

(2) Nonclosed shell nuclei, $(\lambda_0 \mu_0) \neq (00)$: For the general case of nonscalar starting irreps, an $A^{(20)}$ reduced matrix element is the product of a six-$(\lambda \mu)$ U-coefficient[11] times a scalar u(3)-boson matrix element,

$$\begin{aligned}&\langle (n_1' n_2' n_3') \rho'(\lambda' \mu') \| A^{(20)} \| (n_1 n_2 n_3) \rho(\lambda \mu) \rangle \\ &\quad = (-1)^{\lambda' + \mu' - \lambda - \mu} \langle (n_1' n_2' n_3') \| A^{(20)} \| (n_1 n_2 n_3) \rangle \\ &\qquad \times U((\lambda_0 \mu_0)(n_1 - n_2, n_2 - n_3)(\lambda' \mu')(20); \\ &\qquad\quad (\lambda \mu) \rho (n_1' - n_2', n_2' - n_3') \rho'). \end{aligned} \quad (9)$$

The scalar formula (8) is proved by directly verifying the commutation relations, Eqs. (2) and (3). A simple recoupling is required to derive the nonscalar matrix elements (9).

V. DISCUSSION

Comparison of the commutation relations for sp(3, R), Eq. (4), with those for the u(3)-boson algebra, Eq. (3), shows that the u(3)-boson matrix elements approximate the sp(3, R) discrete series provided $N_0 \gg \lambda_0 + \mu_0 + n$. In order to numerically determine the symplectic matrix elements, the solutions to the sp(3, R) commutation relations must be found. This can be achieved efficiently via the method of steepest descent using the u(3)-boson matrix elements as a first approximation. A computer code is available upon request. Note that the matrix elements of sp(3, R) are known analytically in only one special case ($N_0 = \lambda_0$, $\mu_0 = 0$),[12] which does not arise for real nuclei. A recursion formula had been relied on in previous work.[13]

It is interesting to observe that the interacting boson algebra of Arima and Iachello[14] is mathematically isomorphic to our boson (Heisenberg) algebra. The eigenspaces of the harmonic oscillator form the totally symmetric representations of su(6) required in the IBA model. This follows from the observation that $A^{(20)}$ is a six-dimensional boson containing angular momentum zero (s) and angular momentum two (d) bosons. If we take $N_0 = \frac{3}{2}$, $(\lambda_0 \mu_0) = (00)$ and set

$$\begin{aligned} s &= B_0^{(02)}, \quad & s^\dagger &= A_0^{(20)}, \\ d_m &= B_{2m}^{(02)}, \quad & d_m^\dagger &= A_{2m}^{(20)}, \end{aligned} \quad (10)$$

then

$$[s, s^\dagger] = I, \quad [d_{m'}, d_m^\dagger] = (-1)^{m'} \delta_{m', -m} I. \quad (11)$$

and the IBA matrix elements are given from (8) and the Wigner–Eckart theorem. Although our Heisenberg algebra is the dynamical algebra of the IBA su(6) algebra, their physical interpretations are antithetical. In the IBA model, the s and d bosons act within a major oscillator shell, whereas $A^{(20)}$ raises the oscillator eigenvalue by two units.

ACKNOWLEDGMENTS

We would like to acknowledge support by the National Science Foundation and the Science and Engineering Research Council of Canada.

[1]G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **47**, 223 (1981).
[2]D. J. Rowe and G. Rosensteel, Phys. Rev. C **25**, 3236 (1982).
[3]G. Rosensteel and D. J. Rowe, Ann. Phys. **126**, 343 (1980).
[4]K. T. Hecht and D. Braunschweig, Nucl. Phys. A **295**, 34 (1978).
[5]D. R. Peterson and K. T. Hecht, Nucl. Phys. A **344**, 361 (1980).
[6]F. Arickx, Nucl. Phys. A **284**, 264 (1977).
[7]F. Arickx, J. Broeckhove, and E. Deumens, Nucl. Phys. A **318**, 269 (1979).
[8]R. Godement, *Seminaire Cartan* (Ecole Normale Superieure, Paris, 1958).
[9]G. Rosensteel and D. J. Rowe, Int. J. Theor. Phys. **16**, 63 (1977).
[10]R. M. Thrall, Am. J. Math. **64**, 371, (1942).
[11]K. T. Hecht, Nucl. Phys. **62**, 1 (1965).
[12]A. U. Klimyk, "Representations of the Groups Sp(n, R) and Sp(n) in the u(n)-Basis," Preprint ITP-81-37P, Kiev, 1981.
[13]G. Rosensteel, J. Math. Phys. **21**, 924 (1980).
[14]A. Arima and F. Iachello, Ann. Phys. **111**, 201 (1978).

On the binary collision expansion and resummations of it

Antoine Royer
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 1 February 1983; accepted for publication 25 March 1983)

It is pointed out that when there are only two interaction terms in the Hamiltonian, the binary collision expansion (BCE) for the scattering operator can be resummed into a closed form. When there are more than two interaction terms, one can resum the BCE into a continued-fraction-like form by using repeatedly the two interactions resummation. Concise derivations of the BCE are first given, the time-dependent BCE being obtained in a form applicable to both quantum evolution operators and classical frequency modulated oscillators.

PACS numbers: 03.65.Nk, 05.30. – d, 02.30.Mv

I. INTRODUCTION

The so-called "binary collision expansion" (BCE) has been useful in various statistical mechanical problems, as well as being of some interest in itself [1]: it expresses the scattering operator corresponding to a Hamiltonian (or Liouvillian[2]) of the form

$$H = K + V = K + \sum_{\alpha=1}^{N} V_\alpha \tag{1.1}$$

as an infinite expansion in the scattering operators corresponding to each sub-Hamiltonian

$$H_\alpha = K + V_\alpha. \tag{1.2}$$

The main purpose of this paper is to point out that when the Hamiltonian contains only two interaction terms ($N = 2$), the BCE can be resummed into a closed form. When there are more than two interaction terms, one can use the two-interactions-result repeatedly and obtain the scattering operator as a kind of continued fraction.

We start by giving concise derivations of the BCE in both the time and frequency domains. The time dependent BCE is obtained in a form applying to the quantity

$$U(\tau,0) = T_{\leftarrow} \exp\left[-i \sum_{\alpha=1}^{N} \int_0^\tau dt\, V_\alpha(t) \right], \tag{1.3}$$

where $T_{\leftarrow}$ orders the $V_\alpha(t)$ such that the time arguments increase from right to left. If we let

$$V_\alpha(t) = e^{itK} V_\alpha e^{-itK}, \tag{1.4}$$

then

$$U(\tau,0) = e^{i\tau K} e^{-i\tau H} \tag{1.5}$$

is the interaction representation time evolution operator for the Hamiltonian (1.1). But the $V_\alpha(t)$ in (1.3) may also be classical functions (in which case the time ordering is superfluous) or matrices, e.g., $U(\tau,0)$ may represent a classical oscillator or randomly modulated frequency equal to a sum of terms, as is often met in relaxation problems.[3] The above form of the BCE makes manifest its meaning as a "sum over histories," each history corresponding to a partitioning of the time interval $(0,\tau)$ into subintervals during each of which a single one of the $V_\alpha(t)$'s is "turned on."

II. THE BINARY COLLISION EXPANSION

A. Time domain

We have

$$\frac{d}{d\tau} U(\tau,0) = \sum_\alpha \frac{1}{i} V_\alpha(\tau) U(\tau,0), \tag{2.1}$$

$$U(\tau,0) = 1 + \int_0^\tau dt \sum_\alpha \frac{1}{i} V_\alpha(t) U(t,0). \tag{2.2}$$

By applying (2.2) to $U^{(\alpha)}(\tau,t)$ $\equiv T_{\leftarrow} \exp[-i\int_t^\tau dt'\, \Sigma_{\beta(\neq\alpha)} V_\beta(t')]$, while denoting

$$U_\alpha(\tau,t) = T_{\leftarrow} \exp\left[-i \int_t^\tau dt'\, V_\alpha(t') \right], \tag{2.3}$$

we obtain

$$\begin{aligned} U(\tau,0) &= T_{\leftarrow} \{ U_\alpha(\tau,0) U^{(\alpha)}(\tau,0) \} \\ &= T_{\leftarrow} \left\{ U_\alpha(\tau,0) \left[1 + \sum_{\beta(\neq\alpha)} \int_0^\tau dt \frac{1}{i} V_\beta(t) \right. \right. \\ &\quad \left. \left. \times U^{(\alpha)}(t,0) \right] \right\} \\ &= U_\alpha(\tau,0) + \sum_{\beta(\neq\alpha)} \int_0^\tau dt \\ &\quad \times U_\alpha(\tau,t) \frac{1}{i} V_\beta(t) U(t,0). \end{aligned} \tag{2.4}$$

Note that in the integrand in (2.4), only V_α is "turned on" from t to τ. By introducing (2.4) repeatedly into (2.2), we obtain the binary collision expansion

$$\begin{aligned} U(\tau,0) = 1 &+ \sum_\alpha \int_0^\tau dt\, m_\alpha(t,0) \\ &+ \sum_\alpha \sum_{\beta(\neq\alpha)} \int_0^\tau dt \int_0^t dt'\, m_\alpha(t,t') m_\beta(t',0) \\ &+ \sum_\alpha \sum_{\beta(\neq\alpha)} \sum_{\gamma(\neq\beta)} \int_0^\tau dt \int_0^t dt' \int_0^{t'} dt'' \\ &\times m_\alpha(t,t') m_\beta(t',t'') m_\gamma(t'',0) + \ldots, \end{aligned} \tag{2.5}$$

where we denoted

$$m_\alpha(\tau,t) = \frac{1}{i} V_\alpha(\tau) U_\alpha(\tau,t) = \frac{d}{d\tau} U_\alpha(\tau,t). \tag{2.6}$$

Consider, for instance, the integrand of the third term in (2.5): in it, only V_γ is turned on during the time interval $(0,t'')$,

 0022-2488/83/102464-03$02.50

only V_β during (t'',t'), and only V_α during (t',t); whence the interpretation of the BCE as a sum over histories, as mentioned in the Introduction.

Expansion (2.5) may be written more compactly if we define two "operators" λ and Λ as follows[4]:

$$V_\alpha(t)\lambda V_\beta(t') = \begin{cases} V_\alpha(t)V_\beta(t') & \text{if } \alpha \neq \beta, \\ 0 & \text{if } \alpha = \beta, \end{cases} \tag{2.7}$$

$$V_\alpha(t)\Lambda V_\beta(t') = \begin{cases} V_\alpha(t)V_\beta(t') & \text{if } t > t', \\ 0 & \text{if } t < t' \end{cases} \tag{2.8}$$

(λ and Λ are defined as equal to 1 when there are no V's on their left and/or right); thus λ prevents adjacent interactions from being identical, while Λ assures that the interaction on its left is later than that on its right. Inside a product of m_α's, Λ suppresses the overlap between time intervals on its left and right, e.g., if $t > t'$,[5]

$$m_\alpha(t,0)\Lambda m_\beta(t',0) = m_\alpha(t,t')m_\beta(t',0). \tag{2.9}$$

We may now rewrite (2.5) as

$$U(\tau,0) = T_{\leftarrow} \exp\left[\int_0^\tau dt \sum_\alpha m_\alpha(t,0)\lambda\Lambda\right], \tag{2.10}$$

where $T_{\leftarrow}$ orders the *operators* $m_\alpha(t,0)\lambda\Lambda$ such that t increases from right to left. A more direct derivation of (2.10) follows from introducing (2.4) into (2.1):

$$\begin{aligned} \frac{d}{d\tau}U(\tau,0) &= \frac{1}{i}\sum_\alpha V_\alpha(\tau)\left[U_\alpha(\tau,0)\right. \\ &\quad \left. + \frac{1}{i}\sum_{\beta(\neq\alpha)}\int_0^\tau dt\, U_\alpha(\tau,t)V_\beta(t)U(t,0)\right] \\ &= \sum_\alpha m_\alpha(\tau,0)\lambda\Lambda\left[1 + \frac{1}{i}\sum_\beta\int_0^\tau dt\, V_\beta(t)U(t,0)\right] \\ &= \left(\sum_\alpha m_\alpha(\tau,0)\lambda\Lambda\right)U(\tau,0), \end{aligned} \tag{2.11}$$

which is equivalent to (2.10).

B. Frequency domain

When $V_\alpha(t)$ is of the form (1.4),[6] the terms of (2.5) become multiple convolution products [of the operators $e^{itK}m_\alpha(t,0)$] whose Fourier–Laplace transforms are ordinary products; thus, by applying the operator (acting on functions of τ)

$$\hat{F}_\omega \equiv \int_0^\infty d\tau\, e^{i\tau(\omega - K)} \tag{2.12}$$

on (2.5), we get the frequency dependent BCE:

$$\begin{aligned} G(\omega) &= G_0(\omega)\left[1 + \sum_\alpha m_\alpha(\omega)\right. \\ &\quad \left. + \sum_\alpha\sum_{\beta(\neq\alpha)} m_\alpha(\omega)m_\beta(\omega) + \cdots\right] \\ &= G_0\left[1 - \sum_\alpha m_\alpha\lambda\right]^{-1}, \end{aligned} \tag{2.13}$$

where we denoted

$$G(\omega) = \frac{1}{i}\hat{F}_\omega U(\tau,0) = \left(\omega - K - \sum_\alpha V_\alpha\right)^{-1}, \tag{2.14}$$

$$G_0(\omega) = (\omega - K)^{-1}, \quad G_\alpha(\omega) = (\omega - K - V_\alpha)^{-1}, \tag{2.15}$$

$$m_\alpha(\omega) = \hat{F}_\omega m_\alpha(\tau,0) = V_\alpha G_\alpha(\omega). \tag{2.16}$$

An alternative more symmetric form, more convenient for performing resummations,is [this is the $\hat{F}_\omega$ transform of (2.11)]

$$m = \left(1 - \sum_\alpha m_\alpha\lambda\right)^{-1} - 1, \tag{2.17}$$

where

$$m = VG = G_0^{-1}G - 1. \tag{2.18}$$

The above expansions are usually given in terms of the T-matrices

$$T = V + VGV = G_0^{-1}(G - G_0)G_0^{-1} = mG_0^{-1}, \tag{2.19}$$

$$t_\alpha = V_\alpha + V_\alpha G_\alpha V_\alpha = m_\alpha G_0^{-1}, \tag{2.20}$$

for instance,

$$\begin{aligned} G &= \left(\omega - K - \sum_\alpha t_\alpha\lambda\right)^{-1} = G_0 + \sum_\alpha G_0 t_\alpha G_0 \\ &\quad + \sum_\alpha\sum_{\beta(\neq\alpha)} G_0 t_\alpha G_0 t_\beta G_0 + \cdots, \end{aligned} \tag{2.21}$$

$$T = W + WGW \quad \text{with } W = \sum_\alpha t_\alpha\lambda. \tag{2.22}$$

For completeness, let us recall the direct derivation of (2.21) from (2.14): by using the identity $(A - B)^{-1} = A^{-1} + A^{-1}B(A - B)^{-1}$, we get [taking $A = \omega - K$, $B = \Sigma_\alpha V_\alpha$ and then $A = \omega - K - V_\alpha$, $B = \Sigma_{\beta(\neq\alpha)}V_\beta$]

$$G = G_0\left(1 + \sum_\alpha V_\alpha G\right), \tag{2.23}$$

$$G = G_\alpha\left(1 + \sum_{\beta(\neq\alpha)} V_\beta G\right) \tag{2.24}$$

[these are the $\hat{F}_\omega$ transforms of (2.2) and (2.4)]. Repeated substitution of (2.24) into (2.23) yields (2.21); alternatively, substituting (2.24) once into (2.23) and using the operator λ, we obtain (using $V_\alpha G_\alpha = t_\alpha G_0$)

$$G = G_0\left[1 + \sum_\alpha t_\alpha\lambda G_0\left(1 + \sum_\beta V_\beta G\right)\right] = G_0\left(1 + \sum_\alpha t_\alpha\lambda G\right) \tag{2.25}$$

which yields (2.21) on solving for G.

III. RESUMMATIONS OF THE BCE

Consider Eq. (2.17). Because of the operator λ, the quantity $(1 - \Sigma_\alpha m_\alpha\lambda)^{-1}$ is, of course, not the inverse of $1 - \Sigma_\alpha m_\alpha\lambda$, but is just a compact notation for expressing an infinite series. That is, the BCE cannot be resummed into a closed form simply. However, in the special case that the interaction $V = \Sigma_\alpha V_\alpha$ contains only two terms, $V = V_1 + V_2$, a simple resummation is possible. Consider

$$\begin{aligned} m_{12} &= (1 - m_1\lambda - m_2\lambda)^{-1} - 1 \\ &= m_1 + m_2 + m_1m_2 + m_2m_1 + m_1m_2m_1 + \cdots, \end{aligned} \tag{3.1}$$

i.e., the sum of all alternating chains in m_1 and m_2. We distinguish four different contributions to (3.1): the chains which both start and end with m_1 contribute

$$m_1 + m_1m_2m_1 + m_1m_2m_1m_2m_1 + \cdots$$
$$= (1 - m_1m_2)^{-1}m_1 = m_1(1 - m_2m_1)^{-1}; \quad (3.2)$$

the chains starting with m_1 and ending with m_2 contribute

$$m_1m_2 + m_1m_2m_1m_2 + \cdots$$
$$= (1 - m_1m_2)^{-1} - 1 = m_1(1 - m_2m_1)^{-1}m_2. \quad (3.3)$$

There are two other contributions obtained by interchanging m_1 and m_2 above. Summing these four contributions, we get[7]

$$m_{12} = (1 + m_2)(1 - m_1m_2)^{-1}(1 + m_1) - 1. \quad (3.4)$$

For the propagator G_{12} we have (using $m + 1 = G_0^{-1}G$ and $G_1 = G_0 + G_0t_1G_0$)

$$G_{12} = (1 + G_0t_2)\mathscr{G}_{12}(1 + t_1G_0)$$
$$= G_2G_0^{-1}\mathscr{G}_{12}G_0^{-1}G_1 = G_2(G_0^{-1} + \mathscr{T}_{12})G_1, \quad (3.5)$$

where

$$\mathscr{G}_{12} = (\omega - K - t_1G_0t_2)^{-1}, \quad \mathscr{T}_{12} = G_0^{-1}(\mathscr{G}_{12} - G_0)G_0^{-1} \quad (3.6)$$

are the propagator and T-matrix for the "interaction" $t_1G_0t_2$.

When the interaction V contains more than two terms, one can still obtain resummed BCE's by using (3.4) repeatedly, i.e., one first separates $V = \Sigma_\alpha V_\alpha$ into two parts, $V = V_a + V_b$, where V_a and V_b are themselves sums of interactions, and obtains m in terms of m_a and m_b using (3.4); one may then divide the sum V_a into two parts and apply (3.4) to m_a (and likewise with m_b), and so on. One thereby obtains m in a form akin to a continued fraction. For instance, if $V = V_1 + V_2 + V_3$, we have

$$m_{123} = (1 + m_{23})\frac{1}{1 - m_1m_{23}}(1 + m_1) - 1$$
$$= (1 + m_3)\frac{1}{1 - m_2m_3}(1 + m_2)$$
$$\times \frac{1}{1 - m_1(1 + m_3)\dfrac{1}{1 - m_2m_3}(1 + m_2) - m_1}$$
$$\times (1 + m_1) - 1. \quad (3.7)$$

[1] K. M. Watson, Phys. Rev. **103**, 489 (1956); T. D. Lee and C. N. Yang, Phys. Rev. **105**, 1119 (1957); **113**, 1165 (1959); A. J. F. Siegert and Ei Teramoto, Phys. Rev. **110**, 1232 (1958).
[2] U. Fano, Phys. Rev. **131**, 259 (1963).
[3] See, e.g., A. Royer, J. Math. Phys. **24**, 380 (1983), and references therein.
[4] For a usage of the operator Λ in another context, see A. Royer, Phys. Rev. A **7**, 1078 (1973).
[5] $m_\alpha(t,0)\Lambda m_\beta(t',0) = V_\alpha(t)U_\alpha(t,t')U_\alpha(t',0)\Lambda V_\beta(t')U_\beta(t',0)$
$= V_\alpha(t)U_\alpha(t,t')V_\beta(t')U_\beta(t',0)$, since $U_\alpha(t',0)\Lambda V_\beta(t') = V_\beta(t')$.
[6] In the case that the $V_\alpha(t)$ are classical functions analytic in t, we can take $K = -id/dt$.
[7] This is essentially equivalent to Eq. (44) of Ref. 2.

Charged-particle off-shell scattering: common structure of all two-body off-shell scattering quantities for Coulomb plus rational separable potentials

H. van Haeringen
Department of Mathematics, Delft University of Technology, Delft, The Netherlands

(Received 29 July 1982; accepted for publication 23 September 1982)

We establish a common structure of all two-body off-shell scattering quantities (expressed in momentum space or coordinate space) associated with Coulomb plus rational separable potentials. We present expressions in so-called maximal-reduced closed form, including new formulas: (i) for the off-shell Jost state for the Coulomb potential, (ii) for the off-shell Jost function associated with the Coulomb plus Yamaguchi potential, and (iii) for the scattering, regular, and Jost states in coordinate representation for Coulomb plus simple separable potentials for all l.

PACS numbers: 03.65.Nk, 25.10. + s

In charged-particle scattering theory potentials consisting of the sum of the Coulomb potential and a short-range potential play an important role. The two-body Coulomb interaction has been extensively studied, see Ref. 1 and references quoted therein. The additional short-range interaction depends on the nature of the charged particles. In reactions of more than two particles the so-called off-shell (part of the) interaction comes into play. Various two-body scattering quantities have been introduced, off-shell and on-shell: T matrix, Green function (resolvent), physical scattering wave function, regular wave function, Jost state, Jost function, and quantities that are derived from these.

Separable potentials have attracted a great deal of attention mainly because of calculational advantages. Justification of the use of such potentials is provided by the fact that short-range local potentials can be well approximated by rank-N separable potentials. For model calculations with Coulomb plus separable potentials see, e.g., Ref. 2. The Coulomb potential and more generally the sum of the Coulomb potential and a short-range potential is not suitable for separable approximation. The long range $[V(r) \propto r^{-1}, r \to \infty]$ causes certain difficulties and peculiarities which show up as typical singularities in scattering quantities, cf. Eqs. (3). This fact makes any potential with an $\propto r^{-1}$ tail interesting from a mathematical point of view. At the same time these singularities present problems because not all of them can be handled by the existing numerical techniques.[3] Therefore they should first be studied analytically and then taken care of by a combination of analytic and numerical methods.

Clearly it is important to have exact expressions in closed analytic form for the aforementioned scattering quantities associated with Coulomb plus separable potentials. Not only are they useful for model calculations, they also serve as guideline for and check on numerical calculations with Coulomb plus other short-range potentials.

The main purpose of this paper is to reveal the general analytic structure of the aforementioned two-body off-shell scattering quantities associated with a potential that consists of the sum of the Coulomb potential and a so-called rational separable potential, for all partial waves, $V_l = V_{cl} + V_{\mathrm{RS}\,l}$. Important new formulas and known formulas will be presented as illustration of this structure.

We shall use the same notations as in previous related work.[1,4] Throughout we restrict ourselves to rotationally invariant potentials. The Coulomb potential is $V_{cl}(r) = 2k\gamma/r \equiv -2s/r$. A rational separable potential is, roughly speaking, a rank-N separable potential with form factors $|g_l\rangle$ that are in momentum space expressed by p^l times a rational function of p^2. In coordinate space $\langle r|g_l\rangle$ is a finite linear combination of functions of the form $r^{l-n}P^{(i)}(r)\exp(-\beta_i r)$ where $P^{(i)}$ are polynomials. The simplest potential of this large class is obtained by taking a rank-one potential $V_{sl} = -\lambda_l |g_l\rangle\langle g_l|$ with form factor

$$\langle p|g_l\rangle = \langle p|g_{\beta l}\rangle = (2/\pi)^{1/2} p^l (p^2+\beta^2)^{-l-1},$$
$$\langle r|g_l\rangle = (\tfrac{1}{2}ir)^l e^{-\beta r}(rl!)^{-1}.$$

We have introduced the name "simple separable potential" for V_{sl}. The scattering quantities for V_{cl} plus a rational separable potential have a common analytic structure. In effect it suffices to consider only $V_{cl} + V_{sl}$. The generalization from rank-one to rank-N involves in essence only matrix inversion. Therefore we shall restrict ourselves to Coulomb plus simple separable potentials.

We have derived all the aforementioned scattering quantities, as far as yet unknown, in closed analytic form.[4-8] Most of these are in a form which we call "maximal-reduced." This means that no further substantial reduction or simplification is possible. An instructive and interesting example is the expression for the off-shell Jost function for the Coulomb plus Yamaguchi potential given by Eq. (10). The closed form here has been broken up into elementary, "bare" components. Another, simpler example of a maximal-reduced formula is given by Eq. (8).

It is important to realize that the maximal-reduced form is not necessarily identical with the or an optimal form. Clearly the qualification "optimal" depends on the end in view. For the purpose of numerical calculations the maximal-reduced form will usually coincide with the optimal form. Sometimes, especially for abstract discussions or certain derivations and formula manipulations, an implicit expression is much more useful than an explicit one in closed maximal-reduced form. An extreme example: In the context of discussing properties of solutions of differential equations, the scattering wave function should *not* be given in closed form. Its definition—a particular solution of Schrödinger's differential equation—then is the optimal form.

Another interesting point concerns standardization. Often different authors obtain different but equivalent ex-

 0022-2488/83/102467-04$02.50

pressions for one and the same object. As "some closed forms may be more closed than others," it is obviously important to have a prescription for reducing an expression to one particular form, preferably as convenient and as simple as possible.

To become more specific now, we are going to formulate three statements, I–III. The first two concern the maximal-reduced form of the aforementioned scattering quantities (and of any object derived from these) associated with $V_{cl} + V_{sl}$. They hold *a fortiori* for V_{cl} and V_{sl} separately. The third statement concerns the explicit analytic form of the typical on-shell singularity (caused merely by the long range of the Coulomb potential) which we have studied before in detail.[4]

I. The Jost functions, the off-shell Jost functions, and all two-body off-shell scattering quantities expressed in the momentum representation are conveniently expressed in terms of simple (finite) combinations of elementary functions (polynomials, complex powers), gamma functions, and either the hypergeometric function $F_{i\gamma}(z) := {}_2F_1(1,i\gamma;1+i\gamma;z)$ or (especially for general l) the closely related function

$$F^{(l)}(z) := (l+1+i\gamma)^{-1}\,{}_2F_1(1,i\gamma-l;i\gamma+l+2;z),$$
$$l = 0,1,2,.... \tag{1}$$

For the argument z fifteen expressions play a role:

$$AB, B^2, B, Ba, Ba^{-1}, Bb, Bb^{-1}, ba, ba^{-1}, a, a^{-1}, aa', a/a', a'/a, (aa')^{-1},$$

where

$$A := (\alpha + ik)/(\alpha - ik), \quad a := (p-k)/(p+k),$$
$$a' := (p'-k)/(p'+k),$$
$$B := (\beta + ik)/(\beta - ik), \quad b := (q-k)/(q+k).$$

The essential difference between a and b concerns their range in the complex plane: p is always real positive whereas q is in general complex with Im $q \geqslant 0$.

A typical and interesting example is provided by the closed expression for the off-shell Jost function for the Coulomb plus Yamaguchi potential, see Eq. (10). This is a nice illustration of "being in maximal-reduced form." It is clear that no further essential reduction or simplification is possible despite the complexity of this expression.

II. All scattering quantities expressed in the coordinate representation consist of simple (finite) combinations of elementary functions (polynomials, complex powers), gamma functions, and a "new" function V of four variables which is defined, for Re z and Re a positive, by

$$V(a,c;x,-z) := z^{-a}[\Gamma(a)]^{-1} \times \int_0^z t^{a-1}(1+t)^{c-a-1}e^{-xt}\,dt. \tag{2}$$

Special cases of this function include [apart from elementary functions and gamma functions] ${}_2F_1(1,c;1+a;z)$, ${}_1F_1(a;1+a;xz)$, ${}_1F_1(a;c;x)$, and $U(a,c,x)$ (in the notation of Ref. 9). The factors $(1+t)^{c-a-1}$ and e^{-xt} in the integrand in Eq. (2) can be expanded in powers of t and xt, respectively. In this way the function V can be related, e.g., to certain confluent (generalized) hypergeometric functions of two variables: Φ_1 and Ξ, see Ref. 10, p. 225. However, the use of these functions seems not advantageous and hence such expansions constitute no simplification. Rather, it seems that more complicated and less transparent expressions result. We contend that the function V, which is effectively just a primitive of a simple elementary function, is the most elegant and best choice for our maximal-reduced forms. Interesting and illuminating examples can be found in Eqs. (11) and (19).

III. There exists a typical "Coulomb singularity" at the on-shell point: $p = k$ or $q = k$. This singularity is generated *only* by the long tail (range) of the Coulomb potential: $\propto r^{-1}$, $r \to \infty$. It has the same analytic form for all partial waves and for all potentials of the form $V_l = V_{cl} + V_{sl}$ where V_{sl} is any reasonable short-range potential either local or nonlocal (separable), see previous publications.[4] It is convenient to have a simple, "standard" notation. To this end we use two functions, ω and Ω, defined by

$$\omega := \omega(k,q,\gamma) := \left(\frac{q-k}{q+k}\right)^{i\gamma}(e^{\pi\gamma/2}/\Gamma(1+i\gamma)), \tag{3a}$$

where Im $q \geqslant 0$ and k is real positive, and

$$\Omega := \Omega(k,p,\gamma) := \lim_{\epsilon \downarrow 0}\left(\frac{p-k-i\epsilon}{p+k+i\epsilon}\right)^{-i\gamma} \times (e^{\pi\gamma/2}/\Gamma(1-i\gamma)), \tag{3b}$$

where p and k are real positive. Since clearly

$$\lim_{\eta \downarrow 0}\omega(k,p+i\eta,\gamma) = \Omega^*(k,p,\gamma), \quad p>0,\ k>0,$$

ω may be considered as an extension of Ω. We have for the off-shell Jost function and Jost state associated with V_l,

$$\lim_{q \to k}\omega f_l(k,q) = f_l(k) \equiv f_l, \tag{4a}$$

$$\lim_{q \to k}\omega |kql\uparrow\rangle = |kl\uparrow\rangle, \tag{4b}$$

and for the off-shell T matrix and scattering state,

$$\lim_{p \to k}\Omega T_l|p\rangle = T_l|kl\infty\rangle = V_l|kl+\rangle, \tag{5a}$$

$$\lim_{p \to k}\Omega\,|kpl+\rangle = |kl+\rangle. \tag{5b}$$

We point out that the off-shell regular state is continuous at $p = k$:

$$\lim_{p \to k}|kpl\,\textcircled{R}\rangle = |kl\,\textcircled{R}\rangle. \tag{6}$$

These regular states are defined by

$$|kl\,\textcircled{R}\rangle := (1+\bar{G}_l^{\dagger}V_l)|kl\rangle,$$
$$|kpl\,\textcircled{R}\rangle := (1+\bar{G}_l^{\dagger}V_l)|pl\rangle,$$

where $\bar{G}_l^{\dagger}$ is the Green function for the regular state, cf. Ref. 5 and Report 135 in Ref. 4. We have in general [cf. Eq. (15)]

$$|kl\,\textcircled{R}\rangle = |kl+\rangle f_l,$$

$$\lim_{r \to 0}\langle r|kpl\,\textcircled{R}\rangle/\langle r|pl\rangle = 1.$$

In general the basic complexity of the maximal-reduced closed form for a certain scattering quantity is the same for all $l = 0,1,2$. However, for $l = 0$ such an expression often has,...a much more attractive and simple form. A striking

example is the maximal-reduced form for the off-shell Coulomb Jost function:

$$f_c(k,q) = b^{-i\gamma} \equiv \left(\frac{q+k}{q-k}\right)^{i\gamma}, \quad l=0, \tag{7a}$$

$$f_{cl}(k,q) = 1 - c_{l\gamma} A_l((q^2/k^2);\gamma^2) + c_{l\gamma}(q/k)^l b^{-i\gamma} P_l^{(-i\gamma,i\gamma)}(\bar{u}), \quad l=0,1,2,.... \tag{7b}$$

Here $\bar{u} := (q^2+k^2)/(2qk)$, $c_{l\gamma} := \Pi_1^l(1+\gamma^2/n^2)^{-1}$, and A_l is a polynomial of degree l,

$$c_{l\gamma} A_l(q^2/k^2;\gamma^2) = {}_3F_2(-l,1,1,\tfrac{1}{2};1+i\gamma,1-i\gamma;1-q^2/k^2).$$

Clearly, only simple elementary functions play a role here. Other examples involving more complicated functions are provided by the closed forms for $\langle p|kl\uparrow\rangle_c$, $\langle p|T_{cl}|p'\rangle$, and $\langle p|T_{cl}|ql\uparrow\rangle_0$, see Refs. 6 and 7. Examples of the opposite case in which the maximal-reduced closed formula for general l is very similar to the corresponding one for $l=0$ are given by Eq. (19) and (cf. Ref. 1):

$$\langle g_{\alpha l}|G_{cl}|g_{\beta l}\rangle = \frac{(2l+1)!!}{-(2l)!!}\,\frac{(\alpha+\beta)^{-2l-1}}{(\alpha-ik)(\beta-ik)}F^{(l)}(AB). \tag{8}$$

As illustration of the claimed general structure of scattering quantities we present below a set of new formulas which supplement previous results.[4–6] We have

(i) *For $l=0$*: The Jost function $f \equiv f(k)$ associated with the Coulomb plus Yamaguchi potential is given in maximal-reduced form by

$$f = f_c \rho[\rho + B^{-i\gamma}F^{(0)}(B)]^{-1}, \tag{9}$$

where

$$f_c = e^{\pi\gamma/2}/\Gamma(1+i\gamma),$$

$$\rho := (\beta-ik)(\beta^2+k^2)\tau_{cs}^{-1} = (\beta+ik)[\lambda^{-1}(\beta-ik)^2 - (2\beta)^{-1}F^{(0)}(B^2)].$$

For the corresponding off-shell Jost function we have obtained

$$f(k,q) = \frac{b^{-i\gamma}\rho + B^{-i\gamma}F^{(0)}(Bb)(q-k)(q+i\beta)^{-1}}{\rho + B^{-i\gamma}F^{(0)}(B)}, \tag{10}$$

and the Coulomb off-shell Jost state $|kq\uparrow\rangle_c$ is represented in coordinate space by

$$\langle r|kq\uparrow\rangle_c = (2/\pi)^{1/2}\frac{2k}{iq}e^{ikr}\int_{(q-k)/2k}^{\infty}\left(\frac{tb^{-1}}{1+t}\right)^{i\gamma}e^{2ikrt}\,dt. \tag{11}$$

It is interesting to note that

$$\lim_{r\to 0}(2/\pi)^{-1/2}qr\langle r|kq\uparrow\rangle_c = b^{-i\gamma} = f_c(k,q).$$

This is a special case of the general formula

$$f_l(k,q) = \lim_{r\to 0}\langle r|kql\uparrow\rangle/\langle r|ql\uparrow\rangle_0 \tag{12}$$

which holds for any potential, see, e.g., Eq. (2.113) of Ref. 8.

(ii) *For general l ($=0,1,2,...$)*: The Jost function f_l follows from

$$f_l^{-1}f_{cl} = 1 + \tfrac{1}{2}\pi k\tau_{csl}\langle g_l|kl+\rangle_c\,{}_c\langle kl\downarrow|g_l\rangle, \tag{13}$$

the physical scattering wave function is given by

$$\langle r|kl+\rangle = \langle r|kl+\rangle_c - \tau_{csl}\langle r|G_{cl}|g_l\rangle \times \langle g_l|kl+\rangle_c, \tag{14}$$

the so-called regular solution by

$$\langle r|kl\,\textcircled{R}\rangle = \langle r|kl+\rangle f_l, \tag{15}$$

and the Jost solution ("irregular" solution) follows from

$$\langle r|kl\uparrow\rangle f_l^{-1}f_{cl} = \langle r|kl\uparrow\rangle_c - \tau_{csl}\langle r|G_{cl}|g_l\rangle\,\langle g_l|kl\uparrow\rangle_c. \tag{16}$$

These expressions are constructed from (in part) common components, some of which are well known, e.g., [cf. Eq. (8)],

$$\tau_{csl}^{-1} = \lambda_l^{-1} + \langle g_l|G_{cl}|g_l\rangle,$$

$$\langle g_l|kl+\rangle_c = \langle k|g_l\rangle B^{-i\gamma}f_{cl}^{-1} = (2/\pi)^{1/2}k^l(\beta^2+k^2)^{-l-1} \times B^{-i\gamma}e^{-\pi\gamma/2}\Gamma(l+1+i\gamma)/l!. \tag{17}$$

The until now unknown components are given in closed form by

$$\langle g_l|kl\uparrow\rangle_c = {}_c\langle kl\downarrow|g_l\rangle = {}_0\langle kl\downarrow|g_l^c\rangle f_{cl} = f_{cl}(2/\pi)^{1/2}k^{-1}(\beta-ik)^{-1} \times (4k)^{-l}(2l+1)!(l!)^{-2}F^{(l)}(B), \tag{18}$$

and

$$\langle r|G_{cl}|g_l\rangle = \frac{(-1)^l}{l!kr}\left(\frac{2ik^2r}{\beta^2+k^2}\right)^{l+1}\left(\frac{\beta-ik}{\beta+ik}\right)^{i\gamma}e^{ikr} \times \int_0^{(\beta+ik)/-2ik} t^{l+i\gamma}(1+t)^{l-i\gamma}e^{2ikrt}\,dt. \tag{19}$$

Finally, for the off-shell Jost state $|kql\uparrow\rangle$ we have obtained the nontrivial, interesting relation

$$|kql\uparrow\rangle - |kql\uparrow\rangle_c = [f_l(k,q) - f_{cl}(k,q)](k/q)^{l+1}f_{cl}^{-1} \times [|kl\uparrow\rangle_c + G_{cl}|g_l\rangle 2(\pi k)^{-1}\langle g_l|kl+\rangle_c^{-1}]. \tag{20}$$

Here the form factor g_l may be arbitrary.

Summarizing, we have revealed the common analytic structure of all two-body off-shell scattering quantities, expressed either in momentum space or in coordinate space, for Coulomb plus rational separable potentials, for all partial waves. The complete three-dimensional Coulomb scattering quantities have the same structure, as follows from known results. We have discussed maximal-reduced closed forms and proposed a standardization for the presentation of expressions for the scattering quantities. The use of maximal-reduced forms facilitates comparison of different expressions for the same object and offers an optimal form for subsequent numerical calculations which in turn give the connection with experimental data. *New* maximal-reduced closed forms have been given in Eqs. (9)–(11) and (18)–(20). In Eq. (3) we have given a simple formula for the typical on-shell singularity caused merely by the long range of the Coulomb potential.

[1] H. van Haeringen, Phys. Lett. A **82**, 359 (1981).

[2] E. O. Alt, W. Sandhas, and H. Ziegelmann, Phys. Rev. C **17**, 1981 (1978); H. van Haeringen, L. P. Kok, and W. Plessas, J. Phys. G: Nucl. Phys. **7**, L193 (1981); W. Schweiger, W. Plessas, L. P. Kok, and H. van Haeringen, Preprint UNIGRAZ-UTP 02/82, University of Graz, 1982.

[3]L. P. Kok, Nucl. Phys. A **353**, 171c (1981); C. Chandler, *ibid.* **353**, 129c (1981).
[4]H. van Haeringen, Phys. Rev. A **18**, 56 (1978); Nucl. Phys. A **253**, 355 (1975); **327**, 77 (1979); H. van Haeringen and R. van Wageningen, J. Math. Phys. **16**, 1441 (1975); H. van Haeringen, J. Math. Phys. **17**, 995 (1976); **18**, 927 (1977); **19**, 1379 (1978); **20**, 861, 1109, and 2520 (1979); Report 135, University of Groningen, 1979.
[5]H. van Haeringen and L. P. Kok, Phys. Lett. A **82**, 317 (1981); Teor. Mat. Fiz. **50**, 100 (1982) (in Russian).
[6]H. van Haeringen, J. Math. Phys., to appear (five papers).
[7]H. van Haeringen, Reports 82 06–82 12, Delft University of Technology, 1982 (in preparation).
[8]H. van Haeringen, Report 150, Groningen, 1979; Report 82 05, 2nd ed., Delft, 1982.
[9]W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics* (Springer, New York, 1966).
[10]A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.

Nonsingular modified Lippmann–Schwinger equation for two charged particles

J. Dušek
Department of Physics, Institute of Mechanical and Electrical Engineering, Nejedlého sady 14, 306 14 Plzeň, Czechoslovakia

(Received 8 December 1981; accepted for publication 29 October 1982)

We consider the partial waves of the two-body Lippmann–Schwinger equation for the T matrix in the case of the sum of the Coulomb plus a more rapidly vanishing potential. Using the knowledge of explicit forms of the integrals determining the operation of the Lippmann–Schwinger operator on the singularities that characterize the on-shell behavior of the Coulomb-like T matrix, we extract a manifestly nonsingular integral equation from the Lippmann–Schwinger equation. We show that the half- and on-shell inhomogeneous terms and solutions can be made as smooth as the momentum representation partial projection of the non-Coulomb part of the potential.

PACS numbers: 03.80. + r, 11.20. − e, 02.30.Rz

1. INTRODUCTION

It is well known[1] that the most straightforward means of getting two-body scattering amplitudes, half-shell and off-shell T matrices—the Lippmann–Schwinger equation—fails for the Coulomb interaction. This difficulty originates in the nonintegrable singularity of the Lippmann–Schwinger kernel. In the case of potentials vanishing at infinity like r^{-1} and more slowly, the Lippmann–Schwinger operator is not defined in the neighborhood of the energy shell even as a distribution.[2] In the two-body theory the knowledge of analytic forms of many Coulomb quantities enables one in the case of charged particles to bypass this difficulty. The basic means is the Gell-Mann–Goldberger two-potential formalism.[3] The already classic approach uses the known analytic expressions for the coordinate representation pure Coulomb wave functions.[4] By this method it is also possible to obtain the off-shell wave functions[5] and incidentally[6] half- and off-shell T matrices. The momentum representation approach has been applied by van Haeringen *et al.*[7–9] to the case of the potential $V = V_C + V_S$, where V_C is the Coulomb and V_S a separable potential. The two-potential formalism yields the T matrix as the sum

$$T = T_C + T_{SC},$$

T_C being the pure Coulomb T matrix. T_{SC} is calculated in two steps. First the Lippmann–Schwinger equation

$$t_{SC}(z) = V_S + V_S G_C(z) t_{SC}(z) \tag{1}$$

is solved. In Eq. (1) $G_C(z)$ denotes the pure Coulomb Green's function. The short range term T_{SC} is then

$$T_{SC}(z) = [1 + V_S G_C(z)] t_{SC}(z) [1 + G_C(z) V_S]. \tag{2}$$

We have used the same momentum space approach to investigate a more general class of potentials.[10] We have shown this method to be applicable in principle to all potentials vanishing in the coordinate representation at least like $r^{-1-\epsilon}$ for some $\epsilon > 0$. In practical applications, however, the presence of the singular operator $V_S G_C(z)$ in Eqs. (1) and (2) may cause difficulties even if its singularities have been investigated in detail.[11,12]

The operator $V_S G_C(z)$ is given by the pure Coulomb T matrix $T_C(z)$. For the complete Coulomb T matrix several equivalent analytic expressions have been known for a long time.[7,13] Some years ago the s- and p-partial waves were found in closed form.[7,8] Recently we gave a compact expression for an arbitrary l th partial wave.[11] However, no such information is available in the three-body case, i.e., no three-body quantity is known in closed form. The only way to obtain the three-body transition amplitudes is the direct solution of the Faddeev equations.[14] The Faddeev operators are formed by the two-body T matrices, more exactly by the two-body operators $T(z)G_0(z)$. The knowledge of their singular properties in the case of Coulomb-type interactions should enable us to extend the validity of the Faddeev equations to the case of charged particles. We have given a detailed study of the partial wave projections of these operators both for the pure Coulomb[12] and a general Coulomb-type[10] interaction. The two-body case should give a hint how to exploit this piece of information. From this point of view the mentioned method involving the two-potential formula is of little use. It is therefore of interest to study the possibility of regularization of the two-body Lippmann–Schwinger (LS) equation if the interaction is of the Coulomb type.

In this paper we use the notation of Ref. 7. In particular the basis states in momentum representation are normalized so that

$$\langle \mathbf{p}|\mathbf{k}\rangle = \delta(\mathbf{p} - \mathbf{k}) \tag{3}$$

and the l th partial wave is

$$\langle p|k,l\rangle = k^{-2}\delta(p - k). \tag{4}$$

In the coordinate representation the states

$$\langle \mathbf{r}|\mathbf{k}\rangle = (2\pi)^{-3/2} e^{i\mathbf{k}\mathbf{r}} \tag{5}$$

and

$$\langle r|k,l\rangle = (2/\pi)^{1/2} i^l j_l(kr) \tag{6}$$

correspond to the states (3) and (4), respectively. The partial wave projections of a rotationally invariant operator A are defined by

$$\langle p|A_l|p'\rangle = \int d\hat{p}\, P_l(\hat{p}\cdot\hat{p}')\langle \mathbf{p}|A|\mathbf{p}'\rangle, \tag{7}$$

$\hat{p}$ standing for the unit vector $\mathbf{p}/p$. The units are such that $\hbar = 1$ and $2m = 1$, m being the reduced mass of the pair of

 0022-2488/83/102471-10$02.50

particles. In what follows k will be considered to be the square root of the real energy E plus some small imaginary part $\epsilon > 0$:

$$k = (E + i\epsilon)^{1/2}, \quad \mathrm{Im}(k) > 0. \tag{8}$$

For $\epsilon \to 0$ we shall denote

$$k_0 = (E + i0)^{1/2}. \tag{9}$$

We consider a system of two charged particles interacting via a superposition of the pure Coulomb potential

$$V_C(r) = 2k\gamma/r \tag{10}$$

(here we introduce the dimensionless Sommerfeld parameter $\gamma = e_1 e_2/2k$) and a potential V_S which is short ranged in the sense that it is, in the coordinate representation, a spherically symmetric potential vanishing at infinity as

$$V_S(r) \sim r^{-1-\nu} \tag{11}$$

for some $\nu > 0$ and such that the usual scattering theory holds; i.e., the wave operators exist and are complete for the Hamiltonians H_0, $H = H_0 + V_S$. A sufficient condition for this to be true would be[15]

$$\int_1^\infty |V_S(r)|dr + \int_0^1 r|V_S(r)|dr < \infty. \tag{12}$$

However, in what follows we shall work in the l th partial wave projection of the momentum space. Instead of the condition (12) we therefore formulate conditions on the partial wave projections

$$V_{S,l}(p,p') = \langle p,l|V_S|p',l\rangle = \frac{2}{\pi}\int_0^\infty V_S(r) j_l(pr) j_l(p'r) r^2 dr. \tag{13}$$

These conditions are stated as Assumption 1 in Sec. 2. They have been shown[10] to include all local spherically symmetric potentials vanishing for $r \to \infty$ as (11) but they also admit nonlocal potentials. What is most important, the kernels of the partial wave projections of Eq. (1), and consequently also of the partial wave projections of the ordinary LS equation.

$$T(E + i\epsilon) = V + VG_0(E + i\epsilon)T(E + i\epsilon), \tag{14}$$

with $V = V_S$, have been shown[10] to be compact if V_S satisfies these conditions.

In the present paper we are concerned with the LS equation (14) for potentials

$$V = V_C + V_S. \tag{15}$$

Due to the assumed rotational invariance of the potential (15) the T matrix—the solution of Eq. (14)—is also rotationally invariant and its partial waves are defined via Eq. (7). At this point it should be observed that though the partial wave projections of the on-shell Coulomb and Coulomb-type scattering amplitude do not exist[16,17] in the ordinary sense (7), no problems arise in the half- and off-shell case[11]; i.e., the partial waves $\langle p|T_l(E + i\epsilon)|p'\rangle$ exist for $\epsilon \geqslant 0$ for all values $p' \neq p$ and the partial wave series

$$\langle \mathbf{p}|T(E + i\epsilon)|\mathbf{p}'\rangle = \sum_{l=0}^\infty \frac{2l+1}{4\pi} P_l(\hat{p}\cdot\hat{p}')\langle p|T_l(E + i\epsilon)|p'\rangle$$

converges in the ordinary sense. Instead of Eq. (14) we consider its partial wave projections in the momentum representation

$$\langle p|T_l(E + i\epsilon)|p'\rangle = V_l(p,p') + \int_0^\infty \frac{V_l(p,p'')\langle p''|T_l(E + i\epsilon)|p'\rangle}{E - p''^2 + i\epsilon} p''^2 dp'', \tag{16}$$

where

$$V_l(p,p') = V_{S,l}(p,p') + V_{C,l}(p,p'). \tag{17}$$

In Eq. (17) $V_{S,l}$ is defined by Eq. (13) and $V_{C,l}$ is the l th partial wave of the Coulomb potential (10) in the momentum representation

$$V_{C,l}(p,p') = \frac{2k\gamma}{\pi pp'} Q_l\left(\frac{p^2 + p'^2}{2pp'}\right), \tag{18}$$

$Q_l(x)$ being the Legendre function of the second kind.

If $p'^2 \neq E \neq p^2$ the T matrix in Eq. (16) remains off-shell if $\epsilon \to 0$, thus the limit $\epsilon \to 0$ can be taken. From Zorbas' renormalization factor[18] and especially from the analytic expressions[10,11] for the singularities of the partial waves of the off-shell Coulomb and Coulomb-type T matrix it is seen that the solution of Eq. (16) with $\epsilon \to 0$ behaves for $p'^2 \neq E$ and $p^2 \to E$ like

$$\mathrm{const} \times (p^2 - E - i0)^{i\gamma}. \tag{19}$$

[See also Eq. (30) of Sec. 3.] The existence of this singularity is caused by the nonintegrable coincidence of singularities $\ln|E - p''^2|(E - p''^2 + i0)^{-1}$, which appears in the integrand on the RHS of Eq. (16) if $p^2 = E$. The first factor originates in the Coulomb potential (18) if $p^2 = E$ and the second is the free Green's function. This coincidence of singularities, which is characteristic of the Coulomb LS equation, has been discussed already by Veselova.[2] In Sec. 3 we show that the singularity (19) has the property of being conserved by the Lippmann–Schwinger operator $V_l G_0$, where by V_l we mean the l th partial wave projection of the potential V defined by Eqs. (13), (17), and (18). This basic property of the operator $V_l G_0$ is stated in Theorem 1 of Sec. 2. We do not give a proof of this theorem as it is a rather simple consequence of an analogous theorem for the operator $T_l G_0$ proven elsewhere.[10,12] Besides, the simple analytic expression (18) for the Coulomb potential makes it possible to verify Theorem 1 directly. Theorem 1 of Sec. 2 is used in Sec. 3 not only to prove in a way independent of the previous results[10,11] the form (19) of the on-shell singularity of the off-shell T matrix $\langle p|T_l(E + i0)|p'\rangle$ ($p'^2 \neq E$) but also to find the singular part up to terms of the order $o[(p^2 - E)^n]$, where $n = 0, 1, \cdots$, is, in principle, arbitrary. To fully exploit this possibility we introduce in Sec. 2 an alternative stricter requirement on the partial waves of the short-range potential $V_{S,l}(p,p')$. If we require the function $V_{S,l}(p,p')$ to be several times Hölder derivable in both variables p,p' (potentials that fall off exponentially in the coordinate representation are analytic) we are able to split the solution $\langle p|T_l(E + i0)|p'\rangle$ of the LS equation (16) into two terms. In the first one the dependence on the variable p is singular but is given in explicit form, the second term is several times Hölder derivable and is the solution of a nonsingular integral equation. However, even with only a Hölder continuous potential $V_{S,l}(p,p')$ the integral equation for the regular term is free of nonintegrable singularities.

In the case when the initial momentum p' lies on the energy shell $p'^2 = E$ the situation seems to be much more complicated at first sight. First of all for $p' = \sqrt{E}$, if we were to take the limit $\epsilon \to 0$ we should renormalize the T matrix according to the formula[18]

$$\langle p|T_l(E+i0)|k_0\rangle_{\text{h.s.}} = \lim_{\epsilon\to 0} \Gamma^{-1}(1-i\gamma)\left(\frac{\epsilon}{4E}\right)^{-\gamma}\langle p|T_l(E+i\epsilon)|k_0\rangle, \quad (20)$$

where

$$\langle p|T_l(E+i0)|k_0\rangle_{\text{h.s.}} = \langle p|V_l|k_0,l+\rangle$$

is the correct partial wave of the half-shell Coulomb-type T matrix, $\langle r|k_0,l+\rangle$ being the partial wave of the Coulomb-type wavefunction [in the pure Coulomb case we have, e.g.,[19]

$$\langle r|k_0,l+\rangle_c = \sqrt{\frac{2}{\pi}}\frac{\Gamma(l+1+i\gamma)}{(2l+1)!}e^{-(1/2)\pi\gamma}(2k_0r)^l e^{ik_0r} \times {}_1F_1(l+1+i\gamma,2l+2,-2ik_0r)].$$

It is immediately seen that the renormalized limit (20) cannot be performed in Eq. (16) if we take $p' = k_0$ as the inhomogeneous term does not contain the singularity $\epsilon^{i\gamma}$.

On the other hand, if we try to write down the LS equation

$$\langle p|T_l(E+i0)|k_0\rangle_{\text{h.s.}} = V_l(p,k_0) + \int_0^\infty \frac{V_l(p,p')\langle p'|T_l(E+i0)|k_0\rangle_{\text{h.s.}}}{(E-p'^2+i0)}p'^2dp' \quad (21)$$

we find that the integral on the RHS of Eq. (21) is not defined. To see this it is sufficient to write the expression for the l th partial wave of the half-shell pure Coulomb T matrix[11,20]:

$$\langle p|T_{C,l}(E+i0)|k_0\rangle_{\text{h.s.}} = \frac{i}{\pi p}e^{-(1/2)\pi\gamma}\left[\Gamma(1-i\gamma)\frac{\Gamma(l+1+i\gamma)}{\Gamma(l+1-i\gamma)}\binom{l+i\gamma}{l}^{-1} \times P_l^{(i\gamma,-i\gamma)}\left(\frac{p^2+k_0^2}{2pk_0}\right)\frac{(p^2-E-i0)^{i\gamma}}{(p+k_0)^{2i\gamma}} - \Gamma(1+i\gamma)\binom{l-i\gamma}{l}^{-1}P_l^{(-i\gamma,i\gamma)}\left(\frac{p^2+k_0^2}{2pk_0}\right) \times\frac{(p+k_0)^{2i\gamma}}{(p^2-E+i0)^{i\gamma}}\right],$$

where $P_l^{(\alpha,\beta)}(x)$ are the Jacobi polynomials.[21] The second term in the square brackets contains the singularity

$$(p^2-E+i0)^{-i\gamma},$$

which multiplied by the free Green's function $(p^2-E-i0)^{-1}$ is nonintegrable and undefined even as a distribution. This is the reason why the LS equation (21) is undefined.[1]

To treat the LS equation (16) correctly with $p' = k_0$ we have to be careful in taking the limit $\epsilon \to 0$. From the analytic expressions for the singularities of the pure Coulomb and Coulomb-type T matrix[10,11] it can be seen that $\langle p|T_l(E+i\epsilon)|k_0\rangle$ contains the singularity

$$\text{const}\times\left(\frac{i\epsilon}{p^2-E-i\epsilon}\right)^{i\gamma}. \quad (22)$$

In Sec. 2 (Theorem 2) we give the way this singularity is transformed by the LS operator V_lG_0. It appears that the singularity (22) is again conserved. Moreover, a term reproducing the potential $V_l(p,k_0)$ (up to some factor) arises. In Sec. 4 we prove that the off-shell T matrix $\langle p|T_l(E+i\epsilon)|k_0\rangle$ contains the singularity (22) in the neighborhood of the energy shell independently of the previous results[10,11] by showing that a suitably chosen expression containing this singularity satisfies (in the limit $\epsilon \to 0$) Eq. (16) with $p' = k_0$. As we shall see the singularity (22) makes the potential $V_l(p,k_0)$ reappear in the expression for the integral

$$\int_0^\infty \frac{V_l(p,p')\langle p'|T_l(E+i\epsilon)|k_0\rangle}{E-p'^2+i\epsilon}p'^2dp'$$

with the opposite sign. This cancels the inhomogeneous term of Eq. (16) and settles the problem of taking the renormalized limit (20) of the LS equation. The solution of Eq. (16) with $p' = k_0$ is sought in the form of a sum of three terms, the term containing the singularity (22), a term with the singularity (19), and a regular term. The dependence on the variable p of both singular terms can again be determined analytically, the regular term is the solution of the same nonsingular equation as in Sec. 3 with just a different inhomogeneous term.

In Sec. 5 we briefly describe the way physical (on-shell) values are obtained from the solutions of the modified equations. We also discuss the numerical applicability of the new nonsingular equations.

2. LIPPMANN–SCHWINGER OPERATOR V_lG_0

Let us briefly review the consequences of our study of the properties of the operator T_lG_0.[10,12] All results concerning this operator are valid for V_lG_0 if taken in the Born limit. We shall state them in the form of two theorems.

For this purpose we introduce Banach spaces of Hölder functions defined on $(0,\infty)$ $\mathscr{B}_{\theta,\alpha,\mu}(0,\infty)$ as spaces of functions with finite norms:

$$\|\Phi\|_{\theta,\alpha,\mu} = \sup_{\substack{p\in(0,\infty)\\ |h|<1}}(1+p)^\theta\frac{p^\alpha}{1+p^\alpha}\left[|\Phi(p)| + \frac{|\Phi(p+h)-\Phi(p)|}{|h|^\mu}\right],$$

with some $\alpha \geqslant 0$, $\theta \geqslant 0$, and $\mu \in (0,1]$.

The short-range part of the potential V_S will be supposed to satisfy either of the following assumptions.

Assumption 1:

$$V_{S,l}(p,p') \in \mathscr{B}_{\theta_1,\alpha_1,\mu_0}(0,\infty)\times\mathscr{B}_{\theta_2,\alpha_2,\mu_0}(0,\infty),$$

where $\mu_0 \in (0,1]$ and $\theta_1,\theta_2,\alpha_1,\alpha_2$ are arbitrary nonnegative real numbers satisfying

$$\theta_1+\theta_2 = 1+\theta_0, \quad \theta_0\in(0,1); \quad \alpha_1+\alpha_2 = 2-\mu_0.$$

Assumption 2: $V_{S,l}(p,p')$ satisfies Assumption 1 and moreover

$$V_{S,l}(p,p') \in C^{(m)}\times C^{(m)}$$

with Hölder mth derivatives.

It is easily seen[10] that the requirements of smoothness of the potential in momentum representation are equivalent to

restrictions on the asymptotic behavior in coordinate representation. Thus, e.g., Assumption 1 restricts the class of admissible potentials V_S to such that behave for $r\to\infty$ roughly like $r^{-1-\mu_0}$. It has to be pointed out that Assumptions 1 and 2 admit both local and nonlocal potentials.

Theorem 1: Let $\Phi\in\mathscr{B}_{\theta,\alpha,\mu}(0,\infty)$ $(\alpha\in[0,3)$, $\theta\in(0,2)$, $\mu\in(0,1])$, and suppose Assumption 1 holds. Then

$$-\int_0^\infty V_l(p,p')(p'^2-E-i0)^{i\gamma-1}\Phi(p')p'^2dp'$$
$$=\psi_l[\Phi](p,E+i0,0,\gamma)(p^2-E-i0)^{i\gamma}$$
$$+\tau_l[\Phi](p,E+i0,\gamma), \tag{23}$$

where ψ_l is an analytically defined functional (see the Appendix) such that the function $\psi_l[\Phi]$ is Hölder with index μ in the variable p and

$$\psi_l[\Phi](k_0,E+i0,0,\gamma)=\Phi(k_0). \tag{24}$$

The function $\tau_l[\Phi]$ is Hölder with any index $\mu'<\min(\mu,\mu_0)$. For the whole integral (23) the following estimate holds:

$$\left|\int_0^\infty V_l(p,p')(p'^2-E-i0)^{i\gamma-1}\Phi(p')p'^2dp'\right|$$
$$\leqslant C(1+p^{-\alpha'})(1+p)^{-\theta'}, \tag{25}$$

where

$$\alpha'\in[0,2-\mu_0]\wedge\alpha'>\alpha-1;$$
$$\theta'\in[0,2]\wedge\theta'<1+\theta\wedge\theta'<\theta+\theta_0. \tag{26}$$

If Φ is n times Hölder derivable and V_S satisfies Assumption 2 then $\psi_l[\Phi]$ *is* n times Hölder derivable and $\tau_l[\Phi]$ is $\min(m,n)$ times Hölder derivable.

Theorem 2: Let Φ be the same as in Theorem 1. Then

$$\int_0^\infty V_l(p,p')\left(\frac{i\epsilon}{p'^2-E-i\epsilon}\right)^{i\gamma}(p'^2-E-i\epsilon)^{-1}\Phi(p')p'^2dp'$$
$$=-\frac{1}{2}\frac{e^{\pi\gamma}-e^{-\pi\gamma}}{i\gamma}k_0\Phi(k_0)V_l(p,k_0)$$
$$-\psi_l[\Phi](p,E+i0,0,-\gamma)\left(\frac{i\epsilon}{p^2-E-i\epsilon}\right)^{i\gamma}$$
$$+\epsilon^{i\gamma}\sigma_l[\Phi](p,E+i0,\gamma)+o_\epsilon(1), \tag{27}$$

where ψ_l is the same as in Theorem 1 but with opposite sign for γ and $\sigma_l[\Phi]$ is, under Assumption 1, Hölder with any index $\mu'<\min(\mu,\mu_0)$ in the variable p. Under Assumption 2 and if Φ is n times Hölder derivable then $\sigma_l[\Phi]$ is $\min(m,n)$ times Hölder derivable. The sum of the last two terms on the RHS of Eq. (27) satisfies the estimate (25) and (26).

3. OFF-SHELL T MATRIX

In this section we shall derive a nonsingular equation yielding the T matrix fully off the energy shell. It means that in the matrix element $\langle p|T_l(E+i0)|p'\rangle$ the initial momentum $p'\neq k_0$ and there exists a neighborhood (K_1,K_2) of the energy shell such that

$$0<K_1<k_0<K_2<\infty \tag{28}$$

and

$$p'\notin[K_1,K_2]. \tag{29}$$

We shall proceed as follows. We shall first formulate an ansatz on the singular behavior of the T-matrix element $\langle p|T_l(E+i0)|p'\rangle$ in the neighborhood of $p=k_0$. Then the ansatz will be proven by requiring that the T matrix satisfy the Lippmann–Schwinger equation. The proof will automatically yield equation for the nonsingular part of the T matrix on (K_1,K_2). Finally, the equation will be completed for the whole interval $(0,\infty)$ and given a compact form.

Our ansatz will be that on (K_1,K_2) the off-shell Coulomb-like T matrix can be written in the form

$$\langle p|T_l(E+i0)|p'\rangle$$
$$=a_l(p,E+i0,i\gamma)(p^2-E-i0)^{i\gamma}\tilde{t}_{1,l}(p',E+i0)$$
$$+t_{2,l}(p,p',E+i0), \tag{30}$$

where $a_l(p,E+i0,i\gamma)$ and $t_{2,l}(p,p',E+i0)$ are some regular, i.e., at least Hölder, functions in the variable p and $\tilde{t}_{1,l}(p',E+i0)$ is independent on p. [It is easily seen[10,11] that Eq. (30) actually holds but we point out that it will be proven independently using only the properties of the operator V_lG_0 stated in Theorems 1 and 2 of Sec. 2].

The requirement that $\langle p|T_l(E+i0)|p'\rangle$ satisfy the LS equation yields the following.

Theorem 3: For $p\in(K_1,K_2)$ there exists a polynomial of nth degree in the variable p^2-E $a_l^{(n)}(p,E+i0,i\gamma)$ satisfying

$$a_l^{(n)}(k_0,E+i0,i\gamma)=1 \tag{31}$$

and $t_{1,l}(p',E+i0)$ independent of p such that for $p\in(K_1,K_2)$

$$\langle p|T_l(E+i0)|p'\rangle$$
$$=e^{-\pi\gamma}\Gamma(1-i\gamma)(4k_0^2)^{i\gamma}a_l^{(n)}(p,E+i0,i\gamma)(p^2-E-i0)^{i\gamma}$$
$$\times t_{1,l}(p',E+i0)+t_{2,l}^{(n)}(p,p',E+i0) \tag{32}$$

is a solution of the Lippmann-Schwinger equation

$$\langle p|T_l(E+i0)|p'\rangle=V_l(p,p')$$
$$+\int_0^\infty\frac{V_l(p,p'')\langle p''|T_l(E+i0)|p'\rangle}{E+i0-p''^2}p''^2dp'', \tag{33}$$

and that under Assumptions 1 and 2 the function $t_{2,l}(p,p',E+i0)$ is Hölder with an arbitrary index $\mu<\mu_0$ or $\min(m,n)$ times Hölder derivable in the variable p, respectively, and is zero for $p=k_0$:

$$t_{2,l}(k_0,p',E+i0)=0. \tag{34}$$

Proof: For $\epsilon>0$ the LS equation (16) holds for all values of the initial and final momenta.[22] If $p\neq k_0$ and $p'\neq k_0$ the limit for $\epsilon\to0$ exists but has to be taken after the integration. We shall therefore take $\epsilon\to0$ in Eq. (16) but we have to understand this limit in the distributional sense. To see if the expression (30) can satisfy the LS equation we shall substitute it into Eq. (33):

$$(p^2-E-i0)^{i\gamma}a_l(p,E+i0,i\gamma)\tilde{t}_{1,l}(p',E+i0)$$
$$+t_{2,l}(p,p',E+i0)=V_l(p,p')$$
$$+\left(\int_0^{K_1}\int_{K_2}^{\infty}\right)V_l(p,p'')(E-p''^2+i0)^{-1}$$
$$\times\langle p''|T_l(E+i0)|p'\rangle p''^2dp''$$
$$+\int_{K_1}^{K_2}V_l(p,p'')(E-p''^2+i0)^{-1}t_{2,l}(p'',p',E+i0)p''^2dp''$$
$$-\left(\int_{K_1}^{K_2}V_l(p,p'')(p''^2-E-i0)^{i\gamma-1}\right.$$
$$\left.\times a_l(p,E+i0,i\gamma)p''^2dp''\right)\tilde{t}_{1,l}(p',E+i0). \tag{35}$$

If we succeed in finding $a_l(p,E+i0,i\gamma)$ and $\tilde{t}_{1,l}(p,E+i0)$ as stated by the theorem the second integral on the RHS will be nonsingular as well as the first one, in which the on-shell singularities cannot be reached. We shall therefore consider in the first place the last term on the RHS of Eq. (35). Theorem 1 yields

$$-\int_{K_1}^{K_2}V_l(p,p'')(p''^2-E-i0)^{i\gamma-1}a_l(p'',E+i0,\gamma)p''^2dp''$$
$$=\tau_l[a_l](p,E+i0,K_1,K_2)+\psi_l[a_l](p,E+i0,0,\gamma)$$
$$\times(p^2-E-i0)^{i\gamma}, \tag{36}$$

where $\tau_l[a_l]$ is an at least Hölder function assuming a_l is also at least Hölder. The most singular term is thus the second one on the RHS of Eq. (36). After substitution of Eq. (36) into Eq. (35) the singularities proportional to $(p^2-E-i0)^{i\gamma}$ have to cancel if Eq. (35) is to be satisfied. We shall require that the equation

$$a_l(p,E+i0,i\gamma)=\psi_l[a_l](p,E+i0,0,\gamma) \tag{37}$$

be fulfilled at least up to some (arbitrary) nth degree of the Taylor expansion of both sides in the variable p^2-E. If we denote

$$f^{(n)}(p)\equiv[f(p)]^{(n)}=\sum_{j=0}^{n}\frac{1}{j!}\frac{d^jf}{(dp^2)^j}(E)(p^2-E)^j,$$

this means that we seek the solution of the equation

$$a_l^{(n)}(p,E+i0,i\gamma)=[\psi_l[a_l^{(n)}](p,E+i0,0,\gamma)]^{(n)}, \tag{38}$$

satisfying the condition (31). In one of our previous works[12] we have shown (see the Appendix) that the solution of Eq. (38) exists and, if Eq. (31) holds, is unique for any n. If we denote by $a_{l,j}(E,i\gamma)$ the coefficients of the polynomial

$$a_l^{(n)}(p,E+i0,i\gamma)=\sum_{j=0}^{n}a_{l,j}(E,i\gamma)(p^2-E)^j,$$

we can say more exactly that there exists a recursion formula relating all coefficients $a_{l,j}(E,i\gamma)$ for $j>0$ to the zeroth order coefficient $a_{l,0}$, which is chosen equal to one. The formula as well as its application to the calculation of the first-order coefficient are written out in the Appendix. We find, e.g., that

$$a_{l,1}(E,i\gamma)=-\frac{i\gamma+1}{2E}, \tag{39}$$

which is in accordance with the exact expression[11] for a_l.

Having solved Eq. (38) we shall denote

$$u_l^{(n)}(p,E+i0;i\gamma,K_1,K_2)=\tau_l[a_l^{(n)}](p,E+i0;K_1,K_2)$$
$$+R_n[\psi_l[a_l^{(n)}]](p,E+i0,0,\gamma)(p^2-E-i0)^{i\gamma}$$
$$=-\int_{K_1}^{K_2}V_l(p,p'')$$
$$\times(p''^2-E-i0)^{i\gamma-1}a_l^{(n)}(p'',E+i0,i\gamma)p''^2dp''$$
$$-a_l^{(n)}(p,E+i0,i\gamma)(p^2-E-i0)^{i\gamma} \tag{40}$$

and

$$t_{2,l}^{(n)}(p,p',E+i0)=t_{2,l}(p,p',E+i0)$$
$$+R_n[a_l](p,E+i0,i\gamma)(p^2-E-i0)^{i\gamma}\tilde{t}_{1,l}(p',E+i0), \tag{41}$$

where $R_n[f]$ stands for

$$R_n[f]=f-f^{(n)}.$$

On account of Eqs. (40) and (41) we can rewrite Eq. (35) in the following form:

$$t_{2,l}^{(n)}(p,p',E+i0)=V_l(p,p')+\left(\int_0^{K_1}+\int_{K_2}^{\infty}\right)$$
$$\times\frac{V_l(p,p'')\langle p''|T_l(E+i0)|p'\rangle}{E-p''^2+i0}p''^2dp''$$
$$+\int_{K_1}^{K_2}V_l(p,p'')(E-p''^2+i0)^{-1}t_{2,l}^{(n)}(p'',p',E+i0)p''^2dp''$$
$$+u_l^{(n)}(p,E+i0;i\gamma,K_1,K_2)\tilde{t}_{1,l}(p',E+i0). \tag{42}$$

If Eq. (34) is to be satisfied then the RHS of Eq. (42) has to be zero for $p=k_0$. In the Appendix we show that the constants K_1 and K_2 can always be chosen so that

$$u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)\neq 0. \tag{43}$$

By (A3) it is possible to take

$$\tilde{t}_{1,l}(p',E+i0)\equiv e^{-\pi\gamma}\Gamma(1-i\gamma)(2k_0)^{2i\gamma}t_{1,l}(p',E+i0)$$
$$=-\frac{V_l(k_0,p')}{u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)}-\frac{1}{u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)}\left[\left(\int_0^{K_1}+\int_{K_2}^{\infty}\right)V_l(k_0,p'')\right.$$
$$\left.\times(E-p''^2+i0)^{-1}\langle p''|T_l(E+i0)|p'\rangle p''^2dp''+\int_{K_1}^{K_2}V_l(k_0,p'')(E-p''^2+i0)^{-1}t_{2,l}^{(n)}(p'',p',E+i0)p''^2dp''\right]. \tag{44}$$

To complete the proof let us observe that for $p\in(K_1,K_2)$ the first two terms on the RHS of Eq. (42) are under Assumption 1 or 2 Hölder with index μ_0 or m times Hölder derivable, respectively. By Theorem 1 (with $\gamma=0$) the operator $V_l G_0$ with $V_{S,l}$ satisfying Assumption 1 or 2 maps the space of Hölder functions with indices $\mu<\mu_0$ into itself or the space of n times Hölder derivable functions into the space of $\min(m,n)$ times Hölder derivable functions, respectively. This statement can be applied to the third term on the RHS of Eq. (42) which, due to the fact that the last term is n times Hölder derivable, yields the properties of the function $t^{(n)}_{2,l}(p,p',E+i0)$ stated by the theorem.

To get a closed set of equations, Eqs. (42) and (44) must be completed by the equation for $\langle p|T_l(E+i0)|p'\rangle$ with $p\notin[K_1,K_2]$. This equation is easily obtained by substitution of Eq. (42) into Eq. (33) now with $p\notin[K_1,K_2]$. The resulting equations can be given a compact form similar to that of the original LS equation. Let us denote

$$\begin{aligned}
&\mathscr{T}^{(n)}_l(p,p',E+i0)\\
&\quad=\chi_{K_1,K_2}(p)t^{(n)}_{2,l}(p,p',E+i0)\\
&\quad+\left[1-\chi_{K_1,K_2}(p)\right]\langle p|T_l(E+i0)|p'\rangle, \qquad (45)\\
&\mathscr{U}^{(n)}_l(p,E+i0)\\
&\quad=\chi_{K_1,K_2}(p)u^{(n)}_l(p,E+i0;i\gamma,K_1,K_2)\\
&\quad+\left[1-\chi_{K_1,K_2}(p)\right]\int_{K_1}^{K_2}V_l(p,p')(p'^2-E-i0)^{i\gamma-1}\\
&\quad\times a^{(n)}_l(p',E+i0,i\gamma)p'^2dp' \qquad (46)
\end{aligned}$$

where $\chi_{K_1,K_2}(p)=1$ for $p\in[K_1,K_2]$ and is zero elsewhere and

$$\mathscr{V}_l(p,p',E+i0)=V_l(p,p')-\frac{\mathscr{U}^{(n)}_l(p,E+i0)V_l(k_0,p')}{u^{(n)}_l(k_0,E+i0;i\gamma,K_1,K_2)}. \qquad (47)$$

We arrive at the equation

$$\begin{aligned}
\mathscr{T}^{(n)}_l(p,p',E+i0)=\mathscr{V}^{(n)}_l(p,p',E+i0)\\
+\int_0^\infty\mathscr{V}^{(n)}_l(p,p'',E+i0)(E-p''^2)^{-1}\\
\times\mathscr{T}^{(n)}_l(p'',p',E+i0)p''^2dp''. \qquad (48)
\end{aligned}$$

The conditions for the behavior of the potential $V_{S,l}$ on both ends of the interval $(0,\infty)$ have been chosen so that the compactness of the LS operator is ensured.[10] It is easily seen that the Coulomb potential satisfies these conditions too. The noncompactness of the Coulomb-type Lippmann–Schwinger operator results in the singularity of the solution $\langle p|T_l(E+i0)|p'\rangle$ on the energy shell $p^2=E$. In the neighborhood of the energy shell the integrand (48) has been modified by extracting the singularity of the solution in such a way that $(E-p''^2)^{-1}\mathscr{T}^{(n)}_l(p'',p',E+i0)$ has at most the integrable singularity $|p''^2-E|^{\mu-1}$. This singularity remains integrable also if combined with the logarithmic singularity of the modified potential $\mathscr{V}^{(n)}_l(p,p'',E+i0)$, which originates in the pure Coulomb potential (18). As a result Eq. (48) is for any $n\geqslant 0$ manifestly nonsingular.

4. HALF-SHELL T MATRIX

In the previous section we excluded the possibility of the initial momentum reaching its on-shell value. This was because in this case additional singularities appear. However, it appears that the method of Sec. 3 can be applied again with just slight modifications. It is well known[22] that the limit for $\epsilon\to 0$ of the matrix element $\langle p|T_l(E+i\epsilon)|k_0\rangle$ does not exist. Therefore, it is impossible to set $\epsilon\to 0$ from the very beginning. Another reason why we have to take $\epsilon>0$ is that the limit of the Lippmann–Schwinger equation (16) does not exist even after renormalization. We shall formulate, in the same way as in Sec. 3, an ansatz on the form of the T matrix in the neighborhood of the energy shell. This time the energy shell is approached in two ways, for $\epsilon\to 0$ and $p\to k$, and both limits are singular. Subsequently, the ansatz will be proven by substitution into the LS equation. In spite of all the mentioned differences the proof will again yield directly a regular equation for the nonsingular part of the renormalized half-shell T matrix.

In the limit $\epsilon\to 0$ the oscillating factor [see Eq. (20)] $\Gamma(1-i\gamma)(\epsilon/4k_0^2)^{i\gamma}$ appears in the matrix element $\langle p|T_l(E+i\epsilon)|k_0\rangle$.[18,22] Taking this fact into account we shall seek the solution of Eq. (16) with $p=k_0$ in the form

$$\begin{aligned}
\langle p|T_l(E+i\epsilon)|k_0\rangle=&-(i/\pi k_0)\Gamma(1+i\gamma)\\
&\times\Gamma(1-i\gamma)b_l(p,E+i0,i\gamma)\\
&\times\left(\frac{i\epsilon}{p^2-E-i\epsilon}\right)^{i\gamma}+\frac{i}{\pi k_0}\Gamma^2(1-i\gamma)c_l(p,E+i0,i\gamma)\\
&\times s_l(E)\frac{(p^2-E-i0)^{i\gamma}}{(2k_0)^{2i\gamma}}\epsilon^{i\gamma}\\
&+r_l(p,E+i0)\Gamma(1-i\gamma)(2k_0)^{-2i\gamma}\epsilon^{i\gamma}+o_\epsilon(1) \qquad (49)
\end{aligned}$$

for $p\in(K_1,K_2)$, the constants K_1,K_2 being chosen according to (28). The functions b_l,c_l, and r_l are assumed to be at least Hölder. The constant factors are introduced only for convenience (see Sec. 5). We shall prove the following theorem.

Theorem 4: For $p\in(K_1,K_2)$ and $\epsilon\to 0$

$$\begin{aligned}
&\langle p|T_l(E+i\epsilon)|k_0\rangle\\
&\quad=-\frac{i}{\pi k_0}\Gamma(1+i\gamma)\Gamma(1-i\gamma)a^{(n)}_l(p,E+i0,-i\gamma)\\
&\quad\times\left(\frac{i\epsilon}{p^2-E-i\epsilon}\right)^{i\gamma}+\frac{i}{\pi k_0}\Gamma^2(1-i\gamma)\\
&\quad\times a^{(n)}_l(p,E+i0,i\gamma)s_l(E)\frac{(p^2-E-i0)^{i\gamma}}{(2k_0)^{2i\gamma}}\epsilon^{i\gamma}\\
&\quad+r^{(n)}_l(p,E+i0)\Gamma(1-i\gamma)(2k_0)^{-2i\gamma}\epsilon^{i\gamma}+o_\epsilon(1) \qquad (50)
\end{aligned}$$

satisfies Eq. (16) with $p'=k_0$ and $\epsilon\to 0$. In Eq. (50) $a^{(n)}_l(p,E+i0,\pm i\gamma)$ is the same as in Sec. 3 with the opposite and the same sign of γ in the first and second term on the RHS, respectively. Under Assumptions 1 and 2 the function $r^{(n)}_l(p,E+i0)$ is Hölder with an arbitrary index $\mu<\mu_0$, $\min(m,n)$ times Hölder derivable, respectively, and such that

$$r^{(n)}_l(k_0,E+i0)=0. \qquad (51)$$

Proof: In the same way as when proving Theorem 3 we shall substitute the expression (49) into Eq. (16) (with $p'=k_0$). Without writing the obtained equation in detail we

note that the following two relations will be needed:

$$-\int_{K_1}^{K_2} V_l(p,p'')(p''^2-E-i0)^{i\gamma} \times c_l(p'',E+i0,i\gamma)p''^2dp'' = \psi_l[c_l](p,E+i0,0,\gamma)(p^2-E-i0)^{i\gamma} + \tau_l[c_l](p,E+i0;K_1,K_2) \tag{52}$$

and

$$\int_{K_1}^{K_2} V_l(p,p'')\left(\frac{i\epsilon}{p''^2-E-i\epsilon}\right)^{i\gamma}(p''^2-E-i\epsilon)^{-1} \times b_l(p'',E+i0,i\gamma)p''^2dp'' = -\frac{e^{\pi\gamma}-e^{-\pi\gamma}}{2i\gamma}k_0 b_l(k_0,E+i0,i\gamma)V_l(p,k_0) - \psi_l[b_l](p,E+i0,0,-\gamma) \times\left(\frac{i\epsilon}{p^2-E-i\epsilon}\right)^{i\gamma}+\sigma_l[b_l](p,E+i0;K_1,K_2)\epsilon^{i\gamma} + o_\epsilon(1). \tag{53}$$

Equation (52) results from Theorem 1 and Eq. (53) from Theorem 2. Moreover, we shall introduce the half-shell T matrix by[10,11]

$$\langle p|T_l(E+i\epsilon)|k_0\rangle = \Gamma(1-i\gamma)\left(\frac{\epsilon}{4k_0^2}\right)^{i\gamma}\langle p|T_l(E+i0)|k_0\rangle_{\text{h.s.}} + o_\epsilon(1). \tag{54}$$

In the obtained equation we shall require that the terms containing the same type of singularity cancel. First of all two terms on RHS do not contain the singularity $\epsilon^{i\gamma}$, which makes the renormalization of the RHS impossible. These terms are the potential term of the original Eq. (16) and the term arising from the first term on the RHS of Eq. (53). In order that they cancel we have to take

$$b_l(k_0,E+i0,i\gamma)=1. \tag{55}$$

Furthermore, we shall require that

$$b_l(p,E+i0,i\gamma)=\psi_l[b_l](p,E+i0,0,-\gamma) \tag{56}$$

and

$$c_l(p,E+i0,i\gamma)=\psi_l[c_l](p,E+i0,0,\gamma). \tag{57}$$

Taking into account that in the expression (49) c_l forms a product with a still undetermined factor $s_l(E)$ we can choose $c_l(k_0,E+i0,i\gamma)$ arbitrarily. We shall set

$$c_l(k_0,E+i0,i\gamma)=1. \tag{58}$$

Equations (55)–(58) imply

$$b_l(p,E+i0,i\gamma)=a_l(p,E+i0,-i\gamma)$$

and

$$c_l(p,E+i0,i\gamma)=a_l(p,E+i0,i\gamma).$$

As before we shall solve Eqs. (56) and (57) only up to an nth degree of the Taylor expansion in p^2-E, i.e., Eq. (38) will hold. Introducing then

$$r_l^{(n)}(p,E+i0)=r_l(p,E+i0)-(i/\pi k_0)\Gamma(1+i\gamma) \times R_n[a_l](p,E+i0,-i\gamma) \times(p^2-E+i0)^{-i\gamma} + \frac{i}{\pi k_0}\Gamma(1-i\gamma)(2k_0)^{-2i\gamma}s_l(E) \times R_n[a_l](p,E+i0,i\gamma)(p^2-E-i0)^{i\gamma} \tag{59}$$

and

$$v_l^{(n)}(p,E+i0;i\gamma,K_1,K_2)=\sigma_l[a_l^{(n)}(\cdot,E+i0,-i\gamma)](p,E+i0;K_1,K_2)-R_n[\psi_l[a_l^{(n)}(p,E+i0,-i\gamma)]](p,E+i0,0,-\gamma) \times(p^2-E+i0)^{-i\gamma}, \tag{60}$$

we arrive, after multiplying by the renormalization factor $\Gamma^{-1}(1-i\gamma)(\epsilon/4k_0^2)^{-i\gamma}$ and taking the limit $\epsilon\to 0$, at the equation

$$r_l^{(n)}(p,E+i0)=\frac{i}{\pi k_0}\Gamma(1+i\gamma)(2k_0)^{2i\gamma}v_l^{(n)}(p,E+i0;i\gamma,K_1,K_2) + \frac{i}{\pi k_0}\Gamma(1-i\gamma)(2k_0)^{-2i\gamma}s_l(E)u_l^{(n)}(p,E+i0;i\gamma,K_1,K_2) + \left(\int_0^{K_1}+\int_{K_2}^{\infty}\right)V_l(p,p'')(E-p''^2+i0)^{-1}\langle p''|T_l(E+i0)|k_0\rangle_{\text{h.s.}}p''^2dp'' + \int_{K_1}^{K_2}V_l(p,p'')(E-p''^2+i0)^{-1}r_l^{(n)}(p'',E+i0)p''^2dp'', \tag{61}$$

where $u_l^{(n)}$ is defined by Eq. (40) and $\langle p|T_l(E+i0)|k_0\rangle_{\text{h.s}}$ is the Coulomb-like half-shell T matrix introduced by Eq. (54). In order that (51) be satisfied $s_l(E)$ has to be determined by the following relation:

$$\frac{i}{\pi k_0}\Gamma(1-i\gamma)(2k_0)^{-2i\gamma}s_l(E)=-\frac{i}{\pi k_0}\Gamma(1+i\gamma)(2k_0)^{2i\gamma} \times\frac{v_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)}{u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)}-\frac{1}{u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)}\left[\left(\int_0^{K_1}+\int_{K_2}^{K_1}\right)V_l(p,p'') \times(E-p''^2+i0)^{-1}\langle p''|T_l(E+i0)|k_0\rangle_{\text{h.s.}}p''^2dp'' + \int_{K_1}^{K_2}V_l(p,p'')(E-p''^2+i0)^{-1}r_l^{(n)}(p'',E+i0)p''^2dp''\right], \tag{62}$$

where K_1 and K_2 are again supposed to be chosen so that (43) holds.

As far as the dependence on the variable p is concerned Eqs. (61) and (42) differ only by their first terms on the RHS. The first term on the RHS of Eq. (61) is, however, Hölder with index $\mu < \mu_0$, $\min(m,n)$ times Hölder derivable, under Assumptions 1 and 2, respectively. As a result the concluding consideration of the proof of Theorem 3 applies to the present case too. This completes the proof.

Denoting

$$\mathscr{T}_l^{(n)}(p,k_0,E+i0)_{\text{h.s.}} = \chi_{K_1,K_2}(p)r_l^{(n)}(p,E+i0) + [1-\chi_{K_1,K_2}(p)]\langle p|T_l(E+i0)|k_0\rangle_{\text{h.s.}}, \tag{63}$$

$$\mathscr{R}_l^{(n)}(p,E+i0) = \frac{i}{\pi k_0}\Gamma(1+i\gamma)(2k_0)^{2i\gamma}\Big[\chi_{K_1K_2}(p)v_l^{(n)}(p,E+i0;i\gamma,K_1,K_2) + [1-\chi_{K_1,K_2}(p)]\int_{K_1}^{K_2} V_l(p,p'')(p''^2-E+i0)^{-i\gamma-1}a_l^{(n)}(p'',E+i0,-i\gamma)p''^2dp''\Big], \tag{64}$$

and

$$\mathscr{L}_l^{(n)}(p,E+i0) = \mathscr{R}_l^{(n)}(p,E+i0) - \frac{i}{\pi k_0}\Gamma(1+i\gamma)(2k_0)^{2i\gamma} \times v_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)\frac{\mathscr{U}_l^{(n)}(p,E+i0)}{u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)}, \tag{65}$$

with $\mathscr{U}_l^{(n)}$ and $u_l^{(n)}$ defined by Eqs. (40) and (46), we obtain the resulting integral equation in the form

$$\mathscr{T}_l^{(n)}(p,k_0,E+i0)_{\text{h.s.}} = \mathscr{L}_l^{(n)}(p,E+i0) + \int_0^\infty \mathscr{V}_l^{(n)}(p,p',E+i0)(E-p'^2)^{-1} \times \mathscr{T}_l^{(n)}(p',k_0,E+i0)_{\text{h.s.}}\, p'^2dp' \tag{66}$$

where $\mathscr{V}_l^{(n)}$ is given by Eq. (47).

Similarly as in the case of the ordinary Lippmann–Schwinger equation for short-range interactions the modified off- and half-shell equations (48) and (66) differ only by their inhomogeneous terms. Both inhomogeneous terms have the same properties and Eq. (66) is again manifestly nonsingular for any $n \geqslant 0$.

5. CONCLUSIONS

Starting from the Lippman–Schwinger equation which is, in the case of Coulomb-like interactions, highly singular, we have obtained in Secs. 3 and 4 a manifestly nonsingular equation yielding via Eqs. (32) and (45) and (50) and (63) the Coulomb-like off- and half-shell T matrix. This is guaranteed by Eqs. (34) and (51) together with the Hölder property of the solutions because then

$$\left|\frac{\mathscr{T}_l^{(n)}(p'',p',E+i0)}{p''^2-E}\right| \leqslant \mathscr{C}(p',E+i0)|p''^2-E|^{\mu-1}$$

for all $n \geqslant 0$ and the modified potential contains only the logarithmic Coulomb singularity.

For a derivable short-range potential V_S satisfying Assumption 2 the expression in the absolute value above can be made continuous and smooth, which greatly facilitates the numerical treatment. In this case the only singularity left in the integrand is the Coulomb logarithmic one. To treat this singularity, e.g., the subtraction method[23] can be used.

When no long-range interaction is present the Lippmann–Schwinger equation for the off-shell T matrix can be considered as a means of calculation of the half-shell T matrix and, in the same way, the equation for the half-shell T matrix yields, when the final momentum is set equal to its on-shell value, the on-shell T matrix. The same can be said about the equations obtained in Secs. 3 and 4 with the only difference that the transition to on-shell values involves the Coulomb renormalization.[10,11,18] Equations (31), (32), and (34) yield

$$_{\text{h.s.}}\langle k_0|T_l(E+i0)|p'\rangle = t_{1,l}(p',E+i0)$$

and Eqs. (31), (50), and (51) imply[11]

$$t_l(E) = is_l(E)/\pi k_0,$$

where $t_l(E)$ stands for the properly renormalized on-shell value of the Coulomb-like T matrix, i.e.,

$$s_l(E) = \exp[2i\delta_l(E)],$$

where $\delta_l(E)$ is the phase shift of the total Coulomb-type potential $V_C + V_S$. {After the limit $\epsilon \to 0$ has been taken in Eq. (50) the first term on the RHS of Eq. (50) contributes a nonrenormalizable singularity on the energy shell. However, for $p^2 \to E$ this term is independent of l [by Eq. (31)]. As the partial wave decomposition of the on-shell T matrix exists only in a distributional sense which ignores the singularity in the forward direction, and as terms of the partial waves that are independent of l contribute only just to the value of the nondecomposed T matrix in the forward direction, such terms are ignored and are to be left out. See Ref. 11 for a detailed discussion.}

The calculation of the half- and on-shell values is thus performed in analogy with the short-range case in two steps. First Eq. (48) or (66) is solved with a p variable and then the half- or on-shell matrix element is calculated by substituting the solution into Eq. (44) or (62), respectively.

ACKNOWLEDGMENT

The author would like to thank Dr. L. Trlifaj for helpful discussions and for reading the manuscript.

APPENDIX

In Theorem 1 we have introduced the functional ψ_l. It is defined by the following set of relations[12]:

$$\psi_l[\Phi](p,E+i0,0,\gamma)$$
$$=\frac{k_0}{p}\left[P_l\left(\frac{p^2+k_0^2}{2pk_0}\right)\varphi_l[\Phi](p,E+i0,0,\gamma)\right.$$
$$\left.-\frac{i\gamma l(l+1)}{8pk_0^3(1+i\gamma)}(p^2-E)^2\omega_l[\Phi](p,E+i0,0,\gamma)\right], \tag{A1}$$

$$\varphi_l[\Phi](p,E+i0,0,\gamma)=(p^2-E-i0)^{-i\gamma}\lim_{\beta\to 0_+} i\gamma$$
$$\times\int_E^{p^2}P_l\left(\frac{p'^2+k_0^2}{2p'k_0}\right)\Phi(p')(p'^2-E-i0)^{\beta+i\gamma-1}dp'^2, \tag{A2}$$

$$\frac{l(l+1)}{2(i\gamma+1)}(p^2-E-i0)^{i\gamma+1}\frac{1}{2k_0^2}\omega_l[\Phi](p,E+i0,0,\gamma)$$
$$=\int_E^{p^2}\mathcal{P}_l^{(1)}(p,p',k_0;0)\frac{1}{2p'k_0}(p'^2-E-i0)^{i\gamma}\Phi(p')dp'^2, \tag{A3}$$

where

$$\mathcal{P}_l^{(1)}(p,p',k_0;0)=\sum_{j=1}^{l}\frac{1}{j!}P_{l,j}\left(\frac{(p^2+k_0^2)(p'^2+k_0^2)}{4pp'k_0^2}\right)$$
$$\times\left(-\frac{(p^2-k_0^2)(p'^2-k_0^2)}{4pp'k_0^2}\right)^{j-1}, \tag{A4}$$

$P_{l,j}(x)$ standing for

$$P_{l,j}(x)=\left(\frac{d}{dx}\right)^j P_l(x),$$

$P_l(x)$ being the Legendre polynomial.

It is easy to show that if

$$\Phi(p)=\sum_{j=0}^{\infty}\Phi_j(E)(p^2-E)^j,$$

then

$$\psi_l[\Phi](p,E+i0,0,\gamma)=\sum_{j=0}^{\infty}d_j[\Phi](E,0,\gamma)(p^2-E)^j,$$

where

$$d_j[\Phi](E,0,\gamma)=\sum_{m=0}^{j}M_{jm}^{(l)}(E,0,\gamma)(p^2-E)^m.$$

$M_{jm}^{(l)}(E,0,\gamma)$ are the coefficients of the Taylor expansion

$$\psi_l[(\cdot^2-E)^m](p,E+i0,0,\gamma)=\sum_{j=m}^{\infty}M_{jm}^{(l)}(E,0,\gamma)(p^2-E)^j. \tag{A5}$$

Equation (38) then reduces to

$$a_{l,j}(E,i\gamma)=\sum_{m=0}^{j}M_{jm}^{(l)}(E,0,\gamma)a_{l,m}(E,i\gamma),\quad j=1,\ldots,n. \tag{A6}$$

Equations (A1)–(A4) can be used to calculate the coefficients $M_{jm}^{(l)}(E,0,\gamma)$ explicitly. It is especially easy to get

$$M_{jj}^{(l)}(E,0,\gamma)=\frac{i\gamma}{i\gamma+j}. \tag{A7}$$

This shows that for $j>0$ Eq. (A6) can be solved for $a_{l,j}$:

$$a_{l,j}(E,i\gamma)=(1-M_{jj}^{(l)}(E,0,\gamma)^{-1}\sum_{m=0}^{j-1}M_{jm}^{(l)}(E,0,\gamma)a_{l,m}(E,i\gamma). \tag{A8}$$

Let us calculate $a_{l,1}$ as an example. Assuming $a_{l,0}(E,i\gamma)=1$ we have from Eqs. (A6) and (A7)

$$a_{l,1}(E,i\gamma)=(i\gamma+1)M_{1,0}^{(l)}(E,0,\gamma). \tag{A9}$$

Therefore it is sufficient to find $M_{10}^{(l)}(E,0,i\gamma)$, i.e., to get the first degree Taylor expansion $[\psi_l[1](p,E+i0,0,\gamma)]^{(1)}$. By (A1) and (A2)

$$[\psi_l[1](p,E+i0,0,\gamma)]^{(1)}$$
$$=\left[\frac{k_0}{p}P_l\left(\frac{p^2+k_0^2}{2pk_0}\right)\varphi_l[1](p,E+i0,0,\gamma)\right]^{(1)}$$
$$=1-\frac{p^2-E}{2E},$$

i.e.,

$$M_{1,0}^{(l)}(E,0,\gamma)=-(2E)^{-1}. \tag{A10}$$

Equations (A9) and (A10) yield Eq. (39).

In Sec. 3 we assert that there exists K_1,K_2 satisfying (28) and (29) such that (43) holds. To prove it we shall use the definition (40):

$$u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)=\tau_l[a_l^{(n)}](k_0,E+i0;K_1,K_2). \tag{A11}$$

Assuming

$$|K_1-k_0|<\Delta,\quad |K_2-k_0|<\Delta,$$

where $\Delta\to 0$, we decompose the RHS into the contribution of the pure Coulomb part V_C and the short-range part V_S of the potential:

$$\tau_l[a_l^{(n)}]=\tau_{C,l}[a_l^{(n)}]+\tau_{S,l}[a_l^{(n)}].$$

For the short-range term it is easily seen that

$$\tau_{S,l}[a_l^{(n)}](k_0,E+i0;K_1,K_2)$$
$$=-\frac{k_0V_{S,l}(k_0,k_0)}{2i\gamma}[(K_2^2-E-i0)^{i\gamma}$$
$$-(K_1^2-E-i0)^{i\gamma}]+o_\Delta(1) \tag{A12}$$

holds. As far as the pure Coulomb term is concerned we can use the results of a previous paper[12] to obtain after some manipulations $[\psi(z)=\Gamma'(z)/\Gamma(z)]$

$$\tau_{C,l}[a_l^{(n)}](k_0,E+i0;K_1,K_2)$$
$$=-\frac{i}{\pi}[\psi(l+1)-\psi(1)-1/i\gamma-2\ln 2k_0]$$
$$\times[(K_2^2-E-i0)^{i\gamma}-(K_1^2-E-i0)^{i\gamma}]$$
$$-\frac{i}{\pi}[(K_2^2-E-i0)^{i\gamma}\ln(K_2^2-E-i0)$$
$$-(K_1^2-E-i0)^{i\gamma}\ln(K_1^2-E-i0)]$$
$$-(K_1^2-E-i0)^{i\gamma}+o_\Delta(1). \tag{A13}$$

After substitution of Eqs. (A12) and (A13) into Eq. (A11) we get three linearly independent terms, two of which are surely nonzero [the last two terms on the RHS of Eq. (A13)], if we consider them as functions of K_1,K_2. Hence $u_l^{(n)}(k_0,E+i0;i\gamma,K_1,K_2)$ cannot be identically zero as a function of K_1,K_2.

[1]E. Prugovečki and J. Zorbas, J. Math. Phys. **14**, 1398 (1973).
[2]A. M. Veselova, Theor. Math. Phys. **35**, 180 (1978).
[3]M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).
[4]J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[5]F. Gesztesy and B. Thaler, J. Phys. A **14**, 639 (1981).
[6]L. Trlifaj, Czech. J. Phys. B **31**, 969 (1981).
[7]H. van Haeringen and R. van Wageningen, J. Math. Phys. **16**, 1441 (1975).
[8]H. van Haeringen, J. Math. Phys. **18**, 927 (1977).
[9]H. van Haeringen and L. P. Kok, Phys. Lett. A **82**, 317 (1981).
[10]J. Dušek, Czech. J. Phys. B **32**, 1325 (1982).
[11]J. Dušek, Czech. J. Phys. B **31**, 941 (1981).
[12]J. Dušek, Czech. J. Phys. B **32**, 1195 (1982).
[13]S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960); R. A. Mapleton, J. Math. Phys. **2**, 482 (1961); J. Schwinger, *ibid.* **5**, 1606 (1964); L. Hostler, *ibid.* **5**, 1235 (1964); W. F. Ford, Phys. Rev. **133**, B1616 (1964).
[14]L. D. Faddeev, Trudy Mat. Inst. Acad. Nauk SSSR **69** (1963), Engl. transl.: Mathematical Aspects of the Three-Body Problem in Quantum Scattering Theory, Israel Program for Scientific Translation (Jerusalem, 1965).
[15]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. III: Scattering Theory* (Academic, New York, 1979), Chap. XI. 4.
[16]J. R. Taylor, Nuovo Cimento B **23**, 313 (1974).
[17]M. D. Semon and J. F. Taylor, Nuovo Cimento A **26**, 48 (1975).
[18]J. Zorbas, J. Math. Phys. **17**, 498 (1976).
[19]M. R. C. McDowell and J. P. Coleman, *Introduction to the Theory of Ion-Atom Collisions* (North-Holland, Amsterdam, 1970).
[20]H. van Haeringen, J. Math. Phys. **19**, 1379 (1978).
[21]H. Bateman and A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.
[22]J. Zorbas, J. Math. Phys. **18**, 1112 (1977).
[23]W. M. Kloet and J. A. Tjon, Ann. Phys. **79**, 407 (1973).

Theory of light cone cuts of null infinity[a)]

Carlos N. Kozameh and Ezra T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 1 March 1983; accepted for publication 3 June 1983)

Light-cone cuts of null infinity are defined to be the intersection of the light cone of an interior point x^a with the future null boundary of the space-time, i.e., $\mathscr{I}^+$. It is shown how from the knowledge of the set of light-cone cuts of $\mathscr{I}^+$, the interior (conformal) metric can be reconstructed. Furthermore, a differential equation defined only on $\mathscr{I}^+$ is proposed so that (1) the solution space (the parameters defining the set of solutions) is identified with or defines the space-time itself and (2) the solutions themselves yield the light-cone cuts which in turn give metrics conformally equivalent to vacuum solutions of the Einstein equations.

PACS numbers: 04.20. − q, 03.40. + x, 02.40. + m

I. INTRODUCTION

The usual description of differential geometry and general relativity is in terms of local fields, e.g., the metric, the curvature tensor, etc., which satisfy local differential equations as, for example, $R_{ab} = 0$. It is our purpose here to introduce a new nonlocal field, denoted by Z, from which the local fields can be derived.

Basically Z, though it is nonlocal, should be thought of as a function on the bundle of null directions on a space-time, i.e., Z should depend on the space-time points x^a and points on the sphere of null directions which we coordinatize by the complex stereographic coordinates $(\zeta, \bar{\zeta})$. We thus have $Z(x^a;\zeta, \bar{\zeta})$.

Our program can be divided into two parts. In Sec. II we will discuss the geometric meaning of Z and the relationship between Z and the local field g_{ab}. We make the claim that knowledge of the metric (or conformal metric) allows Z to be calculated and, conversely, knowledge of Z permits the calculation of the metric up to conformal factor. In Sec. III we wish to suggest field equations for Z which would be equivalent to the vacuum Einstein equations for g_{ab}.

For simplicity we will confine our discussions to the case of space-times which are asymptotically flat in future (or past) null directions.[1,2] (There is a formulation of our results for a local region of an arbitrary space-time.)

In order to introduce Z, we assume that we have a manifold M with boundary $\mathscr{I}^+$ with a given metric g_{ab} on it. Since we know the metric on the entire manifold, we can (in principle) integrate the null geodesic equations so that all null geodesics are known. From this knowledge it would be possible to construct all the null cones with apex at arbitrary points. This can be described by the equation

$$L(x^a,x'^a) = 0, \tag{1.1}$$

with x^a denoting the apex of the cone and x'^a points on the cone. $L_{,a}$ is a smooth null gradient field except for the apex and conjugate points. Note that x^a and x'^a are on equal footing since for fixed x'^a the points x^a that satisfy Eq. (1.1) form the null cone emanating from x'^a (see Fig. 1). The idea is now to move one of the points, say x'^a, to null infinity, i.e., to $\mathscr{I}^+$. Introducing a standard Bondi coordinate system[3] $(u,\zeta,\bar{\zeta},r)$ in the neighborhood of $\mathscr{I}^+$ (with $r = 0$ on $\mathscr{I}^+$), we write $\mathbf{x}'^a = (u,\zeta,\bar{\zeta},r = 0)$. Equation (1.1) thus becomes

$$L(x^a,\mathbf{x}'^a) = L(x^a,u,\zeta,\bar{\zeta}) = 0, \tag{1.2}$$

or by solving for u

$$u = Z(x^a,\zeta,\bar{\zeta}). \tag{1.3}$$

[Note that the function L is not unique since it can be multiplied by a smooth, nonvanishing $\alpha(x,x')$ and Eq. (1.1) would still hold. However, for $x'^a = \mathbf{x}'^a$ there is a unique choice (relative to the Bondi coordinate), namely, $L(x^a,\mathbf{x}'^a) = u - Z(x^a,\zeta,\bar{\zeta})$ and therefore the cuts are described by $u = Z(x^a,\zeta,\bar{\zeta})$.]

For arbitrary but fixed $u,\zeta,\bar{\zeta}$, Eq. (1.3) describes the past null cone from a point on $\mathscr{I}^+$. An alternative but equally important interpretation of (1.3) comes from the reciprocal meaning of Eq. (1.1); if x^a is held fixed, then Eq. (1.2) describes the intersection of the null cone of x^a with $\mathscr{I}^+$, i.e., as $\zeta,\bar{\zeta}$ are varied, a "cut" of $\mathscr{I}^+$ is defined which *locally* will be a 2-surface and therefore can be described by Eq. (1.3).

{For special cases (e.g., flat space) these light cone cuts (LC cuts) will be S^2. However in general, due to the focusing effects of the space-time curvature the light cones will have self-intersections and hence the cuts themselves will have self-intersections or singularities [see Fig. 2(a),(b)].}

In addition to the two meanings already given to Z, it can also be thought of as a function on the bundle of null directions. The assignment of the $(\zeta,\bar{\zeta})$ to a null direction is via the intersection of the null geodesic (with that direction) and a generator $(\zeta,\bar{\zeta})$ of $\mathscr{I}$. This shows clearly the nonlocality of Z.

We now consider the converse problem; namely of how, if some appropriate $Z(x^a,\zeta,\bar{\zeta})$ is given, does one reconstruct

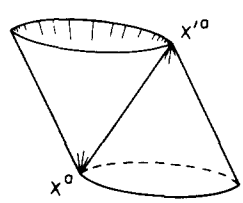


FIG. 1. The points x^a and x'^a are joined by a null geodesic.

[a)] This work has been supported by a grant from the National Science Foundation.

 0022-2488/83/102481-09$02.50

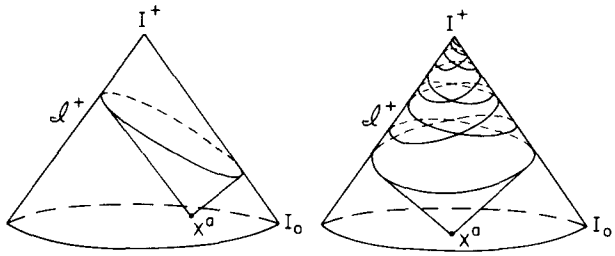


FIG. 2. (a) For Minkowski space the cuts are ellipsoidal figures (with the sphere as special case). (b) For a general space the cut may intersect itself several times.

the conformal metric $g_{ab}(x)$. To avoid at present a discussion of what is meant by the phrase "appropriate Z," consider the situation where we began with a metric and the $Z(x^a,\zeta,\bar{\zeta})$ is found from it. If we now forget the metric and retain knowledge only of the Z, we ask how can the metric be reconstructed.

This question is not hard to answer when we realize that for arbitrary but fixed $(u,\zeta,\bar{\zeta})$, $Z_{,a} \equiv \nabla_a Z$ must be a null vector field on M [because $u - Z(x^a,\zeta,\bar{\zeta}) = 0$ describes a null surface] and hence

$$g^{ab}(x)Z_{,a}(x,\zeta,\bar{\zeta})Z_{,b}(x,\zeta,\bar{\zeta}) = 0. \tag{1.4}$$

At fixed x^a, as $\zeta,\bar{\zeta}$ varies, $Z_{,a}$ sweeps out the null cone. The idea is then to construct the (conformal) g^{ab} at each point from knowledge of all the vectors $Z_{,a}$ at that point. By applying the differential operators[4] $\eth$ and $\bar{\eth}$ several times to Eq. (1.4) enough equations are obtained so that the metric can be written explicitly (up to conformal factor) in terms of the gradient basis

$$Z_{,a},\ \eth Z_{,a},\ \bar{\eth} Z_{,a},\ \eth\bar{\eth} Z_{,a}. \tag{1.5}$$

The resulting expression, though long and complicated, shows that g_{ab} can be obtained from knowledge of the appropriate $Z(x^a,\zeta,\bar{\zeta})$ and therefore we can write

$$g_{ab}(x) = q_{ab}(Z). \tag{1.6}$$

It is geometrically obvious that Z determines the conformal structure of $\tilde{M}$; for the light cones of two points in $\tilde{M}$ are in contact (along a common generator) if the points are null separated, and this occurs if the two corresponding cuts of $\mathscr{I}$ are tangent at some point.

The question which naturally arises now is what happens if an arbitrary $Z(x^a,\zeta,\bar{\zeta})$ is used in Eq. (1.6) for the construction of g_{ab}, rather than the "appropriately" chosen one. In general, the result would be a "metric" which would depend on ζ and $\bar{\zeta}$ as well as x^a. However, by demanding that $\eth g_{ab} = \bar{\eth} g_{ab} = 0$, conditions are imposed on $Z(x^a,\zeta,\bar{\zeta})$ so that the resulting g_{ab} will depend only on x^a and therefore will be a metric for M.

For flat space and Schwarzschild space, the Z is known explicitly,[5,6] and the conditions are, of course, satisfied identically. In the case of self-dual (or anti-self-dual) vacuum space-times, our Z reduces to the Z arising in the theory of H-space,[7] and hence a relatively simple differential equation exists for the determination of the Z, namely,

$$\eth^2 Z = \sigma_B(Z,\zeta,\bar{\zeta}). \tag{1.7}$$

The remarkable fact is that this is an equation for the global cross section of a line bundle over S^2, and at first sight there is no connection between Eq. (1.7) and space-time or the Einstein equations.

In fact,[8] the general (global) solution depends on four complex parameters x^a, i.e., the solution space is four (complex) dimensional. [This is analogous to $\ddot{x} = 0 \Rightarrow x = at + b$ with a two-dimensional (a,b) solution space.] This solution space is identified with (complex) space-time, and hence the solutions can be written as $u = Z(x^a,\zeta,\bar{\zeta})$. Furthermore, the metric obtained from the Z identically satisfies the vacuum Einstein equations, i.e., Eq. (1.7) yields both a manifold and a self-dual vacuum Einstein metric.

The question now is can Eq. (1.7) be generalized to a new equation for Z whose solution space is four-dimensional and, furthermore, which would yield, by our construction, a real metric at x^a. Could this equation be further specialized to yield the real vacuum Einstein equations?

To try to answer these questions, consider an "appropriate" $Z(x^a,\zeta,\bar{\zeta})$. This Z can always be thought of as solutions of the following pair of differential equations:

$$\eth^2 Z = \Lambda(Z,\eth Z,\bar{\eth} Z,\eth\bar{\eth} Z,\zeta,\bar{\zeta}), \quad \bar{\eth}^2 Z = \bar{\Lambda}. \tag{1.8}$$

To see this, we point out that if Z is known so are $\eth,\bar{\eth} Z$, and $\eth\bar{\eth} Z$, and hence we can write

$$x^a = x^a(Z,\eth Z,\bar{\eth} Z,\eth\bar{\eth} Z,\zeta,\bar{\zeta}); \tag{1.9}$$

thus from $\eth^2 Z$ being a known function of $x^a,\zeta,\bar{\zeta}$ we can produce Eq. (1.8). Conversely, if the "appropriate" $\Lambda(Z,\eth Z,\eth Z,\ \eth\bar{\eth} Z,\zeta,\bar{\zeta})$ is given, then the general solution of Eq. (1.8) admits a four-parameter freedom and hence we obtain an appropriate $Z(x^a,\zeta,\bar{\zeta})$. The "appropriateness" conditions on Z turn out to be most easily stated as conditions on Λ.

What remains is to chose a form for Λ so that Eq. (1.8) is equivalent to vacuum Einstein equations and includes H-space, Eq. (1.7), as a particular case.

We will argue in Sec. III that the required equation has the form

$$\eth^2 Z = \sigma_B(Z,\zeta,\bar{\zeta}) - \sigma_Z, \tag{1.10a}$$

$$\bar{\eth}^2 Z = \bar{\sigma}_B - \bar{\sigma}_Z, \tag{1.10b}$$

where σ_Z is to be a "universal" nonlinear functional of Z and σ_B is the free data of the problem. A perturbation scheme for obtaining Z in the weak field limit is presented, and a conjecture is given for the explicit form of σ_Z.

II. KINEMATICS OF THE CUTS

A. Definition of cuts

Since by assumption M is an asymptotically flat space-time,[2] it satisfies $M = \tilde{M} \cup \mathscr{I}$, where $\tilde{M}$ is the physical manifold and $\mathscr{I}$ is the boundary constructed from the completion of the future directed null geodesics of $\tilde{M}$.

This manifold M is provided with a pair (g_{ab},Ω) defined up to an equivalence relation; g_{ab} is a smooth, symmetric

tensor field with Lorentzian signature; Ω is a smooth, nonvanishing scalar field except on the boundary (at $\mathscr{I}$, $\Omega = 0$, $\Omega_{,a} \neq 0, g^{ab}\Omega_{,a}\Omega_{,b} = 0$) and the equivalence relation is given by

$$(g_{ab}, \Omega) \sim (g'_{ab}, \Omega') \Leftrightarrow g'_{ab} = \omega^2 g_{ab}, \quad \Omega' = \omega\Omega$$

for all smooth, nonvanishing ω. Although g_{ab} is defined up to a conformal factor, we can construct fields out of it which are well defined and have important physical meaning. Such examples are Weyl tensor, News tensor (which describe gravitational radiation), etc.

Another well-defined concept is the knowledge of null geodesics of M which (in principle) can be obtained by integrating the null geodesic equation. With the null geodesics we can then construct the null cones N_x with apex x^a. It is known that N_x is a surface-forming collection of points, and that locally (except for conjugate points) N_x can be immersed in a family of null hypersurfaces.

Finally we define the light cone cut C_x, the basic field of our formalism, as $C_x \equiv N_x \cap \mathscr{I}$. It can be shown, using the properties of null geodesics, that if $\tilde{M}$ is singularity-free, then $C_x = C_{x'}$ iff $x^a = x'^a$ so each cut can be labeled unambiguously with the parameter x^a.

[Notice that if the apex x^a is at $\mathscr{I}$, then C_x degenerates into a line, i.e., C_x coincides with one integral line of $n^a \equiv g^{ab}\Omega_{,b}$. Although this situation has been studied in (the complex) H-space, we have not investigated it yet, and for the time being we will restrict the domain of x^a to $\tilde{M}$. In this way we will obtain cuts that locally are cross sections of $\mathscr{I}$.]

As we pointed out in the Introduction, if we attach a Bondi coordinate system $(u,\zeta,\bar{\zeta})$ to $\mathscr{I}$, we can describe C_x locally by the equation

$$u = Z(x^a,\zeta,\bar{\zeta}). \tag{2.1}$$

We summarize here the three different ways $Z(x^a,\zeta,\bar{\zeta})$ can be thought of:

(i) For fixed x^a, Eq. (2.1) describes a piece of the cut C_x; therefore, varying x^a, we obtain a family of cuts (parametrized with x^a).

(ii) $Z(x^a,\zeta,\bar{\zeta})$ can also be thought of as a function on the bundle of null directions at the point x^a, with $\zeta,\bar{\zeta}$ of $\mathscr{I}$ parametrizing the null directions.

(iii) For fixed $(u,\zeta,\bar{\zeta})$, the points x^a that satisfy Eq. (2.1) form the past null cone from a point of $\mathscr{I}$.

According to (iii) $Z_{,a}$ is a null vector, and therefore it obeys

$$g^{ab}(x)Z_{,a}(x,\zeta,\bar{\zeta})Z_{,b}(x,\zeta,\bar{\zeta}) = 0. \tag{2.2}$$

Notice that, by fixing x^a and varying $(\zeta,\bar{\zeta})$, $Z_{,a}$ sweeps the tangent null cone of x^a (or rather a piece of it since Z is only locally defined), and this in turn enables us to construct a conformal metric for $\tilde{M}$ [which, of course, is our original $g_{ab}(x)$]. (See Secs. IIB and IIC.)

B. Metric construction technique

It is clear from these considerations that the (conformal) metric $g_{ab}(x)$ and $Z(x;\zeta,\bar{\zeta})$ should be regarded as equally important fields in the sense that knowledge of the metric allows the construction of the cuts and conversely knowledge of Z allows the construction of $g_{ab}(x)$.

Before showing the explicit construction[9] we may ask can an arbitrary $Z(x^a,\zeta,\bar{\zeta})$ give rise to a metric $g_{ab}(x)$? As we will see, the answer is no. Therefore, the cuts belong to a certain class of functions $Z(x^a,\zeta,\bar{\zeta})$ (capable of producing local fields) in a manifold M. We want to identify this particular class.

To put the question in an appropriate context, assume we have a manifold $M = \tilde{M} \cup \mathscr{I}$ *without metric* and a collection of arbitrary functions $Z(x^a,\zeta,\bar{\zeta})$. We ask if it is possible to define a nonvanishing, symmetric tensor field $q^{ab}(x)$ that satisfies

$$q^{ab}(x)Z_{,a}(x,\zeta,\bar{\zeta})Z_{,b}(x,\zeta,\bar{\zeta}) = 0. \tag{2.3}$$

Clearly the answer in general is no; there are an infinite number of conditions imposed on the ten components of q^{ab}. Therefore, we ask what are the conditions to be imposed on Z such that a nontrivial q^{ab} with signature $(+\ -\ -\ -)$ or $(+\ +\ -\ -)$ exists?

Two types of equations arise in the search for these conditions. The first type actually yield the components of the metric (these equations are denoted by (M.i), $i = 1,2,\cdots$) while the second type [denoted by (C.i)] identify necessary conditions to be satisfied by Z or more precisely by $\eth^2 Z$.

In order to simplify the notation, we introduce a basis θ^i_a (i tetrad index $(0, +, -, 1)$, a tensor index) defined in terms of Z as follows:

$$\theta^0_a = Z_{,a}, \quad \theta^+_a = \eth Z_{,a}, \quad \theta^-_a = \bar{\eth} Z_{,a}, \quad \theta^1_a = \eth\bar{\eth} Z_{,a}. \tag{2.4}$$

(We assume, of course, that for some range of $\zeta,\bar{\zeta}$ these form a linearly independent set.) In terms of this basis we define the metric components $q^{ij} \equiv q^{ab}\theta^i_a\theta^j_b$.

The basic idea is now to apply $\eth$ and $\bar{\eth}$ to q^{ab} and demand $\eth q^{ab} = \bar{\eth} q^{ab} = 0$. Technically, it turns out to be easier to apply this to q^{ij}. (Note that $\eth q^{ij}$ and $\bar{\eth} q^{ij}$ are not zero since the θ^i_a are $\zeta,\bar{\zeta}$-dependent.) In particular, since Eq. (2.3) reads (with the new notation)

$$q^{ab}\theta^0_a\theta^0_b \equiv q^{00} = 0. \tag{M.1}$$

We obtain by applying $\eth$ and $\bar{\eth}$ to q^{00} (using $\eth q^{ab} = 0$),

$$q^{ab}Z_a\,\eth Z_b = q^{ab}\theta^0_a\theta^+_b \equiv q^{0+} = 0, \tag{M.2}$$

$$q^{ab}Z_a\,\bar{\eth} Z_b = q^{ab}\theta^0_a\theta^-_b \equiv q^{0-} = 0. \tag{M.3}$$

Now, by operating on (M.2) with $\bar{\eth}$ or (M.3) with $\eth$, we obtain

$$q^{+-}/q^{01} = -1. \tag{M.4}$$

In this way the trivial coefficients have been exhausted, and new $\eth$ and $\bar{\eth}$ operations provide the six other components plus the conditions we are looking for.

We want to stress the importance of $\eth^2 Z$ (and $\bar{\eth}^2 Z$) since this complex function allows us to determine the remaining components and conditions.

Defining $\Lambda(x^a,\zeta,\bar{\zeta}) \equiv \eth^2 Z$ and denoting $\Lambda_{,a} = \Lambda_i\theta^i_a$, we obtain by applying $\eth$ to (M.2)

$$q^{ab}(\theta^0_a\Lambda_{,b} + \theta^+_a\theta^+_b) = 0 \quad \text{or} \quad \Lambda_i q^{i0} + q^{++} = 0.$$

Since the only nonvanishing q^{i0} is q^{01}, we obtain

$$q^{++}/q^{01} = -\Lambda_1. \tag{M.5}$$

Similarly operating with $\eth$ on (M.3), we obtain

$$q^{--}/q^{01} = -\bar{\Lambda}_1. \tag{M.6}$$

If we now operate with $\eth$ and $\bar{\eth}$ on (M.5) and (M.6), we obtain four independent equations for q^{+1} and q^{-1}. Therefore, two equations must be identities among the Λ_i's. After solving this system, we obtain

$$q^{+1}/q^{01} = -\tfrac{1}{2}(\bar{\eth}\Lambda_1 + \Lambda_1\bar{W}), \tag{M.7}$$

$$q^{-1}/q^{01} = -\tfrac{1}{2}(\eth\bar{\Lambda}_1 + \bar{\Lambda}_1 W), \tag{M.8}$$

and

$$\Lambda_+ = W - \tfrac{1}{2}(\Lambda_1\bar{W} + \bar{\eth}\Lambda_1 + \eth \ln P), \tag{C.1}$$

$$\bar{\Lambda}_+ = \bar{W} - \tfrac{1}{2}(\bar{\Lambda}_1 W + \eth\bar{\Lambda}_1 + \bar{\eth} \ln P) \tag{C.2}$$

with

$$P \equiv (1 - \Lambda_1\bar{\Lambda}_1)^{-1},\ W \equiv (\eth\Lambda_1 - 2\Lambda_-)/\Lambda_1 + \eth \ln P.$$

In obtaining (M.7), (M.8), (C.1), and (C.2), we have used $\eth q^{01} \equiv \eth(q^{ab}\theta^1_a\theta^0_b) = q^{+1} + q^{ab}\theta^0_a\eth\theta^1_b = q^{+1} + q^{ab}\theta^0_a\bar{\eth}\Lambda_b = q^{+1} + q^{01}\Pi_1$, where the coefficient Π_1 (a tetrad component of $\bar{\eth}\Lambda_a$) is defined in terms of Λ_i and $\eth$, $\bar{\eth}$ of Λ_i (the explicit form is given in Appendix A). Therefore, we now have a differential equation for q^{01}. After using conditions (C.1) and (C.2), it simplifies to

$$\eth \ln q^{01} = W, \tag{M.9a}$$

$$\bar{\eth} \ln q^{01} = \bar{W}, \tag{M.9b}$$

from which q^{01} can be obtained. (One can check, using $\eth^2\bar{\Lambda} = \bar{\eth}^2\Lambda$ and after a long calculation, that the integrability condition of Eqs. (M.9) is identically satisfied.) Notice that q^{01} is obtained only up to a conformal factor, i.e., if q^{01} satisfies (M.9) so does $q^{01}e^{\beta(x)}$. This, however, is the conformal freedom we expected.

Finally, taking $\bar{\eth}$ on (M.7) or $\eth$ on (M.8), we obtain

$$q^{11} - q^{ij}\Lambda_i\Lambda_j = \bar{\eth}q^{+1} + \eth q^{-1} + \eth\bar{\eth}q^{01} - 2q^{01},$$

and, since the rest of q^{ij} have been already obtained, we can algebraically solve this equation for q^{11} and denote it by

$$q^{11}/q^{01} = -2\varepsilon(\Lambda_0,\Lambda_1,\Lambda_-),\ \varepsilon = \bar{\varepsilon}. \tag{M.10}$$

The explicit form of ε is long and is not important at this point (see Appendix A).

The last conditions come from operating with $\eth$ and $\bar{\eth}$ on (M.10)

$$\eth\varepsilon + \varepsilon(W - 2\Pi_1) + (1 - \tfrac{1}{2}\Pi_+)(\bar{\eth}\Lambda_1 + \Lambda_1\bar{W}) - \tfrac{1}{2}(\eth\bar{\Lambda}_1 + \bar{\Lambda}_1 W)\Pi_- + \Pi_0 = 0, \tag{C.3}$$

$$\text{c.c.} \tag{C.4}$$

The object we have constructed, $q^{ab} = q^{ij}\theta^a_i\theta^b_j$, satisfies $\eth q^{ab} = \bar{\eth}q^{ab} = 0$ (see Sec. IIC). It therefore is a local metric determined from Z.

Finally, if we want a metric with Lorentzian signature we have to impose $\det(q^{ab}) < 0 \Rightarrow (\det q^{ij})\det(\theta^a_i\theta^b_j) < 0 \Rightarrow \det(q^{ij}) > 0$ since it can be shown that, for a gradient basis with $\theta^a_0\theta^a_1$ real vectors and $\theta^a_+, \theta^a_- = \bar{\theta}^a_+$ complex vectors, $\det(\theta^a_i)$ is always a pure imaginary number. Therefore, we obtain

$$\det(q^{ij}) = (q^{01})^4 P > 0$$

or [using $P \equiv (1 - \Lambda_1\Lambda_1)^{-1}$]

$$P > 0 \Leftrightarrow |\Lambda_1| < 1. \tag{C.5}$$

From now on the metric whose components are given by (M.1)–(M.10) and satisfy conditions (C.1)–(C.5) will be denoted by g^{ab}.

As we mentioned in the Introduction, we can always think of $Z(x^a,\zeta,\bar{\zeta})$ as coming from the solution of

$$\eth^2 Z = \Lambda(\theta^i,\zeta,\bar{\zeta}), \tag{2.5a}$$

$$\bar{\eth}^2 Z = \bar{\Lambda}(\theta^i,\zeta,\bar{\zeta}) \tag{2.5b}$$

with $\theta^i = (Z,\eth Z,\bar{\eth}Z,\eth\bar{\eth}Z)$ for $i = 0, +, -, 1$, respectively.

In this way $M = \tilde{M} \cup \mathscr{I}$ should be regarded as the manifold of solutions of Eqs. (2.5). The conditions (C.1)–(C.5) to be imposed on $Z(x^a,\zeta,\bar{\zeta})$ in order to produce a local field $g^{ab}(x)$ are now stated in terms of derivatives of Λ with respect to θ^i and $(\zeta,\bar{\zeta})$. Note that, in this sense, the most natural way to present our "appropriate $Z(x^a,\zeta,\bar{\zeta})$" is as the solution of Eq. (2.5) with Λ that satisfies conditions (C.1)–(C.5).

C. Alternative approach

In this subsection we want to investigate an alternative approach to the construction of a metric $q_{ab}(x)$.

By assumption Z will be the general (four-parameter) solution of equations

$$\eth^2 Z = \Lambda(Z,\eth Z,\bar{\eth}Z,\eth\bar{\eth}Z,\zeta,\bar{\zeta}), \tag{2.5a'}$$

$$\bar{\eth}^2 Z = \bar{\Lambda}(Z,\eth Z,\bar{\eth}Z,\eth\bar{\eth}Z,\zeta,\bar{\zeta}), \tag{2.5b'}$$

where Λ is a *given* complex function that satisfies identically

$$\eth^2\bar{\Lambda} = \bar{\eth}^2\Lambda \tag{2.6}$$

[namely, the integrability condition of Eqs. (2.5)].

Defining

$$\theta^0 = Z,\ \theta^+ = \eth Z,\ \theta^- = \bar{\eth}Z,\ \theta^1 = \eth\bar{\eth}Z, \tag{2.7}$$

we can transform Eqs. (2.5) into a system of first-order differential equations

$$\eth\begin{pmatrix}\theta^0\\ \theta^+\\ \theta^-\\ \theta^1\end{pmatrix} = \begin{pmatrix}\theta^+\\ \Lambda\\ \theta'\\ \bar{\eth}\Lambda - 2\theta^+\end{pmatrix}, \tag{2.8a}$$

$$\bar{\eth}\begin{pmatrix}\theta^0\\ \theta^+\\ \theta^-\\ \theta^1\end{pmatrix} = \begin{pmatrix}\theta^-\\ \theta^1\\ \bar{\Lambda}\\ \eth\bar{\Lambda} - 2\theta^-\end{pmatrix} \tag{2.8b}$$

or, more compactly,

$$\eth\theta^i = \Delta^i(\theta^j,\zeta,\bar{\zeta}), \tag{2.9a}$$

$$\bar{\eth}\theta^i = \bar{\Delta}^i(\theta^j,\zeta,\bar{\zeta}). \tag{2.9b}$$

The integrability conditions now read

$$\bar{\eth}\Delta^i - \eth\bar{\Delta}^i = 2s_i\theta^i \text{ (no sum over } i), \tag{2.10}$$

where s_i is the spin weight of θ^i. We can easily check that the only nontrivial equation is precisely Eq. (2.6).

Since by assumption the solution is of the form

$$u = Z(x^a,\zeta,\bar{\zeta}),\ a = 1,2,3,4, \tag{2.11}$$

we have as well

$$\theta^i = \theta^i(x^a,\zeta,\bar{\zeta}). \tag{2.12}$$

We will assume the solution space (i.e., the collection of

all points x^a) forms a differentiable manifold, denoted by M.

Notice that we have a natural form basis $\theta^i_a \equiv \theta^i_{,a}$ living in M, where $_{,a}$ means the derivative with respect to x^a. From Eqs. (2.9) we see that θ^i_a satisfies

$$\eth \theta^i_a = \Delta^i_{,a},$$
$$\bar{\eth} \theta^i_a = \bar{\Delta}^i_{,a},$$

or expanding $\Delta^i_{,a}$ in terms of this basis

$$\eth \theta^i_a = \Delta^i_j \theta^j_a, \tag{2.13a}$$
$$\bar{\eth} \theta^i_a = \bar{\Delta}^i_j \theta^j_a, \tag{2.13b}$$

where $\Delta^i_j \equiv \partial \Delta^i / \partial \theta^j$, $\bar{\Delta}^i_j \equiv \partial \bar{\Delta}^i / \partial \theta^j$.

Let us now define a symmetric tensor q^{ab} (eventually to be our metric tensor) as follows:

$$q^{ab} \equiv q^{ij} \theta^a_i \theta^b_j, \tag{2.14}$$

where θ^a_i is the dual basis (i.e., $\theta^a_i \theta^j_a = \delta^j_i$) and q^{ij} are arbitrary functions of $(x^a, \zeta, \bar{\zeta})$ interpreted as the scalar product between θ^i_a and θ^j_b.

Since in general q^{ab} will depend on x^a and $\zeta, \bar{\zeta}$, we raise the question, what are the conditions to be imposed on q^{ij} so that $q^{ab} = q^{ab}(x)$?

We thus want a tensor q^{ab} such that

$$\eth q^{ab} = 0, \quad \bar{\eth} q^{ab} = 0.$$

By applying $\eth$ and $\bar{\eth}$ to Eq. (2.14), we easily see that this is equivalent to the demand that the coefficients q^{ij} satisfy

$$\eth q^{ij} = 2\Delta^{(i}_k q^{j)k}, \tag{2.15a}$$
$$\bar{\eth} q^{ij} = 2\bar{\Delta}^{(i}_k q^{j)k}, \tag{2.15b}$$

where Δ^i_k, Δ^i_k are defined in Eqs. (2.13).

Before studying the solutions to Eqs. (2.15), we first investigate its integrability conditions. They are

$$\bar{\eth}(\Delta^{(i}_k q^{j)k}) - \eth(\bar{\Delta}^{(i}_k q^{j)k}) = s_{ij} q^{ij} \text{ (no sum on } i,j), \tag{2.16}$$

where s_{ij} is the spin weight of q^{ij}. After a brief calculation we rewrite Eq. (2.16) as

$$[(\bar{\eth}\Delta^{(i}_k + \Delta^{(i}_d \bar{\Delta}^{d)}) - (\bar{\eth}\bar{\Delta}^{(i}_k + \bar{\Delta}^{(i}_d \Delta^{d)}_k)] q^{j)k} = s_{ij} q^{ij}. \tag{2.17}$$

This is precisely what we get from integrability conditions (2.10) if we take its gradient, multiply by $q^{ab}\theta^j_b$, symmetrize over i and j, and use $s_{ij} = s_i + s_j$. We thus see that Eq. (2.16) is identically satisfied by virtue of the original integrability conditions Eq. (2.10).

Solutions of Eqs. (2.15) thus do exist. It would be of interest to investigate the general solution of Eqs. (2.15) since this may lead to a generalization of our approach. However, (as we saw earlier in Sec. IIB) we are interested in the particular class of solution such that

$$q^{ab}\theta^0_a \theta^0_b = q^{00} = 0. \tag{2.18}$$

Therefore, we will solve Eqs. (2.15) only for our special case ($q^{00} = 0$) and defer the investigation of the general solution for the future.

We want to obtain the conditions on Λ that characterize this class. They will arise by explicitly solving Eqs. (2.15).

Since q^{00} is given we can put this information on the left side of Eqs. (2.15) and solve algebraically for the q^{ij}'s on the right (the explicit form of Δ^i_j, Δ^i_j is given in Appendix A).

Thus,

$$\left.\begin{aligned} 0 &= \eth q^{00} = 2q^{0+} \\ 0 &= \bar{\eth} q^{00} = 2q^{0-} \end{aligned}\right\} \Rightarrow \begin{cases} q^{0+} = 0 & (2.19) \\ q^{0-} = 0. & (2.20) \end{cases}$$

Since we now know q^{0+}, q^{0-}, we repeat the same construction outlined above to obtain

$$\left.\begin{aligned} 0 &= \eth q^{0+} = q^{++} + q^{01}\Lambda_1 \\ 0 &= \bar{\eth} q^{0-} = q^{--} + q^{01}\bar{\Lambda}_1 \\ 0 &= \bar{\eth} q^{0+} = \eth q^{0-} = q^{+-} + q^{01} \end{aligned}\right\} \Rightarrow \begin{cases} q^{++}/q^{01} = -\Lambda_1 & (2.21) \\ q^{--}/q^{01} = -\bar{\Lambda}_1 & (2.22) \\ q^{+-}/q^{01} = -1. & (2.23) \end{cases}$$

Repeating the same argument for $\eth q^{++}$, $\bar{\eth} q^{++}$, $\eth q^{01}$, $\bar{\eth} q^{01}$, $\eth q^{--}$, $\bar{\eth} q^{--}$, we obtain a differential equation for q^{01} and four independent algebraic equations for q^{+1} and q^{-1}. This leads to two conditions among the Λ_i's. We write the solutions and conditions as

$$q^{+1}/q^{01} = -\tfrac{1}{2}(\bar{\eth}\Lambda_1 + \Lambda_1 \bar{W}), \tag{2.24}$$
$$q^{-1}/q^{01} = -\tfrac{1}{2}(\eth\bar{\Lambda}_1 + \bar{\Lambda}_1 W), \tag{2.25}$$
$$\Lambda_+ = W - \tfrac{1}{2}(\Lambda_1 \bar{W} + \bar{\eth}\Lambda_1 + \eth \ln P), \tag{2.26}$$
$$\bar{\Lambda}_+ = \bar{W} - \tfrac{1}{2}(\bar{\Lambda}_1 W + \eth\bar{\Lambda}_1 + \bar{\eth} \ln P), \tag{2.27}$$

with $P \equiv (1 - \Lambda_1\bar{\Lambda}_1)^{-1}$, $W \equiv (\eth\Lambda_1 - 2\Lambda_-)/\Lambda_1 + \eth \ln P$. The differential equation for q^{01} reads

$$\eth \ln q^{01} = W, \tag{2.28a}$$
$$\bar{\eth} \ln q^{01} = \bar{W}, \tag{2.28b}$$

from which q^{01} can be obtained.

[As we have already proved, $\eth\bar{W} = \bar{\eth}W$ because Eq. (2.16) is satisfied; therefore a solution of Eqs. (2.28) exists.]

Finally, from $\bar{\eth} q^{+1}$ we obtain

$$\bar{\eth} q^{+1} = q^{11} + \tilde{\Pi}_i q^{i+}. \tag{2.29}$$

Since the other q^{ij}'s have been already obtained, we can algebraically solve for q^{11}. We can check, after a long but straightforward calculation, that

$$q^{11} = \bar{\eth} q^{+1} - q^{+i}\tilde{\Pi}_i = \eth q^{-1} - q^{-i}\Pi_i = -2\varepsilon q^{01}, \tag{2.30}$$

where ε is defined in Eq. (M.10). Moreover, using q^{11} of Eq. (2.30) in the equation for $\eth(q^{+1})$ and c.c., we find they are identically satisfied.

The final two equations for $\eth q^{11}$ and $\bar{\eth} q^{11}$ lead to the further conditions on the Λ_i:

$$\eth q^{11} = q^{1i}\Pi_i, \tag{2.31}$$
$$\bar{\eth} q^{11} = q^{1i}\tilde{\Pi}_i, \tag{2.32}$$

which are identical to (C.3) and (C.4)

As we have shown in this subsection, we can duplicate the results of Sec. IIB. However, the starting approaches for obtaining the results were quite different.

In Sec. IIB we began with a function of six variables in a manifold M with boundary $S^2 \times R$ whereas in Sec. IIC we started with a differential equation for the local cross section of a line bundle over S^2.

As we will see in Sec. III we will follow the lines of the later approach to provide field equations for Z.

D. Some examples

As we have shown in Sec. IIB, we can distinguish the "appropriate $Z(x^a, \zeta, \bar{\zeta})$," i.e., the class of cuts that produce

local metrics, as those who satisfy conditions (C.1)–(C.5). We now want to present some known members of this class of Z.

For Minkowski space and Schwarzschild space the cuts can be constructed explicitly by integrating the null geodesic equation, constructing the cones and finding their intersection with $\mathscr{I}$. Since the cuts in these cases are coming from a metric, the conditions (C.1)–(C.5) must be satisfied identically.

The Schwarzschild case has been analyzed in detail by Joshi *et al.*,[6] and will be presented elsewhere.

In Minkowski space the intersection of null cones with $\mathscr{I}$ can be described by the function[5,6]

$$u = Z^M(x^a,\zeta,\bar{\zeta}) = x^a l_a(\zeta,\bar{\zeta}), \tag{2.33a}$$

where

$$l_a \equiv (1+\zeta\bar{\zeta},\zeta+\bar{\zeta},(\zeta-\bar{\zeta})/2i,-1+\zeta\bar{\zeta})/P_0,$$
$$P_0 = \tfrac{1}{2}(1+\zeta\bar{\zeta}), \tag{2.33b}$$

is defined from the spherical harmonics $Y_{0,0}, Y_{1,m}$, $m = 1,0,-1$. Since $\eth^2 Y_{00} = \eth^2 Y_{1m} = 0$, then Z^M satisfies

$$\eth^2 Z^M = \bar{\eth}^2 Z^M = 0 \tag{2.34}$$

and therefore $\Lambda^M = 0$ for this particular case. The basis θ^i_a has the form

$$\theta^0_a = l_a,\ \theta^+_a = \eth l_a,\ \theta^-_a = \bar{\eth} l_a,\ \theta^1_a = \eth\bar{\eth} l_a.$$

For fixed $(\zeta,\bar{\zeta})$, θ^i_a is the conventional parallelly propagated null tetrad $l_a, m_a, \bar{m}_a$, and $n_a - l_a$ respectively. Since $\Lambda^M = 0$, the conditions (C.1)–(C.5) are trivially satisfied and the only nonvanishing components of (M.1)–(M.10) are

$$g^{+-}/g^{01} = -1,\ g^{11}/g^{01} = -2 \text{ with } g^{01} = \omega^2(x); \tag{2.35}$$

therefore, we obtain the usual expression of the (conformal) Minkowski metric in tetrad language

$$g_{ab} = 2\omega^2(l_{(a}n_{b)} - m_{(a}\bar{m}_{b)}). \tag{2.36}$$

Our last example is the cuts Z^H arising from the theory of H-space.[7] In order to check that Z^H belongs to our class, we have to generalize slightly our formalism. This is done by complexifying $(\zeta,\bar{\zeta})\to(\zeta,\tilde{\zeta})$; therefore, $(\eth,\bar{\eth})\to(\eth,\tilde{\eth})$, $(\Lambda,\bar{\Lambda})\to(\Lambda,\tilde{\Lambda})$, real functions$\to$spin weight zero complex functions,[4] etc. It is easy to check that all our equations still hold after this procedure.

In H-space theory Z^H satisfies the following differential equation:

$$\eth^2 Z^H = \sigma_B(Z^H,\zeta,\tilde{\zeta}), \tag{2.37}$$

where σ_B is an arbitrary function of three variables and $\tilde{\eth}^2 Z$ is not given. This is an equation for the global cross section of a line bundle over S^2. The general regular solution admits a four (complex) parameter freedom and hence $Z^H = Z^H(x^a,\zeta,\tilde{\zeta})$. In this context the manifold M arises as the space of solutions x^a.

Since Eq. (2.37) plus global regularity conditions determine $Z^H, \tilde{\Lambda}^H$ is then *defined* by $\tilde{\Lambda}^H \equiv \tilde{\eth}^2 Z^H$. However, we can drop global conditions and demand that Z^H satisfy the following pair of local equations:

$$\eth^2 Z^H = \sigma_B(Z^H,\zeta,\tilde{\zeta}),$$
$$\tilde{\eth}^2 Z^H = \tilde{\Lambda}^H(Z^H\eth Z^H,\tilde{\eth} Z^H,\eth\tilde{\eth} Z^H,\zeta,\tilde{\zeta}). \tag{2.38}$$

Since Z^H produces a local metric, we know that conditions (C.1)–(C.4) are identically satisfied. Nevertheless, we would like to see what equations are produced from applying (C.1)–(C.4) to (2.38).

Since $\Lambda^H = \sigma_B(Z,\zeta,\tilde{\zeta})$, then $\Lambda_+ = \Lambda_- = \Lambda_1 = 0$, $\Lambda_0 = \dot{\sigma}_B$. Defining $2\mathscr{F} \equiv \tilde{\Lambda}_1$ to match the notation in the theory of H-space,[7] we find

$$0 = W \Leftrightarrow \eth g^{01} = 0, \quad \tilde{W} = 0 \Leftrightarrow \tilde{\eth}\mathscr{F} = \tilde{\Lambda}_-, \tag{C.1'}$$

$$\tilde{\Lambda}_+ = -\eth\mathscr{F}, \tag{C.2'}$$

$$\eth^3\mathscr{F} = 4\dot{\sigma}_B\,\eth\mathscr{F} + 2\,\eth\dot{\sigma}_B - 2\,\tilde{\eth}\dot{\sigma}_B, \tag{C.3'}$$

$$\eth(\tilde{\Lambda}_0 + \dot{\sigma}_B\mathscr{F}^2 - \tfrac{1}{2}\mathscr{F}\,\eth^2\mathscr{F} + \tfrac{1}{4}(\eth\mathscr{F})^2 + \tfrac{1}{2}\,\eth\tilde{\eth}\mathscr{F} - 3\mathscr{F}) = 0$$
$$\Rightarrow \tilde{\Lambda}_0 = 3\mathscr{F} - \dot{\sigma}_B\mathscr{F}^2 + \tfrac{1}{2}\mathscr{F}\eth^2\mathscr{F} - \tfrac{1}{4}(\eth\mathscr{F})^2 - \tfrac{1}{2}\eth\tilde{\eth}\mathscr{F}, \tag{C.4'}$$

which are well-known identities from H-space theory.[10] We have thus elucidated the role of these equations as being required to produce a local (H-space) metric.

Following the procedure described earlier, we have

$$g^{00} = 0, \tag{M.1'}$$
$$g^{0+} = 0, \tag{M.2'}$$
$$g^{0-} = 0, \tag{M.3'}$$
$$g^{+-} = -1, \tag{M.4'}$$
$$g^{++} = 0, \tag{M.5'}$$
$$g^{+1} = 0, \tag{M.6'}$$
$$g^{--} = -2\mathscr{F}, \tag{M.7'}$$
$$g^{-1} = -\eth\mathscr{F}, \tag{M.8'}$$
$$g^{11} = -2(1 - \mathscr{F}\dot{\sigma}_B + \tfrac{1}{2}\eth^2\mathscr{F}). \tag{M.9'}$$

Equations (M.1')–(M.9') give precisely the components of the metric of H-space and (C.1)–(C.4) assure $\eth q^{ij} = \tilde{\eth} q^{ij} = 0$.[7,10]

We want to emphasize that this metric automatically satisfies the vacuum Einstein equations ,i.e., Eq. (2.37) is the vacuum Einstein equations for self-dual Weyl tensors.[7]

III. FIELD EQUATIONS FOR THE CUTS

Although the class of cuts defined by the solution of Eq. (2.5), where Λ satisfies conditions (C.1)–(C.5), allows us to define metrics in the manifold of solutions M, in general the metrics $g_{ab}(x)$ will not satisfy Einstein vacuum equations. Therefore, our next step should be to choose a form for Λ so that Eq. (2.5) is equivalent to Einstein equations.

In order to obtain a geometric understanding of Λ, we will recall results originally obtained by Sachs[11] regarding the asymptotic shear of null surfaces near $\mathscr{I}^+$. (In Appendix B we offer an alternative proof of the theorem using the available intrinsic structure at $\mathscr{I}$.)

Assume that one has a Bondi slicing of $\mathscr{I}^+$ (with associated coordinates $(u,\zeta,\bar{\zeta})$ and an asymptotic shear $\sigma_B(u,\zeta,\bar{\zeta})$ for the associated Bondi null surfaces; then the asymptotic shear of any other cut of $\mathscr{I}^+$ [given by $u = \alpha(\zeta,\bar{\zeta})$] is obtained from

$$\sigma_\alpha(\zeta,\bar\zeta) = \sigma_B(\alpha(\zeta,\bar\zeta),\zeta,\bar\zeta) - \eth^2\alpha(\zeta,\bar\zeta)$$

or

$$\eth^2\alpha = \sigma_B(\alpha,\zeta,\bar\zeta) - \sigma_\alpha. \tag{3.1}$$

If we now identify the cut $u = \alpha(\zeta,\bar\zeta)$ with members of the family of light cone cuts, Eq. (3.1) takes the form

$$\eth^2 Z = \sigma_B(Z,\zeta,\bar\zeta) - \sigma_Z. \tag{3.2}$$

From Eqs. (2.5) and (3.2) we see that Λ can be interpreted [for each light cone (l.c.) cut $Z(x^a,\zeta,\bar\zeta)$] as the difference between the Bondi shear and the shear associated with the l.c. cut.

The idea now, to construct field equations for Z, is to treat $\sigma_B(u,\zeta,\bar\zeta)$ as free characteristic data and to try to write σ_Z as some "universal" *functional* of Z with the result that Eq. (3.2) is a nonlinear differential equation for Z. As we said before, in this equation no mention whatsoever is made of the manifold—the equation is for local cross sections of a line bundle over S^2. The solution space x^a defines the manifold while the solutions themselves, $Z(x^a,\zeta,\bar\zeta)$, determine the metric on the manifold.

The last question to be answered is how can σ_Z be written as a "universal" functional of Z. We would like to present (i) an argument for the construction of σ_Z using the four-dimensional geometry and (ii) a conjecture for the explicit form of σ_Z as a function of $Z,\eth Z,\bar\eth Z$, and $\eth\bar\eth Z$.

(i) If we consider Z to be a known "appropriate" function of x^a, then the local metric, connection, and curvature tensor can be written as functionals of Z. In particular, we could write the optical parameters ρ and σ[12] associated with each light cone (with apex x^a) as functions of Z. In terms of these parameters, we can write the geodesic deviation equation for a null cone as[12,13]

$$D\rho = \rho^2 + \sigma\bar\sigma + \phi_{00}, \tag{3.3}$$

$$D\sigma = 2\rho\sigma + \psi_0, \tag{3.4}$$

where

$$\sigma = M^a M^b \nabla_a L_b,\ \rho = M^a \bar M^b \nabla_a L_b, \tag{3.5}$$

$$\phi_{00} = \tfrac{1}{2}R_{ab}L^a L^b,\ \psi_0 = C_{abcd}L^a M^b L^c M^d, \tag{3.6}$$

$$L^a = g^{ab}L_b,\ D = L^a\nabla_a,\ L_a = L_{,a} \tag{3.7}$$

[the null cone is given by $L(x^a,x'^a) = 0$],

and M is a complex vector lying on the cone that satisfies $M^a M_a = 0$, $M^a\bar M_a = -1$.

The choice $M^a n_a = 0$, $L^a n_a = 1$ $(n_a = \Omega_{,a})$ at $\mathscr{I}^+$ guarantees σ_Z and σ_B pick the same conformal factor under rescaling of n^a.

Finally by integrating Eqs. (3.3) and (3.4) we would obtain the asymptotic σ_Z as a function of Z. The Einstein equations go into the choice of ϕ_{00}, if for example, $\phi_{00} = 0$ or if conformally equivalent to zero, we would have the vacuum equations.

Though the idea described here to determine Λ is more a program than an exact theory, we, nevertheless, feel it is essentially correct. However, we do not yet know how to carry out this program explicitly. Thus, instead of trying to derive the full expression for σ_z, our next step will be to define a perturbation scheme for small σ_B. Notice that if $\sigma_B = 0$, then Z^M of Eq. (2.6) is a solution of Eq. (3.2) and $\sigma_z = 0$ for this case. Therefore, if we multiply the free data σ_B by a strength factor ϵ and consider situations for which $\epsilon \ll 1$, we can implement a perturbation procedure for Z in powers of ϵ. We write

$$\begin{aligned} Z &= Z_0 + \epsilon Z_1 + \epsilon^2 Z_2 + \cdots, \\ \sigma &= \sigma_0 + \epsilon\sigma_1 + \epsilon^2\sigma_2 + \cdots, \\ \rho &= \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \cdots, \\ C_{abcd} &= \epsilon C_{1abcd} + \cdots, \\ s &= s_0 + \epsilon s_1 + \epsilon^2 s_2 + \cdots, \\ \psi_0 &= \epsilon\psi_{01} + \epsilon^2\psi_{02} + \cdots, \text{ etc.} \end{aligned} \tag{3.8, 3.9}$$

Order 0: Solving Eqs. (3.3) and (3.4) for Minkowski space, we find $\sigma_0 = 0$, $\rho_0 = -1/s_0$

$$\eth^2 Z_0 = 0 \Rightarrow Z_0 = Z^M = x^a l_a. \tag{3.10}$$

Order ϵ: With Z_0 we construct η_{ab}; from the Bianchi identities,[14] we obtain C'_{abcd} as a function of x^a and σ_B. We now solve Eqs. (3.3) and (3.4) for σ_1 and ρ_1:

$$D\rho_1 = 2\rho_0\rho_1, \tag{3.11}$$

$$D\sigma_1 = 2\rho_0\sigma_1 + \psi_{01}, \tag{3.12}$$

from which we obtain $\rho_1 = 0$, $\sigma_1 = (1/s_0^2)\int_0^{s_0}(s'^2\psi_{01})\,ds'$.

Finally $\sigma_Z = \lim_{s_0\to\infty}\epsilon(s_0^2\sigma_1) = \epsilon\int_0^\infty(s'^2\psi_{01})\,ds' \equiv \epsilon\sigma_1(Z_0)$. Therefore, we write

$$\eth^2 Z_I = \sigma_B(Z_0,\zeta,\bar\zeta) - \sigma_1(Z_0), \tag{3.13}$$

and, solving for Z_I, we obtain the first nontrivial approximation

$$Z_1 = Z_0 + \epsilon Z_I, \tag{3.14}$$

from which (applying the procedure described in Sec. II) the metric can be obtained:

$$g'_{ab} = \eta_{ab} + \epsilon h'_{ab}. \tag{3.15}$$

Following a similar procedure as the one outlined above, we can find the higher order terms of Z. In this way Z can be obtained without the explicit knowledge of σ_Z.

(ii) We would like now to present a conjecture for the explicit choice of the function σ_Z. Its proof (or disproof) is for the future.

Beginning with H-space [$\bar H$-space] theory, we recall that solutions of equation

$$\eth^2 Z^H = \sigma_B(Z^H,\zeta,\bar\zeta), \tag{3.16a}$$

$$[\bar\eth^2 Z^{\bar H} = \bar\sigma_B(Z^H,\zeta,\bar\zeta)] \tag{3.16b}$$

yield a $Z^H(x^a,\zeta,\bar\zeta)$ $[Z^H(x^a,\zeta,\bar\zeta)]$ such that the resulting metric satisfies the vacuum Einstein equations and has a Weyl tensor which is self-dual [anti-self-dual], i.e., is an H-space [$\bar H$-space]. Assuming a solution of (3.16a) [(3.16b)] of the form $Z^H = Z^H(x^a,\zeta,\bar\zeta)$ $[Z^{\bar H} = \cdots]$, we can calculate $\bar\eth^2 Z^H$ and eliminate the x^a via $Z^H,\eth Z^H,\bar\eth Z^H,\eth\bar\eth Z^H$, yielding

$$\bar\eth^2 Z^H = -\bar\sigma^*(Z^H,\eth Z^H,\bar\eth Z^H,\eth\bar\eth Z^H,\zeta,\bar\zeta) \tag{3.17a}$$

and similarly

$$[\eth^2 Z^{\bar H} = -\sigma^*(Z^{\bar H},\eth Z^{\bar H},\cdots)]. \tag{3.17b}$$

The point to be emphasized is that $\bar{\sigma}^*$ $[\sigma^*]$ is a uniquely determined function of $Z^H, \eth Z^H, \bar{\eth} Z^H$, and $\eth\bar{\eth} Z^H$ obtained from σ_B. Our conjecture now is that

$$\sigma_Z(Z,\eth Z,\bar{\eth} Z,\eth\bar{\eth} Z,\zeta,\bar{\zeta}) \equiv \sigma^* \tag{3.18a}$$

$$[\bar{\sigma}_Z = \bar{\sigma}^*]; \tag{3.18b}$$

thus our *conjectured* vacuum Einstein equations are

$$\eth^2 Z = \sigma_B(Z,\zeta,\bar{\zeta}) - \sigma^*(Z,\eth Z,\bar{\eth} Z,\eth\bar{\eth} Z,\zeta,\bar{\zeta}), \tag{3.19a}$$

$$\bar{\eth}^2 Z = \bar{\sigma}_B(Z,\zeta,\bar{\zeta}) - \bar{\sigma}^*(Z,\eth Z,\bar{\eth} Z,\eth\bar{\eth} Z,\zeta,\bar{\zeta}). \tag{3.19b}$$

Several nice (required) features of (3.19) are (1) they satisfy the integrability conditions identically, (2) if either $\bar{\sigma}_B$ or σ_B were zero they would produce either $Z = Z_H$ *or* $Z = Z_{\bar{H}}$, and (3) if we consider the linearized version of (3.19), i.e., where we substitute on the right side of (3.19) the flat space $Z = Z_0 = x^a l_a$, the resulting $Z = Z_0 + Z_1$, $Z_1 = Z_1^H + Z_1^{\bar{H}}$ yields automatically the linearized vacuum solutions of the Einstein equations. This is easily seen to be a consequence of the fact we are simply superposing *linearized* self-dual and anti-self-dual metrics.

IV. CONCLUSION

To conclude this work, we would like to make a few comments intended to clarify our point of view and to express its connection with other works.

(1) Though we hope that sometime in the future we will be able to produce an equation for Z, i.e., the explicit form for σ_Z in

$$\eth^2 Z - \sigma_B(Z,\zeta,\bar{\zeta}) = -\sigma_Z,$$

which would include properties of sources obtained from a stress tensor, our point of view for the present is much more modest. We wish to find the form for σ_Z only for the pure vacuum case, in which case it will be completely determined by the characteristic data on future (or past) null infinities, namely the Bondi shear, $\sigma_B(u,\zeta,\bar{\zeta})$. The type of solutions we envisage are those analogous to the pure radiation (i.e., half retarded minus half advanced) solutions of the Maxwell equations. These solutions are completely determined by their characteristic data on $\mathscr{I}^+$. (There is a formulation of both Maxwell and Yang–Mills theory completely analogous to our present formulation of general relativity.[5])

(2) We also wish to point out the work we have discussed here is very much connected with the twistor program of Penrose.[15] The main point of contact appears to be the fact that the cut function Z can be expressed as the envelope of twistor lines in complexified $\mathscr{I}$. The connection appears in particular to be intimately associated with the so called "googly" graviton construction.

(3) As our final comment, we would like to say that one of our original motivations for this work (and the similar work for gauge theories[5]) was to try to formulate a (nonlinear) classical scattering theory, so that data given on $\mathscr{I}^-$ could be (somehow) directly translated into data on $\mathscr{I}^+$. An immediate desire having been to find the relationship between the two "Hilbert spaces" of $\mathscr{I}^+$ and $\mathscr{I}^-$ in the Ashtekar approach to quantum gravity.[16] A secondary motivation, which was much more nebulous, was to find a new variable for general relativity that was fundamentally associated with light cones for use in quantum gravity. The idea here is that quantum gravity should in some fashion "smear out" light cones. Z seems to be in some sense an ideal variable for this purpose. It certainly remains to be seen if we will be successful on any of these long range goals.

APPENDIX A

We want to derive here several of the formulae used earlier in the paper. Expanding $\Lambda_{,a}$ [Λ from Eq. (2.5)] in the basis θ^i_a, we obtain

$$\Lambda_{,a} = \Lambda_0\theta^0_a + \Lambda_+\theta^+_a + \Lambda_-\theta^-_a + \Lambda_1\theta^1_a \equiv \Lambda_i\theta^i_a, \tag{A1}$$

$$\bar{\Lambda}_{,a} = \bar{\Lambda}_0\theta^0_a + \bar{\Lambda}_+\theta^-_a + \bar{\Lambda}_-\theta^+_a + \bar{\Lambda}_1\theta^1_a \equiv \tilde{\Lambda}_i\theta^i_a. \tag{A2}$$

We now wish to obtain the expansion of $\bar{\eth}\Lambda_a$ and c.c. in terms of Λ_i and $\bar{\Lambda}_i$. Operating with $\bar{\eth}$ on Λ_a, $\eth$ on $\bar{\Lambda}_a$, we obtain

$$\begin{aligned}\bar{\eth}\Lambda_a = {} & (\bar{\eth}\Lambda_0 + \Lambda_-\bar{\Lambda}_0)\theta^0_a + (\bar{\eth}\Lambda_1 + \Lambda_+ + \bar{\Lambda}_1\Lambda_-)\theta^1_a \\ & + (\bar{\eth}\Lambda_+ + \Lambda_-\bar{\Lambda}_-)\theta^+_a \\ & + (\bar{\eth}\Lambda_- + \Lambda_0 - 2\Lambda_1 + \bar{\Lambda}_+\Lambda_-)\theta^-_a + \Lambda_1\eth\bar{\Lambda}_a\end{aligned}$$

and c.c.

Solving this pair of equations (note the $\eth\bar{\Lambda}_a$ on right side) for $\bar{\eth}\Lambda_a$ and $\eth\bar{\Lambda}_a$, we can write the solution as

$$\begin{aligned}\bar{\eth}\Lambda_a/P = {} & (\rho + \Lambda_1\bar{\rho})\theta^0_a + (\delta + \Lambda_1\bar{\beta})\theta^+_a \\ & + (\beta + \Lambda_1\bar{\delta})\theta^-_a + (\eta + \Lambda_1\bar{\eta})\theta^1_a\end{aligned} \tag{A3}$$

and c.c. (A4)

with $P \equiv (1 - \Lambda_1\bar{\Lambda}_1)^{-1}$ and

$$\eta = \bar{\eth}\Lambda_1 + \Lambda_+ + \bar{\Lambda}_1\Lambda_-, \tag{A5}$$

$$\rho = \bar{\eth}\Lambda_0 + \Lambda_-\bar{\Lambda}_0, \tag{A6}$$

$$\delta = \bar{\eth}\Lambda_+ + \Lambda_+\bar{\Lambda}_+, \tag{A7}$$

$$\beta = \Lambda_0 + \bar{\eth}\Lambda_- - 2\Lambda_1 + \bar{\Lambda}_+\Lambda_-. \tag{A8}$$

Using Eqs. (A3)–(A8), we define $\Pi_i, \tilde{\Pi}_i$ the tetrad components of $\bar{\eth}\Lambda_{,a}$ and $\eth\bar{\Lambda}_a$ by

$$\bar{\eth}\Lambda_a = \Pi_i\theta^i_a, \tag{A9}$$

$$\eth\bar{\Lambda}_a = \tilde{\Pi}_i\theta^i_a. \tag{A10}$$

(Notice $\tilde{\Pi}_i = \Pi_i$ for $i = 0,1$ and $\tilde{\Pi}_+ = \bar{\Pi}_-$, $\tilde{\Pi}_- = \bar{\Pi}_+$. The same rule applies for $\bar{\Lambda}_i$ and $\tilde{\Lambda}_i$.) The coefficients $\Lambda_i, \tilde{\Lambda}_i$, $\Pi_i, \tilde{\Pi}_i$ are used to define the nontrivial components of $\Delta^i_j, \tilde{\Delta}^i_j$ [Eq. (2.15)], namely

$$\Delta^0_i = \delta_i{}^+, \quad \Delta_i{}^+ = \Lambda_i, \tag{A11a}$$

$$\Delta_i^- = \delta_i^1,\ \Delta_i^1 = \Pi_i - 2\delta_i^+,$$
$$\tilde{\Delta}_i^0 = \delta_i^-,\ \tilde{\Delta}_i^+ = \delta_i^1,$$
$$\tilde{\Delta}_i^- = \tilde{\Lambda}_i,\ \tilde{\Delta}_i^1 = \tilde{\Pi}_i - 2\delta_i^-. \tag{A11b}$$

Finally the explicit form of q^{11} [Eq. (M.10)] is given by

$$q^{11}/q^{01} = -2[1 - \mathrm{Re}(\Lambda_0\bar{\Lambda}_1)]/(1 - \Lambda_1\bar{\Lambda}_1)$$
$$\{ -\mathrm{Re}[\bar{\eth}(\bar{\eth}\Lambda_1 + \Lambda_1\bar{W})$$
$$+ (\bar{\eth}\Lambda_1 + \Lambda_1\bar{W})(\bar{W} + \bar{\Lambda}_1\Lambda_+ + \Lambda_1\bar{\Lambda}_-)$$
$$+ 2\Lambda_1\Lambda_+\bar{\Lambda}_- + \Lambda_+\bar{\Lambda}_+ + \Lambda_-\bar{\Lambda}_-]\}/(1 - \Lambda_1\bar{\Lambda}_1) \tag{A12}$$

APPENDIX B

We want to give here an alternative proof of Eq. (3.1) using the intrinsic geometry of $\mathscr{I}^+$.

We first summarize the intrinsic structure of $\mathscr{I}$.

(i) Since the restriction $\mathbf{n}^a$ of $n^a = g^{ab}\Omega_{,b}$ to $\mathscr{I}$ is complete and the manifold of orbits of $\mathbf{n}^a$ is diffeomorphic to S^2, $\mathscr{I}$ is a trivial bundle with base space S^2 and fiber R.

(ii) The restriction $\mathbf{g}_{ab}$ is a degenerate metric with $\mathbf{g}_{ab}V^a = 0$ iff $V^b = \alpha\mathbf{n}^b$; therefore, $\mathbf{g}_{ab} = \Pi^*(h_{ab})$, where Π^* is the lift operation to $\mathscr{I}$ and h_{ab} is the metric of S^2.

(iii) It is useful to introduce the tetrad fields of $\mathscr{I}$ in the following way:

(a) Choose an arbitrary cross section of $\mathscr{I}$, then slide it up and down by the integral curves of $\mathbf{n}^a$. This yields cross sections $u = \text{const}$ such that $\mathscr{L}_{\mathbf{n}}u = 1$ or defining $\mathbf{l}_a \equiv u_{,a}$ we have $\mathbf{n}^a\mathbf{l}_a = 1$. (b) Introduce a complex vector field $\mathbf{m}^a$ tangent to the cross sections such that $\mathscr{L}_{\mathbf{n}}\mathbf{m}^a = 0$, $\mathbf{m}^a\mathbf{m}_a = 0$, $\mathbf{m}^a\bar{\mathbf{m}}_a = -1$ with $\mathbf{m}_a = \mathbf{g}_{ab}\mathbf{m}^b$. Coordinatizing the cross sections with $\zeta,\bar{\zeta}$, we can set $\mathbf{m}_a = (\sqrt{2P})^{-1}\zeta_{,a}$, $\mathbf{m}^a = \sqrt{2P}\,\partial/\partial\bar{\zeta}$, $g_{ab} = \zeta_{,(a}\bar{\zeta}_{,b)}/P^2$ with $P = P(\zeta,\bar{\zeta})$. Then $(\mathbf{n}^a,\mathbf{m}^a,\bar{\mathbf{m}}^a)$ and $(\mathbf{l}_a,\mathbf{m}_a\bar{\mathbf{m}}_a)$ constitute the tetrad fields of $\mathscr{I}$.

(iv) Finally it can be shown[17] that the intrinsic connection D_a can be completely described if we give its action on $\mathbf{l}_a$ and $\mathbf{m}_a$:

$$D_a\mathbf{l}_b = \sigma_B\bar{\mathbf{m}}_a\bar{\mathbf{m}}_b + \bar{\sigma}_B\mathbf{m}_a\mathbf{m}_b + \tfrac{1}{2}\mathbf{g}_{ab}\gamma$$
$$\equiv \sigma^B_{ab} + \tfrac{1}{2}\mathbf{g}_{ab}\gamma \text{ with } \sigma_B = \sigma_B(u,\zeta,\bar{\zeta}),$$
$$D_a\mathbf{m}_b = \Pi^*(\hat{D}_a\hat{m}_b), \text{ where } \hat{D}_a \text{ is the connection of } S^2$$
and $h_{ab} = 2\hat{m}_{(a}\hat{\bar{m}}_{b)}$.

We now present the geometric objects we are interested in. Assume we introduce a cross section on $\mathscr{I}$ which is locally described by $u' = 0$. Since u' is not unique, we want to obtain a canonical choice. If we use coordinates $(u,\zeta,\bar{\zeta})$, then the cross section is characterized by the equation $u = \alpha(\zeta,\bar{\zeta})$; therefore, the canonical u' is given by $u' \equiv u - \alpha(\zeta,\bar{\zeta})$ and

$$u' = 0 \Leftrightarrow u = \alpha(\zeta,\bar{\zeta}). \tag{B1}$$

Denoting by u'_a the gradient of u', we can easily see

$$D_a u'_b = D_a\mathbf{l}_b - D_aD_b\alpha \tag{B2}$$

or, taking the trace-free part of this equation, we obtain at $u' = 0$

$$(D_aD_b - \tfrac{1}{2}\mathbf{g}_{ab}D^cD_c)\alpha = \sigma^B_{ab} - \sigma^\alpha_{ab}, \tag{B2}$$

where σ_{ab} is the trace-free part of $D_a u'_b$ or $D_a l_b$ and is usually called the shear tensor associated with each cut. Multiplying by $\mathbf{m}^a\mathbf{m}^b$ (and $\bar{\mathbf{m}}^a\bar{\mathbf{m}}^b$) and defining $\sigma_\alpha = \mathbf{m}^a\mathbf{m}^bD_a u'_b$ (and c.c.), we get

$$\mathbf{m}^a\mathbf{m}^bD_aD_b\alpha = \sigma_B - \sigma_\alpha \text{ (and c.c.)}. \tag{B3}$$

Since $\alpha = \alpha(\zeta,\bar{\zeta})$, then by defining the $\eth,\bar{\eth}$ operators in the usual way we obtain the known result[11]

$$\eth^2\alpha = \sigma_B(\alpha,\zeta,\bar{\zeta}) - \sigma_\alpha \text{ (and c.c.)}. \tag{B4}$$

[1] R. Penrose, *Battelle Recontres*, edited by B. de Witt and J. A. Wheeler (Benjamin, New York, 1968), p. 121.
[2] R. Geroch, *Asymptotic Structure of Space-Time*, edited by P. Esposito and L. Witten (Plenum, New York, 1976).
[3] H. Bondi, Proc. Roy. Soc. (London) A **269**, 21 (1962).
[4] E. T. Newman and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[5] S. L. Kent and E. T. Newman, J. Math. Phys. **24**, 949 (1983).
[6] P. Joshi, C. N. Kozameh, and E. T. Newman, J. Math. Phys. **24**, 2490 (1983).
[7] M. Ko, M. Ludvigsen, E. T. Newman, and K. P. Tod, "The Theory of *H*-space," Phys. Rep. **71**, 53 (1981).
[8] R. Hansen, E. T. Newman, R. Penrose, and K. P. Tod, Proc. Roy. Soc. (London) A **363**, 445 (1978).
[9] Another reconstruction technique can be found in F. A. Pirani and A. Schild, *Perspectives in Geometry and Relativity*, edited by B. Hoffman (Indiana U. P., Bloomington, 1966), p. 291.
[10] M. Ko, E. T. Newman, and K. P. Tod, *Asymptotic Structure of Space-Time* edited by P. Esposito and L. Witten (Plenum, New York, 1976).
[11] R. K. Sachs, Proc. Roy. Soc. (London) **270**, 103 (1962).
[12] R. K. Sachs, Proc. Roy. Soc. (London) A **264**, 309 (1961).
[13] S. Hawking and G. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973).
[14] The gauge freedom in the solution of the Bianchi identities is fixed by allowing only half advanced minus half retarded solutions.
[15] R. Penrose and M. A. McCallum, "Twistor Theory," Phys. Rep. **6**, 242 (1973).
[16] Abhay Ashtekar, J. Math. Phys. **22**, 2885 (1981).
[17] C. N. Kozameh and E. T. Newman, Gen. Rel. Grav. **15**, 475 (1983).

Light cone cuts of null infinity in Schwarzschild geometry[a)]

P. S. Joshi, C. N. Kozameh, and E. T. Newman
Department of Physics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 1 March 1983; accepted for publication 3 June 1983)

Light cone cuts of future null infinity in Schwarzschild geometry are studied here. The future null cone from an arbitrary apex in the space-time has been constructed, and its intersection with $\mathscr{I}^+$ is obtained. Knowledge of the cuts yields a great deal of information about the interior of the space-time. In particular, we use it to reconstruct the Schwarzschild metric up to a conformal factor.

PACS numbers: 04.20. − q, 04.30. + x

I. INTRODUCTION

The purpose of this paper is twofold. Working with Schwarzschild geometry we are first of all interested in constructing (by integrating the null geodesic equations) the (future) null cones N_x from arbitrary space-time points x^a and in particular in finding the intersections C_x of these cones with future null infinity, i.e., $C_x = N_x \cap \mathscr{I}^+$. These intersections or light cone cuts of $\mathscr{I}^+$ are (locally) 2-surfaces, which are uniquely labeled by the apex of the cone, i.e., x^a. If Bondi coordinates $(u,\zeta,\bar{\zeta})$ are used to coordinatize $\mathscr{I}^+$, the cuts can be given in the form $u = Z(x^a,\zeta,\bar{\zeta})$. It is clear that knowledge of the set of all these cuts yields a great deal of information about the interior of the space-time. In particular, for the second purpose, we will show that from the $Z(x^a,\zeta,\bar{\zeta})$ one can reconstruct the Schwarzschild metric up to a conformal factor. This result is a particular example of a general theory of light cone cuts for asymptotically flat space-times which has recently been developed.[1]

In Sec. II we give a brief review of the general theory of extracting the metric from the $Z(x^a,\zeta,\bar{\zeta})$. In Sec. III the null geodesic equations are integrated and the cut function $u = Z(x^a,\zeta,\bar{\zeta})$ is obtained. Unfortunately, the function is sufficiently complicated so that it must be given parametrically and in terms of elliptic integrals. In Sec. IV we discuss the reconstruction of the Schwarzschild metric from the Z while in Sec. V we describe several examples of other information that can be extracted from the Z.

II. THEORY OF LIGHT CONE CUTS

Given an asymptotically flat space[2] M with (conformal) metric g_{ab} and null boundary $\mathscr{I}^+$, we can (in principle) integrate the null geodesic equations and obtain all the null geodesics in M.

In particular, we can construct the null cone N_x with apex x^a. N_x is a collection of points that, except for conjugate points, is surface-forming when considered locally.

N_x can also be described by the equation

$$L(x^a,x'^a) = 0, \tag{2.1}$$

where x^a is the apex of the cone and x'^a are points on the null surface. [Notice that fixing x'^a the set of points x^a that satisfy Eq. (2.1) form the null cone emanating from x'^a].

We now define the light cone cut C_x as $C_x \equiv N_x \cap \mathscr{I}^+$. In terms of the function L, the cuts are described by

$$L(x^a, \mathbf{x}'^a) = 0, \tag{2.2}$$

where we use the notation $\mathbf{x}'^a$ when x^a is constrained to lie on $\mathscr{I}^+$.

Locally the light cone cuts will be 2-surfaces and therefore if Bondi coordinates[3] $(u,\zeta,\bar{\zeta})$ are used to coordinatize $\mathscr{I}^+$ we can solve for u in Eq. (2.2) and write locally

$$u = Z(x^a,\zeta,\bar{\zeta}). \tag{2.3}$$

We can think of $Z(x^a,\zeta,\bar{\zeta})$ in three different ways

(i) For fixed x^a, Eq. (2.3) describes the cut C_x; therefore, varying x^a, we obtain a four-parameter familty of cuts.

(ii) $Z(x^a,\zeta,\bar{\zeta})$ can also be thought of as a function on the bundle of null directions at points x^a of M, with ζ, $\bar{\zeta}$ of $\mathscr{I}^+$ parametrizing the null directions.

(iii) For fixed $(u,\zeta,\bar{\zeta})$ on $\mathscr{I}^+$ the points x^a that satisfy Eq. (2.3) form the past null cone from a point of $\mathscr{I}^+$.

According to the third statement, $Z_{,a}(x,\zeta,\bar{\zeta})$ is a null vector, and therefore it obeys

$$g^{ab}(x)Z_{,a}(x,\zeta,\bar{\zeta})Z_{,b}(x,\zeta,\bar{\zeta}) = 0. \tag{2.4}$$

Notice that for fixed x^a if we vary $(\zeta,\bar{\zeta})$, $Z_{,a}$ sweeps the tangent null cone of x^a, and this in turn enables us to construct a metric conformally related to g^{ab}.

Our reconstruction technique consists in applying the different operators[4] $\eth$ and $\bar{\eth}$ several times to Eq. (2.4) until enough equations are obtained so that the metric can be written explicitly in terms of the gradient basis $Z_{,a}, \eth Z_{,a}, \bar{\eth} Z_{,a}, \eth\bar{\eth} Z_{,a}$.

In order to simplify the notation we define $\theta^i_{\ a}$ [i tetrad index $(0, +, -, 1)$ a tensor index] as follows:

$$\theta^0_{\ a} = Z_{,a}, \quad \theta^+_{\ a} = \eth Z_{,a}, \quad \theta^-_{\ a} = \bar{\eth} Z_{,a}, \quad \theta^1_{\ a} = \eth\bar{\eth} Z_{,a}. \tag{2.5}$$

In terms of this basis we define the tetrad components of the metric

$$g^{ij} = g^{ab}\theta^i_{\ a}\theta^j_{\ b}.$$

Equation (2.4) reads in this new notation

$$g^{ab}\theta^0_{\ a}\theta^0_{\ b} = g^{00} = 0. \tag{2.6}$$

Applying $\eth$ and $\bar{\eth}$ to g^{00}, we obtain

$$g^{ab}\theta^0_{\ a}\theta^+_{\ b} = g^{0+} = 0, \tag{2.7}$$

a) Supported by NSF Grant PHY800823.

 0022-2488/83/102490-08$02.50

$$g^{ab}\theta^0_a\theta_b^- = g^{0-} = 0. \tag{2.8}$$

Operating on (2.7) with $\bar{\delta}$ or (2.8) with δ, we obtain

$$g^{+-}/g^{01} = -1. \tag{2.9}$$

In this way the trivial components of g^{ab} have been obtained. New δ and $\bar{\delta}$ operations will provide the six other components we are looking for.

We want to stress the importance of $\delta^2 Z$ (and c.c.) since this complex function alone allows us to determine the remaining g^{ij}'s. First we expand $\delta^2 Z_{,a}$ in terms of θ^i_a as

$$\delta^2 Z_{,a} = \Lambda_0\theta^0_a + \Lambda_+\theta^+_a + \Lambda_-\theta^-_a + \Lambda_1\theta^1_a \equiv \Lambda_i\theta^i_a. \tag{2.10}$$

[We emphasize that if Z is known, then so are $\delta^2 Z_{,a}$ and θ^i_a and hence (2.10) leads to an algebraic determination of the Λ_i.]

We then write the following set of equations:

$$Z^a\delta^2 Z_a = -g^{++} = g^{01}\Lambda_1, \tag{2.11}$$

$$\delta Z^a\delta^2 Z_a = \delta(g^{++}) = g^{++}\Lambda_+ + g^{+-}\Lambda_- + g^{+1}\Lambda_1, \tag{2.12}$$

$$\bar{\delta}Z^a\delta^2 Z_a = -\delta(g^{01}) - g^{+1} = g^{+-}\Lambda_+ + g^{--}\Lambda_- + g^{+1}\Lambda_1, \tag{2.13}$$

and c.c.

The right side of these equations are just the contractions of (2.10) with $Z^a, \delta Z^a$, and $\bar{\delta}Z^a$ whereas the center is obtained from the extreme left by commuting δ derivatives and using $g^{0+} = g^{0-} = 0$ and $\delta g^{ab} = 0$.

Equation (2.11) and its conjugate yield g^{++}/g^{01} and g^{--}/g^{01}. Now by inserting (2.11) and c.c. into (2.12), (2.13), and c.c. we obtain a set of four algebraic equations from which g^{+1}/g^{01}, g^{-1}/g^{01}, $\delta g^{01}/g^{01}$, and $\bar{\delta}g^{01}/g^{01}$ can be obtained.

The last component g^{11}/g^{01} could be derived from $\delta\bar{\delta}Z^a\delta^2 Z_a$. However, it is simpler to obtain it from $\delta^2 Z^a\bar{\delta}^2 Z_a$. By multiplying (2.10) with its conjugate and some manipulation, we obtain

$$\bar{\delta}^2 Z^a\delta^2 Z_a = g^{11} + 2g^{01} - [\bar{\delta}(g^{+1}) + \delta(g^{-1}) + \delta\bar{\delta}(g^{01})] = g^{ij}\Lambda_i\tilde{\Lambda}_j, \tag{2.14}$$

with $\tilde{\Lambda}_i = \bar{\Lambda}_i$ for $i = 0, 1$, $\tilde{\Lambda}_+ = \bar{\Lambda}_-$, and $\tilde{\Lambda}_- = \bar{\Lambda}_+$. Since the remaining components have been obtained, we can algebraically solve for g^{11}/g^{01}. We would like at this point to make two comments.

(i) The explicit form of g^{01} is irrelevant since we are looking for a conformal metric, that is, for fixed $(\zeta,\bar{\zeta})$, g^{01} becomes an overall conformal factor.

(ii) There is a one-to-one algebraic relationship between $(1/g^{01})(g^{++}, g^{--}, g^{+1}, g^{-1}, g^{11}, \delta g^{01}, \bar{\delta}g^{01}) \equiv q_{ij}$ and $(\Lambda_1,\bar{\Lambda}_1,\Lambda_+,\bar{\Lambda}_+,\Lambda_-,\bar{\Lambda}_-,\Lambda_0\bar{\Lambda}_1 + \Lambda_0\Lambda_1) \equiv \lambda_\mu$ via Eqs. (2.11)–(2.14). This correspondence can be used in two ways: (a) to obtain q^{ij} if λ_μ is known as it is our case or (b) to obtain λ_μ if q^{ij} is the data. This property will be used in Sec. IV.

The coordinate components of the metric are now obtained from $g^{ab} = g^{ij}\theta^a_i\theta^b_j$ with g^{ij} of (2.6)–(2.14) and θ^a_i satisfying $\theta^a_i\theta^i_b = \delta^a_b$. [Notice that if we fix $(\zeta,\bar{\zeta})$, θ^i_a is an ordinary coordinate basis.]

III. ASYMPTOTIC CUT DESCRIPTION FOR SCHWARZSCHILD GEOMETRY

Schwarzschild space-time is among the simplest nontrivial asymptotically flat solutions to Einstein's equations. It was therefore natural to begin the study of light cone cuts with this example. Using the known Schwarzschild metric, knowledge of all null geodesics of the space-time can be obtained. This provides a complete description of the light cone from an arbitrary apex in the space-time. The intersection of this cone with $\mathscr{I}^+$ generates the light cone cut. The nontrivial feature represented by the light cones in this geometry, as compared to flat space, is the angular deflection of null rays because of the curvature. As a result, the cut function Z in this geometry is described by rather complicated elliptic integrals and must be given parametrically as opposed to the elementary functions obtained in the case of a flat space-time.

Before considering the Schwarzschild situation in detail, we will discuss briefly the light cone cuts of infinity in flat space-times. This permits special insights into the situation by virtue of its simplicity and clarity.

Throughout our discussion (for both the flat and Schwarzschild cases) we shall use the null polar coordinates $(u,r,\zeta,\bar{\zeta})$, where ζ and $\bar{\zeta}$ are complex stereographic coordinates on the sphere defined by $\zeta = e^{i\phi}\cot(\theta/2)$. Let x^a be the Minkowskian coordinates (t,x,y,z) for an arbitrary apex, then the null cone about $\mathring{x}^a$ can be described as

$$u = u(\mathring{x}^a,\zeta,\bar{\zeta},r), \tag{3.1}$$

which, in the limit as $r\to\infty$ becomes

$$u = Z(\mathring{x}^a,\zeta,\bar{\zeta}). \tag{3.2}$$

To work out Z for the flat space-time situation, we rewrite the usual Minkowskian metric with signature $(+,-,-,-)$ in the null polar coordinates $(u,r,\zeta,\bar{\zeta})$ [Note that when a two-point function, e.g., (3.2) is used the interior points are denoted with $u_0,r_0,\zeta_0,\bar{\zeta}_0$]:

$$ds^2 = du^2 + 2\,du\,dr - r^2\frac{d\zeta\,d\bar{\zeta}}{P_0^2}, \tag{3.3a}$$

where $u = t - r$ is the retarded time and $P_0 = (1 + \zeta\bar{\zeta})/2$. We shall introduce here new coordinates[5]

$$u' = (1/\sqrt{2})u, \quad r' = \sqrt{2}r, \tag{3.3b}$$

which are more convenient for the study of asymptotic structure. Then (3.3a) can be written as

$$ds^2 = 2du^2 + 2\,du\,dr - \frac{1}{2}r^2\frac{d\zeta\,d\bar{\zeta}}{P_0^2}, \tag{3.3c}$$

where we have suppressed the primes. Since null geodesics are conformally invariant, there is an inherent conformal freedom available and we shall transform (3.3c) to

$$d\hat{s}^2 = \Omega^2 ds^2 = 4l^2 du^2 - 4\,du\,dl - \frac{d\zeta\,d\bar{\zeta}}{P_0^2}, \tag{3.4}$$

where we have introduced a new coordinate $l = r^{-1}$ and the

conformal factor has been chosen to be $\Omega = \sqrt{2}l$. Working out the geodesic equations of the space-time, we have (with the dot denoting derivative with respect to s, the affine parameter),

$$\begin{aligned} &2l^2\dot{u} - \dot{l} = 1, \\ &\ddot{u} + l\dot{u}^2 = 0, \\ &\ddot{\zeta}(1+\zeta\bar{\zeta}) - 2\bar{\zeta}\dot{\zeta}^2 = 0, \\ &\ddot{\bar{\zeta}}(1+\zeta\bar{\zeta}) - 2\dot{\zeta}\dot{\bar{\zeta}}^2 = 0, \\ &4l^2\dot{u}^2 - 4\dot{u}\dot{l} - \dot{\zeta}\dot{\bar{\zeta}}/P_0^2 = 0, \end{aligned} \tag{3.5}$$

where the last equation corresponds to $ds^2 = 0$. Restricting ourselves momentarily to the equatorial plane $\theta = \pi/2$ for the sake of simplicity, (3.5) can be written as

$$2l^2\dot{u} - \dot{l} = 1, \tag{3.6a}$$

$$\ddot{u} + l\dot{u}^2 = 0, \tag{3.6b}$$

$$\ddot{\varphi} = 0, \quad \dot{\varphi} = b, \tag{3.6c}$$

$$l^2\dot{u}^2 - \dot{u}\dot{l} = \dot{\varphi}^2 = b^2/4. \tag{3.6d}$$

We shall now eliminate the parameter s from above. For that, substituting (3.6a) in (3.6d) and then solving both these equations for u gives $l^2 = 1 - b^2l^2$, i.e.,

$$\dot{l} = \pm\sqrt{1 - b^2l^2}, \tag{3.7a}$$

$$ds = \pm dl/\sqrt{1 - b^2l^2}. \tag{3.7b}$$

{It should be noted that $\dot{l} < 0$ corresponds to a null ray moving away from the origin [$\dot{l} = -(1/r^2)\dot{r}$]. Next, if $\dot{l} > 0$ initially, then the ray moves initially towards the origin of the coordinate system ($r = 0$, $l = \infty$) and after reaching a minimum r_m (i.e., $\dot{l} = \sqrt{1 - b^2l_m^2} = 0$) it begins to move outwards and again $\dot{l} < 0$. For the sake of definiteness we shall choose here rays such that initially $\dot{l} < 0$; however, by considering the other sheet $\dot{l} > 0$ as well, we can span the full light cone of null rays from our starting point. We shall return to this point later.}

Next, using (3.6a) and (3.7a), and (3.7b), we can write

$$du = -\frac{dl}{2l^2\sqrt{1 - b^2l^2}} + \frac{dl}{2l^2}, \tag{3.7c}$$

and, from (3.6c) and (3.7b), we have

$$d\varphi = \frac{-b\,dl}{\sqrt{1 - b^2l^2}}. \tag{3.7d}$$

If the apex of the cone is at $l = l_0$, $u = u_0$, $\phi = \phi_0$, integrating (3.7c) and (3.7d) from l_0 to an arbitrary l gives the equations for (one sheet of) the light cone as

$$u - u_0 = -\frac{1}{2}\int_{l_0}^{l}\frac{dl}{l^2\sqrt{1 - b^2l^2}} + \frac{1}{2}\int_{l_0}^{l}\frac{dl}{l^2}, \tag{3.8a}$$

$$\varphi - \varphi_0 = \int_{l_0}^{l}\frac{-b\,dl}{\sqrt{1 - b^2l^2}}, \tag{3.8b}$$

where b (the initial direction) ranges from 0 to l_0^{-1}. For the sake of simplicity we choose the apex on the $\phi = 0$ axis, i.e., $\phi_0 = 0$. Now by taking the limit as the null rays escape to infinity, i.e., $l \to 0$, (3.8a) and (3.8b) provide us with a portion of the cut at infinity of the rays coming from u_0, l_0, ϕ_0. [At the moment we are describing only the rays in the equatorial plane.] Integration of (3.8) from l_0 to 0 yields

$$u - u_0 = \frac{1 - \sqrt{1 - b^2l_0^2}}{2l_0} \tag{3.9}$$

and

$$\varphi = \arcsin(bl_0). \tag{3.10}$$

Note that for fixed apex (3.10) yields a one-to-one relation between the initial direction b and the final angular position ϕ on $\mathscr{I}^+$. By eliminating b from (3.9) and (3.10), we obtain the portion (equatorial plane) of the light cone cut,

$$u = u_0 + (1/2l_0)(1 - \cos\varphi). \tag{3.11}$$

We emphasize (for later use) that had (3.9) and (3.10) been functionally more complicated, we might have to consider them as defining (3.11) parametrically. (Note that for the sheet $\dot{l} < 0$ which we have been considering, $\cos\phi$ will be positive because ϕ will always be in the first or fourth quadrant in this situation. For the other sheet corresponding to $\dot{l} > 0$ initially, $\cos\phi$ will be negative.)

The portion of Z given by (3.10) describes, as we have mentioned, only an S^1 worth of null rays intersecting $\mathscr{I}^+$, since we have restricted ourselves to the equatorial plane. However, because of spherical symmetry the full cut, which is topologically S^2, can be generated by rotating this plane.

For that, first it will be convenient to define a null vector l^a [for all values of the stereographic coordinates $(\zeta,\bar{\zeta})$] given by

$$l^a = \frac{1}{\sqrt{2}}\left(1, \frac{\zeta + \bar{\zeta}}{1 + \zeta\bar{\zeta}}, \frac{i(\bar{\zeta} - \zeta)}{1 + \zeta\bar{\zeta}}, \frac{\zeta\bar{\zeta} - 1}{1 + \zeta\bar{\zeta}}\right). \tag{3.12}$$

As $(\zeta,\bar{\zeta})$ more over S^2, l^a sweeps out the null cone of directions.

The situation we have considered is that of all null geodesics in the equatorial plane with apex on the $\phi = 0$ axis. Now consider two unit vectors at the origin, one pointing to the apex and the other to the "final point" of the geodesic, i.e., $\vec{n}_A = (1,0,0)$, $\vec{n}_F = (\cos\phi, \sin\phi, 0)$. We have $\vec{n}_A \cdot \vec{n}_F = \cos\phi$. If now we perform an arbitrary rigid rotation so that the apex is moved to a new direction

$$n'_A = (\sin\theta_0\cos\varphi_0,\ \sin\theta_0\sin\varphi_0, \cos\theta_0) \tag{3.13a}$$

and the final point to

$$n'_F = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta), \tag{3.13b}$$

then we have $\vec{n}_A \cdot \vec{n}_F = \vec{n}'_A \cdot \vec{n}'_F$ or

$$\cos\hat{\phi} = \cos\theta_0\cos\theta + \sin\theta_0\sin\theta\cos(\varphi - \varphi_0), \tag{3.14a}$$

where we have now used $\hat{\phi}$ to designate the angle between $\vec{n}_A$ and $\vec{n}_F$. Using stereographic coordinates $(\zeta,\bar{\zeta})$ and $(\zeta_0,\bar{\zeta}_0)$ instead of the θ and ϕ 's and (3.12), we have [using

$$(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) = \left(\frac{\zeta + \bar{\zeta}}{1 + \zeta\bar{\zeta}}, \frac{i(\bar{\zeta} - \zeta)}{1 + \zeta\bar{\zeta}}, \frac{\zeta\bar{\zeta} - 1}{1 + \zeta\bar{\zeta}}\right)\Big],$$

$$\cos\hat{\phi} = 1 - 2l^a\mathring{l}_a, \tag{3.14b}$$

where $\mathring{l}_a$ is the same as (3.12) but with $(\zeta_0,\bar{\zeta}_0)$ instead of $(\zeta,\bar{\zeta})$. Substituting this back in (3.11) yields

$$u = u_0 + \frac{1}{l_0}(l^a \mathring{l}_a) = Z(u_0, l_0, \zeta_0, \bar{\zeta}_0, \zeta, \bar{\zeta}), \tag{3.15}$$

the general equation for the light cone cut for Minkowski space where the apex has coordinates $(u_0, l_0, \zeta_0, \bar{\zeta}_0)$.

If one chooses to use Minkowski coordinates x^a, which are related to the null polar coordinates $(u, r, \zeta, \bar{\zeta})$ by

$$\mathring{x}^a = u_0 t^a + r_0 \mathring{l}^a(\zeta_0, \bar{\zeta}_0), \tag{3.16}$$

where t^a is a constant vector with $t^a t_a = 2$ and $t^a l_a = 1$, then the light cone cut (3.15) can also be written as

$$u = \mathring{x}^a l_a(\zeta, \bar{\zeta}), \tag{3.17}$$

a well-known result.[5]

Having considered the flat space-time situation, we now return to Schwarzschild geometry.

The Schwarzschild metric in (u, r, θ, ϕ) coordinates, where $u = t - r - 2m\log(r - 2m)$ is the retarded time, is given as

$$ds^2 = (1 - 2m/r)du^2 + 2\,du\,dr - r^2(d\theta^2 + \sin^2\theta\, d\varphi^2). \tag{3.18}$$

As before, we make coordinate transformations (3.3b), use stereographic coordinates $\zeta, \bar{\zeta}$, and conformally transform the metric by $\Omega = r^{-1} = \sqrt{2}l$ which gives

$$d\hat{s}^2 = \Omega^2 ds^2 = 4(l^2 - 2\sqrt{2}ml^3)\,du^2 - 4\,du\,dl - \frac{d\zeta\, d\bar{\zeta}}{P_0^2}. \tag{3.19}$$

The new coordinate l is now finite at infinity and $\mathscr{I}^+$ is described by the hypersurface $l = 0$, which corresponds to $r = \infty$ for (3.18).

The Lagrangian for the geodesics is written as

$$\mathscr{L} = 2(l^2 - 2\sqrt{2}ml^3)\dot{u}^2 - 2\dot{u}\dot{l} - \dot{\zeta}\dot{\bar{\zeta}}/2P_0^2, \tag{3.20}$$

where dot denotes differentiation w.r.t. an affine parameter s along null geodesics. The equations for null geodesics are then given as:

$$2(l^2 - 2\sqrt{2}ml^3)\dot{u} - \dot{l} = 1, \tag{3.21a}$$

$$\ddot{u} + 2(l - 3\sqrt{2}ml^2)\dot{u} = 0, \tag{3.21b}$$

$$\ddot{\zeta}(1 + \zeta\bar{\zeta}) - 2\bar{\zeta}\dot{\zeta}^2 = 0, \tag{3.21c}$$

$$\ddot{\bar{\zeta}}(1 + \zeta\bar{\zeta}) - 2\zeta\dot{\bar{\zeta}}^2 = 0, \tag{3.21d}$$

$$4(l^2 - 2\sqrt{2}ml^3)\dot{u}^2 - 4\dot{u}\dot{l} - \dot{\zeta}\dot{\bar{\zeta}}/P_0^2 = 0, \tag{3.21e}$$

where (3.21e) corresponds to $ds^2 = 0$. Though, in principle, all the null geodesics of the space-time are obtained from (3.21), we begin our investigation with those in the equatorial plane $\theta = \pi/2$. From a fixed apex this yields an S^1 worth of geodesics. Now from these and the spherical symmetry of the situation we will generate *all* the null geodesics from an *arbitrary* apex by a rigid rotation as was done in the Minkowski case. For $\theta = \pi/2$ we have $\zeta = e^{i\phi}$, and (3.21) becomes

$$2(l^2 - 2\sqrt{2}ml^3)\dot{u} - \dot{l} = 1, \tag{3.22a}$$

$$\ddot{u} + 2(l - 3\sqrt{2}ml^2)\dot{u}^2 = 0, \tag{3.22b}$$

$$\ddot{\varphi} = 0, \quad \dot{\varphi} = b, \tag{3.23}$$

$$(l^2 - 2\sqrt{2}ml^3)\dot{u}^2 - \dot{u}\dot{l} = \dot{\varphi}^2 = b^2/4, \tag{3.24}$$

with (3.22b) an identity from the other equations. Combining the first and last equations of above, the following relations can be easily deduced:

$$\dot{u} = \frac{1 + \dot{l}}{2(l^2 - 2\sqrt{2}ml^3)}, \tag{3.25a}$$

$$\dot{l} = \pm(2\sqrt{2}mb^2l^3 - b^2l^2 + 1)^{1/2} = \pm\sqrt{A}, \tag{3.25b}$$

$$ds = \pm\frac{dl}{\sqrt{A}}, \tag{3.25c}$$

$$d\varphi = \pm\frac{b\,dl}{\sqrt{A}}. \tag{3.25d}$$

Before integrating (3.25), we note that the null rays coming from an arbitrary apex are divided into two sets (the two sheets of A) defined by (3.25b), i.e., those given initially by $\dot{l} < 0$ and $\dot{l} > 0$. For the first set, i.e., those initially with $\dot{l} < 0$, the geodesics continue with a decreasing l (increasing r) until intersection with $\mathscr{I}^+$. For the rays which begin with $\dot{l} > 0$ (the second set), i.e., those rays with initially increasing l (decreasing r), some reach a maximum l (when $A = 0$), then begin to move outwards, and eventually also intersect $\mathscr{I}^+$. For others, depending on b, they continue towards increasing l and eventually cross the horizon and do not reach $\mathscr{I}^+$. We will not be concerned with the later rays.

For a fixed apex (say at $l = l_0 < 1/3\sqrt{2}m$) the null rays, on each sheet, are characterized by the value of the impact parameter b. For the first set, ($\dot{l} < 0$), the range of b is from $b = 0$ to a maximum, b_m, where the b_m is determined by $A = b_m^2(2\sqrt{2}ml_0^3 - l_0^2) + 1 = 0$, i.e.

$$b_m^2 = 1/(l_0^2 - 2\sqrt{2}ml_0^3). \tag{3.26}$$

For the second sheet ($\dot{l} > 0$), the range is again from some b_m to $b = 0$, but now there is a critical value b_c, such that for all $b < b_c$ the rays continue past the horizon. To determine b_c, we want the smallest b so that A has a real positive root, l_c. By plotting A against l it is easily calculated that l_c is a double root and $l_c = 1/3\sqrt{2}m$, with $b_c = 3\sqrt{6}m$. Thus, on the second sheet, our range for b is $b_c < b < b_m$. Note that a ray beginning at $l = l_0$ with $b = b_c$ approaches asymptotically the well-known (unstable) orbit $l = l_c$.

We now restrict ourselves (for the time being) to the family of null geodesics from first sheet. Then (3.25) can be written as

$$\dot{u} = \frac{1 - \sqrt{A}}{2(l^2 - 2\sqrt{2}ml^3)} = \frac{b^2(1 - \sqrt{A})}{2(1 - A)} = \frac{b^2}{2(1 + \sqrt{A})}, \tag{3.27a}$$

$$\dot{l} = -\sqrt{A}, \tag{3.27b}$$

$$ds = -\frac{dl}{\sqrt{A}}, \tag{3.27c}$$

i.e.,

$$du = -\frac{b^2 dl}{2\sqrt{A}(1 + \sqrt{A})}. \tag{3.27d}$$

Integration of (3.27d) then yields

$$u = u_0 - \frac{1}{2}\int_{l_0}^{l}\frac{b^2\,dl'}{\sqrt{A}} + \frac{1}{2}\int_{l_0}^{l}\frac{b^2\,dl'}{1 + \sqrt{A}} = u(u_0, l_0, b, l). \tag{3.28}$$

It should be noted that since A is a cubic, both the integrals appearing on rhs of (3.28) are elliptic integrals, and they can be reexpressed as combinations of the standard elliptic integrals of first, second, or third types. This ellipticity is the consequence of $m \neq 0$ in A. When $m = 0$, these become elementary integrals and (3.28) reduces to the flat-space situation (3.8).

Returning to the remaining geodesic equation (3.23) with the aid of (3.27c) and choosing the apex on $\phi = 0$ axis, we obtain

$$\varphi = \int_{l_0}^{l} \frac{-b\,dl}{\sqrt{A}} \equiv \Phi(l_0,l,b). \tag{3.29}$$

We can now, in (3.28) and (3.29), pass to the limit $l = 0$ ($r = \infty$) obtaining

$$u = u_0 - \frac{1}{2}\int_{l_0}^{0} \frac{b^2\,dl'}{\sqrt{A}} + \frac{1}{2}\int_{l_0}^{0} \frac{b^2\,dl'}{1+\sqrt{A}} \equiv U(u_0,l_0,b), \tag{3.30a}$$

$$\hat{\varphi} = -\int_{l_0}^{0} \frac{b\,dl'}{\sqrt{A}} \equiv \Phi(l_0,b). \tag{3.30b}$$

Equations (3.30) define implicitly (by eliminating b between them) a function

$$u = u(u_0,l_0,\hat{\varphi}), \tag{3.31}$$

which yields a portion of the light cone cut of $\mathscr{I}^+$ (the equatorial plane). However, because of the spherical symmetry of the problem, the angle ϕ can be considered as the angle between a vector pointing (from the origin) towards an arbitrary apex with angular position $(\zeta_0,\bar{\zeta}_0)$ and the final angular position on $\mathscr{I}^+$, $(\zeta,\bar{\zeta})$. As in the Minkowski space case we have (3.14b)

$$\cos\hat{\varphi} = 1 - 2l_a(\zeta,\bar{\zeta})\mathring{l}^a(\zeta_0,\bar{\zeta}_0). \tag{3.32}$$

Thus with (3.32), Eq. (3.31) has the form

$$u = Z(u_0,l_0,\zeta_0,\bar{\zeta}_0,\zeta,\bar{\zeta}), \tag{3.33a}$$

and becomes the light cone cut function of $\mathscr{I}^+$ from $(u_0,l_0,\zeta_0,\bar{\zeta}_0)$.

Unfortunately, we cannot solve either (3.30a) or (3.30b) explicitly for b, and we must give (3.33) parametrically as

$$u = u_0 + \frac{1}{2}\int_0^{l_0} \frac{b^2\,dl}{\sqrt{A}} - \frac{1}{2}\int_0^{l_0} \frac{b^2\,dl}{1+\sqrt{A}}, \tag{3.33b}$$

$$\hat{\varphi} \equiv \arccos(1 - 2l_a\mathring{l}^a) = \int_0^{l_0} \frac{b\,dl}{\sqrt{A}}. \tag{3.33c}$$

We now consider briefly the second sheet ($\dot{l} > 0$ initially). The integration procedure is slightly more complicated than in the $\dot{l} < 0$ case. We must first integrate (3.25) and (3.23) with the $+\sqrt{A}$ (for fixed b) to the "bounce" point l_b [$A(l_b) = 0$] after which we return to the first sheet ($-\sqrt{A}$) and integrate to $l = 0$. Performing these operations, we obtain the cut function for the second sheet:

$$u = Z(u_0,l_0,\zeta_0,\bar{\zeta}_0,\zeta,\bar{\zeta}), \tag{3.34a}$$

or, parametrically,

$$u = u_0 + \frac{1}{2}\int_{l_0}^{l_b} \frac{2b^2\,dl}{\sqrt{A}(1-A)} + \frac{1}{2}\int_0^{l_0} \frac{b^2\,dl}{\sqrt{A}(1+\sqrt{A})}, \tag{3.34b}$$

$$\hat{\varphi} = \arccos(1 - 2l_a\mathring{l}^a) = 2\int_{l_0}^{l_b} \frac{b\,dl}{\sqrt{A}} + \int_0^{l_0} \frac{b\,dl}{\sqrt{A}}. \tag{3.34c}$$

Though the equations of the cut function Z, i.e., (3.33) or (3.34), appear to be very complicated, it will turn out (next section) that they can be manipulated with relative ease and that from them the Schwarzschild metric (up to conformal factor) can be reconstructed.

IV. RECONSTRUCTION OF METRIC

In the previous section we obtained the light cone cut function for Schwarzschild space-time, while in Sec. II we gave the general description for the construction of the conformal metric from the light cone cut function. In principle, we could stop at this point—the general theory is there as well as a specific cut function, so the metric could be calculated. However, for two reasons we would like to show how the calculation is done. First of all it will serve as a check on both the general formalism and also on the correctness of our derivation of the Schwarzschild Z. Second, it forces us to compute certain explicit expressions that have intrinsic interest on their own. These will be discussed in Sec. V.

Working with the first sheet, we begin with $Z = Z(u_0,l_0,\zeta_0,\bar{\zeta}_0,\zeta,\bar{\zeta})$ (expressed parametrically) in Eq. (3.33), and wish to obtain explicit expressions for the following quantities which are needed for computation of the metric;

$$Z_{,a} = (Z_{,u_0}, Z_{,l_0}, Z_{,\zeta_0}, Z_{,\bar{\zeta}_0}), \tag{4.1a}$$

$$\eth Z_{,a}, \quad \bar{\eth} Z_{,a}, \quad \eth\bar{\eth} Z_{,a}, \tag{4.1b}$$

$$\eth^2 Z_{,a}, \quad \bar{\eth}^2 Z_{,a}. \tag{4.1c}$$

In addition (though these quantities are not needed for the calculation of the metric) we wish to obtain, for later use,

$$\eth^2 Z, \quad \bar{\eth}^2 Z. \tag{4.1d}$$

Before beginning these calculations we first give some notation and relationships from flat space tetrad calculus which will be needed later. In addition to the vector $l_a(\zeta,\bar{\zeta})$, Eq. (3.12), we have three related vectors[6] forming a tetrad (for each value of ζ, $\bar{\zeta}$) namely

$$\begin{aligned} l_a &= l_a(\zeta,\bar{\zeta}), \\ m_a &= \eth l_a, \\ \bar{m}_a &= \bar{\eth} l_a, \\ n_a &= l_a + \eth\bar{\eth} l_a, \end{aligned} \tag{4.2}$$

where l and n have spin weights zero and m and $\bar{m}$ spin weights 1 and -1, respectively. They have nonvanishing scalar products among themselves,

$$l^a n_a = -m^a \bar{m}_a = 1.$$

The vectors satisfy the identities

$$\begin{aligned} &\eth m^a = \bar{\eth}\bar{m}^a = 0, \\ &\bar{\eth} m^a = \eth \bar{m}^a = n^a - l^a, \\ &\eth n^a = -m^a, \\ &\bar{\eth} n^a = -\bar{m}^a. \end{aligned} \tag{4.3}$$

Returning now to (4.1), we see immediately from (3.33) that

$$Z_{,u_0} = 1. \tag{4.4}$$

The calculation of $Z_{,l_0}$ is just a bit more tricky as u depends explicitly on l_0 (upper limit of integration) and implicitly on l_0 via b.

We thus have from (3.33b) that

$$Z_{,l_0} = \partial_{l_0} Z + Z_{,b}\, \partial_{l_0} b,$$

$$Z_{,l_0} = \frac{1}{2} \frac{b^2}{\sqrt{A} + A}\bigg|_{l=l_0} + \partial_{l_0} b \frac{1}{2} \int_0^{l_0} \frac{b}{A\sqrt{A}}\, dl. \tag{4.5}$$

By differentiating (3.33c) with respect to $l_0[(\zeta,\zeta_0) = \text{const}]$, we obtain

$$\frac{b}{\sqrt{A}}\bigg|_{l=l_0} + \partial_{l_0} b \int_0^{l_0} \frac{dl}{A\sqrt{A}} = 0. \tag{4.6}$$

Substituting (4.6) into (4.5) and simplifying, we obtain

$$Z_{,l_0} = \frac{-b^2}{2(1+\sqrt{A})}\bigg|_{l_0}. \tag{4.7}$$

In a similar fashion we calculate $Z_{,\zeta_0}$ and $Z_{,\bar{\zeta}_0}$ with the result that $Z_{,a}$ is

$$\begin{aligned} &Z_{,u_0} = 1, \\ &Z_{,l_0} = -\left[b^2/2(1+\sqrt{A})\right]_{l_0}, \\ &Z_{,\zeta_0} = bG/2(1+\zeta_0\bar{\zeta}_0), \\ &Z_{,\bar{\zeta}_0} = b\bar{G}/2(1+\zeta_0\bar{\zeta}_0), \end{aligned} \tag{4.8}$$

where the G and $\bar{G}$ arose in the differentiation of the arccos in (3.33c) and have the form

$$G = \left(\frac{l_a \mathring{m}^a}{l_b \bar{\mathring{m}}^b}\right)^{1/2} = -\left(\frac{(\bar{\zeta}-\bar{\zeta}_0)(1+\zeta\bar{\zeta}_0)}{(\zeta-\zeta_0)(1+\zeta_0\bar{\zeta})}\right)^{1/2}, \tag{4.9}$$

and $G\bar{G} = 1$.

At this point we have an ideal check on the accuracy of our expression for Z. If we calculate the norm of $Z_{,a}$ in (4.8) using the Schwarzschild metric (for either sheet), we obtain the identity

$$g^{ab} Z_{,a} Z_{,b} \equiv 0, \tag{4.10}$$

for all $(\zeta,\bar{\zeta})$. This shows that, in the Schwarzschild spacetime, $Z = \text{const.}$ is, for arbitrary $(\zeta,\bar{\zeta})$, in fact a null surface and that, at a fixed point $(u_0,l_0,\zeta_0,\bar{\zeta}_0)$, $Z_{,a}$ describes the tangent space null cone as $(\zeta,\bar{\zeta})$ varies. Since knowledge of the local cones determines the conformal metric,[7] we could stop at (4.10). We nevertheless continue to the next level.

By applying $\eth,\bar{\eth}$, and $\eth\bar{\eth}$, respectively to (4.8), we obtain the remaining tetrad vectors:

$\eth Z_{,a}$:

$$\begin{aligned} &\eth Z_{,u_0} = 0, \\ &\eth z_{,l_0} = -b\eth b/2\sqrt{A}, \\ &\eth z_{,\zeta_0} = \eth(bG)/2(1+\zeta_0\bar{\zeta}_0), \\ &\eth Z_{,\bar{\zeta}_0} = \eth(b\bar{G})/2(1+\zeta_0\bar{\zeta}_0), \end{aligned} \tag{4.11}$$

$\bar{\eth} Z_{,a}$:

$$\begin{aligned} &\bar{\eth} Z_{,u_0} = 0, \\ &\bar{\eth} Z_{,l_0} = -b\bar{\eth} b/2\sqrt{A}, \\ &\bar{\eth} Z_{,\zeta_0} = \bar{\eth}(bG)/2(1+\zeta_0\bar{\zeta}_0), \\ &\bar{\eth} Z_{,\bar{\zeta}_0} = \bar{\eth}(b\bar{G})/2(1+\zeta_0\bar{\zeta}_0), \end{aligned} \tag{4.12}$$

$\eth\bar{\eth} Z_{,a}$:

$$\begin{aligned} &\bar{\eth}\eth Z_{,u_0} = 0, \\ &\bar{\eth}\eth Z_{,l_0} = -(\eth b\, \bar{\eth} b + bA\eth\bar{\eth} b)/2A\sqrt{A}, \\ &\bar{\eth}\eth Z_{,\zeta_0} = \bar{\eth}\eth(bG)/2(1+\zeta_0\bar{\zeta}_0), \\ &\bar{\eth}\eth Z_{,\bar{\zeta}_0} = \bar{\eth}\eth(b\bar{G})/2(1+\zeta_0\bar{\zeta}_0). \end{aligned} \tag{4.13}$$

The final quantity needed is $\eth^2 Z_{,a}$:

$$\begin{aligned} &\eth^2 Z_{,u_0} = 0, \\ &\eth^2 Z_{,l_0} = -((\eth b)^2 + bA\eth^2 b)/2A\sqrt{A}, \\ &\eth^2 Z_{,\zeta_0} = \eth^2(bG)/2(1+\zeta_0\bar{\zeta}_0), \\ &\eth^2 Z_{,\bar{\zeta}_0} = \eth^2(b\bar{G})/2(1+\zeta_0\bar{\zeta}_0), \end{aligned} \tag{4.14}$$

and its complex conjugate $\bar{\eth}^2 Z_{,a}$.

As we discussed in Sec. II, the basic quantities needed to obtain the metric from the Z, are the four Λ_0, Λ_+, Λ_-, and Λ_1 defined by

$$\eth^2 Z_{,a} = \Lambda_0 Z_{,a} + \Lambda_+\, \eth Z_{,a} + \Lambda_-\, \bar{\eth} Z_{,a} + \Lambda_1\, \eth\bar{\eth} Z_{,a}.$$

Now by using (4.8), (4.11), (4.12), (4.13), (4.14), and several identities involving the tetrad vectors l, m, $\bar{m}$, n generated by (4.9) and by applying the $\eth$ and $\bar{\eth}$ operator to $G\bar{G} = 1$, we calculate the Λ_i's (a long, tedious calculation) and obtain

$$\Lambda_0 = 0, \quad \Lambda_1 = B_1/B_2, \tag{4.15a}$$

$$\begin{aligned} \Lambda_- = {} & (\eth^2 b\, \eth G - \eth b\, \eth^2 G)/B_3 \\ & + B_1(\eth b\, \bar{\eth}\eth G - \eth G\, \bar{\eth}\eth b)/B_2 B_3 \\ & + [b^2 \eth b\, (\eth G)^2 - b^2 A \eth b\, (\eth G)^2 + b\, (\eth b)^2 G \eth G \\ & - G^2 (\eth b)^3]/bB_2, \end{aligned} \tag{4.15b}$$

$$\begin{aligned} \Lambda_+ = {} & (\bar{\eth} b\, \eth^2 G - \eth^2 b\, \bar{\eth} G)/B_3 \\ & - B_1(\bar{\eth} b\, \bar{\eth}\eth G - \bar{\eth} G\, \bar{\eth}\eth b)/B_2 B_3 \\ & - (b\, \eth b\, \bar{\eth} b\, G\, \eth G - G^2\, \bar{\eth} b\, (\eth b)^2 + b^2\, \eth b\, \eth G\, \bar{\eth} G \\ & - b^2 A\, \bar{\eth} b\, (\eth G)^2/bB_2, \end{aligned} \tag{4.15c}$$

where

$$\begin{aligned} B_1 &= G^2 (\eth b)^2 - b^2 A\, (\eth G)^2, \\ B_2 &= G^2\, \eth b\, \bar{\eth} b - b^2 A\, \eth G \bar{\eth} G, \\ B_3 &= \bar{\eth} b\, \eth G - \eth b\, \bar{\eth} G, \end{aligned} \tag{4.15d}$$

From the Λ_i's we could, via Eqs. (2.11)–(2.14) and conjugates, reconstruct the tetrad components of the Schwarzschild metric, namely, $g^{ij} = g^{ab}\theta^i_{\ a}\theta^j_{\ b}$ and then the g^{ab}. It actually turns out to be simpler (since we are only verifying that our method yields the already known Schwarzschild metric) to begin with the known g^{ab} and calculate the g^{ij} with the use of the tetrad $\theta^i_{\ a}$. From the same set, Eqs. (2.11)–(2.14) we can calculate the λ_μ and compare with (4.15) and thereby verify the reconstruction of the g^{ab}.

We have in fact carried out this calculation which is relatively long but straightforward. The details shed no new light. However, for completeness we present the g^{ij}:

$$g^{00} = g^{0+} = g^{0-} = 0, \tag{4.16a}$$

$$g^{01} = \frac{\eth b\, \bar{\eth} b}{4A} + \frac{b^2 \eth G\, \bar{\eth}\bar{G}}{4}, \tag{4.16b}$$

$$g^{++} = -\frac{(\eth b)^2}{4A} - \frac{b^2 \eth G\, \eth \bar{G}}{4}, \tag{4.16c}$$

$$g^{+1} = \frac{A-1}{4bA^2}(\eth b)^2(\bar{\eth} b) - \frac{\eth b\, \bar{\eth}\eth b}{4A} - \frac{1}{8}\bar{\eth}(b^2\, \eth G\, \eth \bar{G}), \tag{4.16d}$$

$$g^{11} = \frac{A-1}{4b^2A^3}(\eth b\, \bar{\eth} b + bA\, \bar{\eth}\eth b)^2 - \frac{1}{4}\bar{\eth}\eth(bG)\bar{\eth}\eth(b\bar{G}), \tag{4.16e}$$

and rest of the components are given by

$$g^{-1} = \overline{g^{+1}}, \quad g^{+-} = -g^{01}, \quad g^{--} = \overline{g^{++}}. \tag{4.16f}$$

V. DISCUSSION

In this final section we would like to discuss a series of different items, all connected to the main theme of the paper, some intimately connected, others much more loosely.

(1) We would like to return to a question that was just touched upon in Sec. III, namely, the integration of the null geodesic equations for a particular value of the impact parameter $b = b_c = \sqrt{6}\cdot 3m$. We pointed out then that these rays can be thought of (i) on the first sheet as if they had come from the unstable orbit $l = 1/3\sqrt{2}m$ moving towards $\mathscr{I}^+$ while (ii) for the second sheet they would represent rays moving towards the same unstable orbit and eventually remaining at $l = 1/3\sqrt{2}m$. A point of interest for us is that the elliptic integrals that appear in Sec. II for this case reduce to elementary integrals [the polynomial $A(l)$ factors in this case so that $A = 54m^2(l - 1/3\sqrt{2}m)^2(2\sqrt{2}ml + 1/3)$]. Because of this (on sheet I) one can, for specific directions, obtain explicit expressions for Z and Φ. We have not bothered to evaluate them.

A second point (concerning sheet II) also related to $b = b_c$ is that one can exactly integrate (3.25d) from l_0 to $l = 1/3\sqrt{2}m - \epsilon$, $\epsilon > 0$, with the result

$$\varphi = \log\left\{\frac{2(1 - 3\sqrt{2}ml_0)}{\sqrt{2}\epsilon m\left[3\sqrt{2}ml_0 + 3(2\sqrt{2}ml_0 + 1/3)^{1/2} + 2\right]}\right\}. \tag{5.1}$$

Thus for fixed l_0 the geodesic will have an arbitrarily large deflection (or number of rotations) as $\epsilon \to 0$. In a similar manner, if $b = 3\sqrt{6}m + \delta$ (δ small but positive), a ray from $l = l_0 < 1/3\sqrt{2}m$ will come close to $l = 1/3\sqrt{2}m$ but "bounce" and move towards $\mathscr{I}^+$. Again in this case one can easily approximate the integrals and study analytically the multiple encirclements of the black hole.

(2) Quantities that have intrigued us but for which we have not yet found a simple geometric interpretation are the (multiple) periods of the elliptic integrals involved in the evaluation of both u and ϕ, in (3.38) or (3.34). Presumably a related question is the meaning of the analytic continuation of the integrals into the complex b-plane. They seem to have a relationship to the analytic continuation of Schwarzschild geometry to its Euclidean version and to Hawking radiation effects.[8]

Related also is the inversion of the integrals so that one has elliptic functions instead of elliptic integrals. This can be clearly done for the ϕ integral, but we have not succeeded in doing so for the u integral.

(3) If we now return to Eqs. (3.33) or (3.34) and apply the $\eth$ and $\bar{\eth}$ operators several times (now without taking the gradient), we obtain (for the first sheet)

$$\eth^2 Z = b\, \eth\hat{G} + \hat{G}\, \eth b, \tag{5.2}$$

$$\bar{\eth}\eth Z = b\, \bar{\eth}\hat{G} + \hat{G}\, \bar{\eth} b, \tag{5.3a}$$

$$\bar{\eth} Z = b\bar{\hat{G}}, \tag{5.3b}$$

$$\eth Z = b\hat{G}, \tag{5.3c}$$

$$Z = Z(u_0, l_0, \zeta_0, \bar{\zeta}_0, \zeta, \bar{\zeta}), \tag{3.33a'}$$

where $\hat{G}$ is given as

$$\hat{G} = \left[\frac{(\bar{\zeta} - \bar{\zeta}_0)(1 + \bar{\zeta}\zeta_0)}{(\zeta - \zeta_0)(1 + \bar{\zeta}_0\zeta)}\right]^{1/2}. \tag{5.3d}$$

[The relationship between G defined by (4.9) and $\hat{G}$ above should be noted. Whereas G arises from taking $\eth_0$ of (3.32), $\hat{G}$ arises by taking $\eth$ of the same expression.]

Since the right hand side of (5.2) is a function of $(u_0, l_0, \zeta_0, \bar{\zeta}_0)$, as are the right sides of (5.3), we could view (5.2) as an equation of the form

$$\eth^2 Z = -\sigma_z(Z, \eth Z, \bar{\eth} Z, \bar{\eth}\eth Z, \zeta, \bar{\zeta}), \tag{5.4}$$

but given parametrically by (5.2) and (5.3). Equations (5.4) and their conjugates can be considered as the "Einstein equations" for the Schwarzschild geometry in the following sense: They appear to have a unique four-parameter set of solutions given by (3.33), where the solution space itself, i.e., $(u_0, l_0, \zeta_0, \bar{\zeta}_0)$ defines the Schwarzschild manifold and the solutions themselves yield the Schwarzschild metric.

The quantity σ_z or (from a slightly simpler point of view) the right side of (5.2) has a simple geometric meaning, namely, it is the asymptotic shear of the null cone with apex $(u_0, l_0, \zeta_0, \bar{\zeta}_0)$. This is seen from the Sachs theorem[9] on the transformations of asymptotic shear and the fact that the Bondi shear for Schwarzschild space-time is zero.

(4) A further interesting piece of technical information is obtainable from (5.3c) and (5.2). We have wondered what would be the consequences of choosing a null goedesic beginning at $(u_0, l_0, \zeta_0, \bar{\zeta}_0)$ which possesses a conjugate point at $\mathscr{I}^+$, on the local geometry of Z in the neighborhood of the conjugate point. This can be investigated by choosing the origin of the ray in the equitorial plane $\theta = \pi/2$ on the line $\phi = 0$, then it is easy to see from the symmetry of the situation that there will exist a geodesic such that the point on $\mathscr{I}^+$ along the negative x axis, i.e., $\theta = \pi/2$ and $\phi = \pi$ will be a conjugate point. Notice that from (5.3c) we have

$$\eth Z = b\hat{G} = b\sqrt{\bar{\zeta}/\zeta}, \tag{5.5}$$

where we have used $(\theta_0 = \pi/2, \phi_0 = 0) \to (\zeta_0, \bar{\zeta}_0) = 0$. If we now have $\zeta + \bar{\zeta} \to \infty$ (for $\phi \to \pi$, the antipodal point to the

positive x axis), we have the result that $\eth Z = be^{i2\phi}$, where $\zeta = |\zeta|e^{-i\phi}$, i.e., the value of Z depends not on the final $(\zeta,\bar{\zeta})$ but on the direction of approach to the final point. The behavior of the surface defined by $u = Z(x^a,\zeta,\bar{\zeta})$ near the conjugate point is thus like a cusp. From this we see that $\eth^2 Z$ at the conjugate point is singular and hence the asymptotic shear σ_z at a conjugate point is singular. This result gives us a general warning that our construction of metrics from the cut function breaks down at conjugate points.

(5) It is shown in the general theory of light cone cuts[1] that a necessary condition for Lorentzian signature of the metric is

$$P^{-1} = 1 - \Lambda_1\bar{\Lambda}_1 > 0. \tag{5.6}$$

If we calculate (5.6) for Schwarzschild geometry, we obtain

$$P^{-1} = b^2 A\left(\frac{\bar{\eth} b\,\eth G}{G} + \frac{\eth b\,\bar{\eth}\bar{G}}{\bar{G}}\right)^2, \tag{5.7}$$

which is greater than zero for null rays that *can* escape to infinity.

(6) A question that remains to be considered is what is the twistor description of the theory of light-cone cuts in general and in particular for the Schwarzschild geometry. It is clear that for each x^a or cut there is a two-parameter family of twistor lines on $C\mathscr{I}^+$ whose envelope is the analytic continuation of the cut function into the complex $\mathscr{I}^+$. It should be investigated whether this view allows a deeper and clearer understanding of the cut function as it did in the special case of $\mathscr{H}$-space theory.

[1] C. N. Kozameh and E. T. Newman, J. Math. Phys. **24**, 2481 (1983).
[2] R. Geroch, *Asymptotic Structure of Space-Time*, edited by P. Esposito and L. Witten (Plenum, New York, 1976), p. 1.
[3] H. Bondi, Proc. Roy. Soc. (London), A **269**, 21 (1963).
[4] E. T. Newman and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[5] A. R. Exton, E. T. Newman, and R. Penrose, J. Math. Phys. **10**, 1566 (1969).
[6] S. L. Kent and E. T. Newman, J. Math. Phys. **24**, 949 (1983).
[7] S. Hawking and G. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1973), p. 61.
[8] J. B. Hartle and S. W. Hawking, Phys. Rev. D **13**, 2188 (1976).
[9] R. K. Sachs, Proc. Roy. Soc. (London) **270**, 103 (1962).

All null orbit type D electrovac solutions with cosmological constant

Alberto García D.[a] and Humberto Salazar I.
Departamento de Física, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Apartado Postal 14-740, 07000 México D.F., Mexico

(Received 1 July 1982; accepted for publication 13 May 1983)

All null orbit type D solutions of the Einstein–Maxwell equations with λ are obtained. There are only two families of solutions depending upon whether the complex expansion of the electromagnetic eigenvectors aligned along the double DP vectors is different or equal to zero; they are the null orbit solution with complex expansion, and the five-parameter free of complex expansion null orbit solution, respectively.

PACS numbers: 04.20.Jb

I. INTRODUCTION AND EQUATIONS OF THE PROBLEM

In 1981 Debever and McLenaghan[1] obtained, without explicit integration, the general metric structures of spacetimes V_4 which are solutions of the Einstein–Maxwell equations with λ,

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R + \lambda g_{\mu\nu} = f_{\mu\alpha}f^{\alpha}_{\nu} - \tfrac{1}{4}g_{\mu\nu}f_{\alpha\beta}f^{\alpha\beta},$$
$$f_{\mu\nu;}{}^{\nu} = 0 = f_{[\mu\nu;\alpha]},$$

and which fulfill the following conditions.

(i) The Weyl tensor $C_{\alpha\beta\mu\nu}$ is of Petrov type D, i.e., there exist two real null principal directions, Debever–Penrose (DP) vectors, l and n such that, at each point of V_4,

$$l^{\mu}l^{\nu}C_{\alpha\mu\nu[\beta}l_{\gamma]} = 0 = n^{\mu}n^{\nu}C_{\alpha\mu\nu[\beta}n_{\gamma]}.$$

(ii) The electromagnetic field tensor $f_{\alpha\beta}$ is nonsingular and its principal null directions are aligned along the DP directions of the Weyl tensor

$$l^{\mu}f_{\mu[\alpha}l_{\beta]} = 0 = n^{\mu}f_{\mu[\alpha}n_{\beta]}.$$

(iii) The invariants of the Weyl tensor and the tracefree Ricci tensor $C_{\mu\nu} := R_{\mu\nu} - \tfrac{1}{4}g_{\mu\nu}R$ satisfy one of the inequalities

$$C_{\alpha\beta\mu\nu}C^{\alpha\beta\mu\nu} \neq \tfrac{4}{3}C_{\mu\nu}C^{\mu\nu},$$

or

$$C^{*}_{\mu\nu\alpha\beta}C^{\mu\nu\alpha\beta} \neq 0.$$

They established, among others results, a theorem which applies to the $\mathscr{D}$ class of solutions, i.e., to solutions of the Einstein–Maxwell equations above satisfying conditions (i)–(iii).

Theorem 1: Every solution in $\mathscr{D}$ admits at least a two-parameter orthogonally transitive abelian isometry group. If the orbits of the group are non-null, the group is invertible and there exists a coordinate system (u, v, w, x) such that the metric and the self-dual Maxwell field have the form

$$ds^2 = -e(L\,du + M\,dv)^2 + eR^2dw^2 + e(N\,du + P\,dv)^2 - eT^2dx^2,$$
$$F = B[R(L\,du + M\,dv)\wedge dw - eT(N\,du + P\,dv)\wedge dx], \tag{T1}$$

where L, M, and R are real-valued functions, B is a complex-valued function, and N, P, and T are functions satisfying $\overline{N} = -eN, \overline{P} = -eP, \overline{T} = eT$, where all these functions are independent of the coordinates u and v, and where $e = 1$ (spacelike orbits) or $e = -1$ (timelike orbits). If the orbits are null, the group is not invertible and there exists a system of coordinates (u, v, w, x) in which the metric and the self-dual Maxwell field have the form

$$ds^2 = 2R\,dw(L\,du + M\,dv) - (N\,du + P\,dv)^2 - T^2dx^2,$$
$$F = B[-R(L\,du + M\,dv)\wedge dw + iT(N\,du + P\,dv)\wedge dx], \tag{T2}$$

where L, M, N, P, R, and T are real-valued functions and B is a complex-valued function, all independent of the coordinates u and v.

The main purpose of the present work is to give the complete set of type D electrovac solutions with λ possessing an isometry group with null orbits. According to the theorem above, in this case, the metric modified to signature $+2$ and the 2-form of the electromagnetic field can be given as

$$g = (N\,du + P\,dv)^2 + T^2dx^2 + 2R\,dy(L\,du + M\,dv),$$
$$\omega = -(\mathscr{E} + i\check{\mathscr{B}})\{R(L\,du + M\,dv)\wedge dy + iT(N\,du + P\,dv)\wedge dx\}, \tag{1.1}$$

where the real structural functions N, P, T, R, L, M, E, and $\check{\mathscr{B}}$, constrained to the Einstein–Maxwell field equations, are all independent of the Killingian variables u and v. The functions $\mathscr{E}$ and $\check{\mathscr{B}}$ are the electromagnetic invariants defined by

$$\mathscr{F} = \tfrac{1}{4}f_{\mu\nu}f^{\mu\nu} + \tfrac{1}{4}\check{f}_{\mu\nu}f^{\mu\nu} =: -\tfrac{1}{2}(\mathscr{E} + i\check{\mathscr{B}})^2.$$

Working in the null tetrad formalism the metric g and ω from (1.1) can be given as

$$g = 2e^1\otimes e^2 + 2e^3\otimes e^4,$$
$$\omega = (\mathscr{E} + i\check{\mathscr{B}})(e^1\wedge e^2 + e^3\wedge e^4), \tag{1.2}$$

where

$$e^1 = (1/\sqrt{2})(N\,du + P\,dv + iT\,dx), \quad e^2 = (\overline{e^1}),$$
$$e^3 = R\,dy, \quad e^4 = L\,du + M\,dv. \tag{1.3}$$

The congruences e^3 and e^4, aligned along the double DP directions, via the Sachs–Goldberg theorem (fulfillment of (i)–(iii), are geodesic and shearfree, i.e.,

[a] Also at Sección de Graduados, Escuela Superior de Ingeniería Mecánica y Eléctrica del Instituto Politécnico Nacional, México D.F., Mexico.

 0022-2488/83/102498-06$02.50

$$\Gamma_{424} = \Gamma_{422} = 0 = \Gamma_{313} = \Gamma_{311}, \tag{1.4}$$

where $\Gamma_{abc} = \Gamma_{[ab]c}$ are the components of the connection 1-forms $\Gamma_{ab} = \Gamma_{abc}e^c$, which are defined by the first Cartan equations

$$de^a = e^b \wedge \Gamma^a{}_b. \tag{1.5}$$

Under the present alignment of e^3 and e^4 along the DP vectors, the only nonvanishing independent component of the traceless Ricci tensor is

$$C_{12} = -(\mathscr{E}^2 + \check{\mathscr{B}}^2), \tag{1.6}$$

the conformal curvature C_{abcd} is characterized by the only nonvanishing curvature quantity $C^{(3)}$, and the curvature constant $R = -4\lambda$, λ being the cosmological constant.

The second structure equations, into which the Einstein equations with an aligned electromagnetic field are built in, can be written as

$$\begin{aligned}
\mathscr{A} &= d\Gamma_{42} + \Gamma_{42} \wedge (\Gamma_{12} + \Gamma_{34}) = (\tfrac{1}{2}C^{(3)} + \tfrac{1}{3}\lambda)e^3 \wedge e^1, \\
\mathscr{B} &= d\Gamma_{31} + (\Gamma_{12} + \Gamma_{34}) \wedge \Gamma_{31} = (\tfrac{1}{2}C^{(3)} + \tfrac{1}{3}\lambda)e^4 \wedge e^1, \\
\mathscr{C} &= d(\Gamma_{12} + \Gamma_{34}) + 2\Gamma_{42} \wedge \Gamma_{31} \\
&= [C^{(3)} - \tfrac{1}{3}\lambda - (\mathscr{E}^2 + \check{\mathscr{B}}^2)]e^1 \wedge e^2 \\
&\quad + [C^{(3)} - \tfrac{1}{3}\lambda + (\mathscr{E}^2 + \check{\mathscr{B}}^2)]e^3 \wedge e^4.
\end{aligned} \tag{1.7}$$

The Maxwell equations of the problem studied are

$$\begin{aligned}
&d \ln(\mathscr{E} + i\check{\mathscr{B}})^{1/2} + \Gamma_{314}e^1 + \Gamma_{423}e^2 \\
&\quad - \Gamma_{312}e^3 - \Gamma_{421}e^4 = 0.
\end{aligned} \tag{1.8}$$

They simply state that the 2-form ω of the electromagnetic field, formula (1.1) or (1.2), is a closed form.

From Eqs. (1.5), by substituting there e^a from (1.3), one readily reads the connection components Γ_{abc}. The components Γ_{424} and Γ_{421} are identically equal to zero. The components Γ_{422}, Γ_{313}, and Γ_{311} amount to

$$\begin{aligned}
\Gamma_{422} &= (i/2TZ)(P_xN - PN_x), \\
\Gamma_{313} &= (1/\sqrt{2}RZ)(M_yL - L_yM), \\
\Gamma_{311} &= (1/2R)[T_y/T - (1/Z)(N_yM - P_yL)],
\end{aligned} \tag{1.9}$$

where $Z := NM - PL$. The vanishing conditions of these quantities, Eqs. (1.4), yield

$$P = p(y)N, \quad M = m(x)L, \quad T = f(x)N(m(x) - p(y)). \tag{1.10}$$

Thus without loss of generality, the metric (1.1) satisfying the conditions (1.4) can be written as

$$\begin{aligned}
g = {} & N^2(du + p(y)dv)^2 + N^2(m(x) - p(y))^2 f^2(x)dx^2 \\
& + 2R\, dy(du + m(x)dv),
\end{aligned} \tag{1.11}$$

where N and R are functions of x and y only, and m, f, and p are functions of their arguments.

From the integrability condition of Maxwell equations (1.8) two conditions arise.

(i) From the real part one infers

$$\partial_x\partial_y \ln \frac{R}{N^2(m-p)} = 0;$$

consequently,

$$R = \psi(x)\phi(y)N^2(m-p). \tag{1.12}$$

By redefining the coordinates x and y, one can always set $\phi(y) = 1$, and $\psi(x) = f(x)$.

(ii) The imaginary part leads to

$$\partial_x \frac{m_x}{m-p} - \partial_y \frac{p_y}{m-p} = 0, \tag{1.13}$$

the integration of which is a straightforward process.

By introducing the definitions

$$H^2(x,y) := R^{-1}, \quad K(x) := f^{-1}, \tag{1.14}$$

one brings the metric (1.11) into the form

$$\begin{aligned}
g = H^{-2}\Big(& \frac{K}{m-p}(du + p\,dv)^2 \\
& + \frac{m-p}{K}dx^2 + 2\,dy(du + m\,dv)\Big).
\end{aligned} \tag{1.15}$$

The information contained in the second Cartan equation (1.6) can be summarized as follows: The vanishing of the coefficients multiplying the 2-form $e^3 \wedge e^2$ implies

$$\text{from } \mathscr{A}: \quad H_{xx} - \frac{1}{4}\frac{p_{yy} + m_{xx}}{m-p}H = 0, \tag{1.16}$$

$$\text{from } \mathscr{B}: \quad H_{yy} + \frac{1}{4}\frac{p_{yy} + m_{xx}}{m-p}H = 0, \tag{1.17}$$

$$\text{from } \mathscr{C}: \quad H_{xy} - \frac{1}{2}\frac{m_xH_y - p_yH_x}{m-p} = 0, \tag{1.18}$$

$$p_yH_y + m_xH_x - \tfrac{1}{2}(m_{xx} + p_{yy})H = 0. \tag{1.19}$$

Two out of the three equations with second derivatives of H occur to be independent when p_y and m_x are not simultaneously equal to zero. This fact can be established by differentiating (1.19) and the use of Eq. (1.13) together with its third derivatives. Therefore under the condition imposed on p_y and m_x, one can consider as independent equations for H the ones given by (1.16), (1.17), and (1.19), remembering always the necessity of (1.13). If $p_y = 0 = m_x$, one arrives at two different branches of solutions; the de Sitter space [see below formulas (1.26), (1.30)–(1.32)] outside of the D's solutions, and the null orbit exceptional type D solution [see further (3.12), which does not satisfy the (iii) condition].

The equation for the scalar curvature amounts to

$$\begin{aligned}
-4\lambda(m-p) = {} & H^2K_{xx} - 3K_x\partial_xH^2 + 2K[-\partial_x\partial_xH^2 \\
& + 2H^{-2}(\partial_xH^2)^2] := D_xK,
\end{aligned} \tag{1.20}$$

which permits us to determine the function K once the m, p, and H are known.

The curvature quantities $C^{(3)}$ and C_{12} are correspondingly

$$\begin{aligned}
C^{(3)} = {} & -\frac{2}{3}\lambda - \frac{1}{2}\frac{KH^2}{m-p}\{\partial_x\partial_x \ln(m-p)H^{-2} \\
& + \tfrac{1}{2}\partial_x \ln[(m-p)H^{-2}]\partial_x \ln K^2H^{-2}(m-p)^{-1} \\
& + \tfrac{1}{2}\partial_y \ln(m-p)\cdot\partial_y[\ln(m-p) \\
& + 2i \ln K(m-p)^{-2}]\},
\end{aligned} \tag{1.21}$$

$$\begin{aligned}
C_{12} = {} & \frac{1}{2}\frac{KH^2}{m-p}\Big\{-\frac{p_{yy}}{m-p} + \frac{1}{2}\frac{K_{xx}}{K} \\
& - \frac{K_x}{K}\Big(\frac{m_x}{m-p} + \frac{H_x}{H}\Big) + 2\frac{m_x}{m-p}\frac{H_x}{H}\Big\}.
\end{aligned} \tag{1.22}$$

In this way, the problem of finding all null orbit electrovac type D solutions satisfying (i)–(iv) reduces to integrating the system, denoted subsequently by $\mathscr{S}$, determined by Eq. (1.13) (with m_x and p_y not vanishing simultaneously), (1.16), (1.17), (1.19), and (1.20). The integrability of the Maxwell equations (1.8), which now can be written as

$$d\ln(\mathscr{E}+i\check{\mathscr{B}})H^{-2}(m-p)-i\times\left[\frac{p_y}{m-p}dx+\frac{m_x}{m-p}dy\right]=0, \tag{1.23}$$

is guaranteed by Eq. (1.13).

Before starting the integration of the quoted equations, let us demonstrate a theorem:

Within the class of null orbit D' solutions, if the congruence e^4 is twist-free (Im $\Gamma_{312}=0$), then it is also nondiverging (Re $\Gamma_{312}=0$), and inversely.

Proof: The complex rotation of e^4 is given by

$$\Gamma_{312}=H\left(H_y+\frac{1}{2}\frac{p_y}{m-p}H\right)+\frac{i}{2}\frac{H}{m-p}m_x. \tag{1.24}$$

Let the divergence be zero, i.e.,

$$H_y+\frac{1}{2}\frac{p_y}{m-p}H=0.$$

Differentiating the expression above with respect to y, and substituting H_{yy} from (1.17) into the obtained relation one arrives, by virtue of (1.13), at

$$H(m_x)^2=0, \tag{1.25}$$

therefore $m_x=0$, i.e., Im $\Gamma_{312}=0$.

Inversely, suppose that the twist of e^4 is zero, i.e.,

$$m_x=0\Rightarrow m=m_0=\text{const.} \tag{1.26}$$

Hence, from (1.13), one has

$$p_{yy}+(p_y)^2/(m_0-p)=0. \tag{1.27}$$

Substituting p_{yy} into (1.19), one obtains

$$p_y\left(H_y+\frac{1}{2}\frac{p_y}{m_0-p}H\right)=0. \tag{1.28}$$

Thus one has the two possibilities

$$\text{(i)}\quad H_y+\frac{1}{2}\frac{p_y}{m_0-p}H=0\Rightarrow \text{Re}\,\Gamma_{312}=0 \tag{1.29}$$

and

$$\text{(ii)}\quad p_y=0\Rightarrow p=p_0=\text{const.} \tag{1.30}$$

In this last branch, the equations for H are $H_{xx}=H_{xy}=H_{yy}=0$; therefore

$$H=a+bx+cy, \tag{1.31}$$

where a, b, and c are constants. Entering with m, p, and H into Eqs. (1.24) and (1.20), one arrives at

$$\text{Re}\,\Gamma_{312}=c(a+bx+cy),$$
$$-4\lambda(m_0-p_0)=(a+bx+cy)^2K_{xx}-6b(a+bx+cy)K_x+12b^2K. \tag{1.32}$$

Hence if one demands Re Γ_{312} to be different from zero, then c is a nonvanishing constant. This, in turn, implies that K is a constant structural function yielding, together with H, to the de Sitter space with all curvature quantities equal to zero except the scalar curvature $R=-4\lambda$, and, therefore, outside of the studied class of type D solutions.

Thus the class of null orbit electrovac type D solutions can be divided into two branches: the free of complex expansion D's and the D solutions with one DP vector possessing simultaneously twist and divergence.

Incidentally, branch (ii) with $\Gamma_{312}=0$ is not empty. It contains an exceptional null orbit type D solution; see Formula (3.12), and the static solution given by Eqs. (3.9) and (3.10) with l equal to zero.

II. INTEGRATION FOR ALL NULL ORBIT D SOLUTIONS WITH COMPLEX EXPANSION

In this section, all solutions of the system $\mathscr{S}$ within the class with nonvanishing Γ_{312} are determined. They are exhibited in a canonical system of coordinates together with their curvature quantities referred to the null tetrad (1.3).

The starting point in the integration of $\mathscr{S}$, provided that $m_x\neq 0$, is Eq. (1.13). A trivial solution, corresponding to a flat space, is obtained when p is a constant. Thus cases of interest arise only if p_y is also different from zero.

Equation (1.13), differentiated with respect to x and y, leads to a separable equation

$$m_{xxx}/m_x=-\alpha=-p_{yyy}/p_y, \tag{2.1}$$

where α is a separation constant. Integrating the above relations, one arrives at

$$m_{xx}+\alpha m=\beta,\quad p_{yy}-\alpha p=\gamma, \tag{2.2}$$

where β and γ are constants.

Two possible branches of solutions depending upon whether α is different or equal to zero ought to be considered.

Case $\alpha=0$: The general solutions fulfilling Eqs. (1.13) and (2.2) are

$$m=(\beta/2)x^2+\kappa_1x+\kappa_2,$$
$$p=-(\beta/2)y^2+\epsilon_1y+\epsilon_2, \tag{2.3}$$

with constants β, κ_i, and ϵ_i ($i=1,2$) constrained to

$$\kappa_1^2-2\beta\kappa_2+\epsilon_1^2+2\beta\epsilon_2=0.$$

Thus with $m_x\neq 0\neq p_y$, the constant β has to be different from zero.

Subjecting the coordinates in the metric (1.15) to the transformation

$$\{x,y,v,u\}\rightarrow\{x-\kappa_1/\beta,\,y+\epsilon_1/\beta,\,(2/\beta)v,$$
$$u-(1/\beta^2)(\epsilon_1^2+2\beta\epsilon_2)v\},$$

without loss of generality, one can set

$$m=x^2,\quad p=-y^2. \tag{2.4}$$

The equations for the function H, taking into account (2.4), amount to

$$H_{xx}=H_{yy}=yH_y-xH_x=0. \tag{2.5}$$

Hence H has the general form

$$H=\mu+\nu xy, \tag{2.6}$$

where μ and ν are constants.

Entering with the obtained H into the equation for K, (1.20), one arrives at

$$-4\lambda(x^2+y^2) = (\mu+\nu xy)^2 K_{xx} - 6\nu y(\mu+\nu xy)K_x + 12\nu^2 y^2 K. \quad (2.7)$$

By comparing coefficients to powers of y one obtains

$$\begin{aligned} -4\lambda &= \nu^2(x^2K_{xx} - 6xK_x + 12K), \\ 0 &= \mu\nu(xK_{xx} - 3K_x), \\ -4\lambda x^2 &= \mu^2 K_{xx}. \end{aligned} \quad (2.8)$$

The system above has two different solutions.

If the cosmological constant λ is present, then both parameters μ and ν are different from zero. The structural function K is then given by

$$K = -(\lambda/3)\mu^{-2}(x^4 + \mu^2\nu^{-2}). \quad (2.9)$$

The metric (1.20) with m and p from (2.4), H from (2.6), and K from (2.9), after a suitable scaling of coordinates, can be always written as

$$g = \frac{1}{(1-xy)^2}\left\{\frac{(-\lambda/3)(1+x^4)}{x^2+y^2}(du - y^2dv)^2 + \frac{x^2+y^2}{(-\lambda/3)(1+x^4)}dx^2 + 2\,dy(du + x^2dv)\right\}, \quad (2.10)$$

with λ strictly negative to have the physical Minkowskian signature $(+\ +\ +\ -)$, $\lambda < 0$.

Maxwell equations (1.23) yield

$$d\ln(\mathscr{E} + i\check{\mathscr{B}})H^{-2}(y+ix)^2 = 0 \Rightarrow \mathscr{E} + i\check{\mathscr{B}} = (e+ig)/(y+ix)^2(1-xy)^2,$$

where e and g are real constants of integration. According to (1.6), one has

$$C_{12} = -(e^2+g^2)(1-xy)^4/(x^2+y^2)^2,$$

which on the other hand, by virtue of (1.22), amounts to

$$C_{12} = -\frac{\lambda}{3}\frac{(1-xy)^4}{(x^2+y^2)^2};$$

therefore the real constants e, g, and λ ought to fulfill

$$(\lambda/3) = e^2 + g^2. \quad (2.11)$$

Consequently, the cosmological constant is to be positive, $\lambda > 0$! This contradiction implies that physically this kind of solution does not exist. This formal "solution" was given in Ref. 2; the misunderstanding arises from the wrong interpretation of the equality (7.17) of Ref. 2 which does not hold for real quantities.

If the cosmological constant is now equal to zero, one obtains the null orbit D solution with complex expansion. Equations (2.8) with $\lambda = 0$ permit two possible solutions:

$$\mu \neq 0 = \nu: \quad K = \kappa_0 + \kappa_1 x, \quad (2.12)$$

and

$$\mu = 0 \neq \nu: \quad K = \kappa_0 x^3 + \kappa_1 x^4, \quad (2.13)$$

where κ_0 and κ_1 are integration constants.

The metric (1.20), with K from (2.12), can be written as

$$g = \frac{x^2+y^2}{2nx - (e^2+g^2)}dx^2 + \frac{2nx-(e^2+g^2)}{x^2+y^2}\times(du - y^2dv)^2 + 2\,dy(du + x^2dv), \quad (2.14)$$

Maxwell equations (1.23) imply

$$\mathscr{E} + i\check{\mathscr{B}} = (e+ig)(y+ix)^{-2}, \quad (2.15)$$

where e and g are real constants. Hence the 2-form ω, which accompanies the metric (2.14), amounts to

$$\omega = -(e+ig)d\left[\frac{1}{y+ix}(du + ixy\,dv)\right]. \quad (2.16)$$

The solution given by (2.14) and (2.16) is characterized by

$$\begin{aligned} C^{(3)} &= -2ni/(y+ix)^3 + 2(e^2+g^2)/((y+ix)^2(y-ix)), \\ C_{12} &= -(e^2+g^2)/(y^2+x^2)^2, \\ \Gamma_{312} &= -1/(y+ix). \end{aligned} \quad (2.17)$$

The evaluation of these quantities can be carried out from expressions (1.21), (1.22), or (1.6), and (1.24), respectively.

The parameters n, e, and g are interpretable as magnetic mass (NUT parameter), electric and magnetic charges, respectively.

The metric with K from (2.13) is reducible, by means of the transformation

$$\{x, y, u\ v\} \to \{x^{-1},\ -y^{-1},\ \nu^2 v,\ \nu^2 u\},$$

to the metric (2.14). The corresponding ω is equivalent to the one given by (2.16).

The metric structure (2.14) and (2.16) has been presented in Ref. 1, Eq. (2.26), as the first known example of null orbit electrovac type D solutions. It should be noted that this solution is a special case of a solution obtained by Leroy[3] [with $b = 0$ in Eq. (3.36) of Ref. 3] under different hypothesis, and it also occurs as a special case of a solution obtained by Debever[4] [with $b = 0$ in Eq. (2.12) of Ref. 4].

Case $\alpha \neq 0$: The general solutions satisfying Eqs. (1.13) and (2.2) are given by

$$\begin{aligned} m &= A_1 e^{\sqrt{-\alpha}x} + A_2 e^{-\sqrt{-\alpha}x} + \beta/\alpha, \\ p &= B_1 e^{\sqrt{\alpha}y} + B_2 e^{-\sqrt{\alpha}y} + \beta/\alpha, \end{aligned} \quad (2.18)$$

with the constants A_i and B_i $(i = 1, 2)$ constrained to

$$A_1A_2 - B_1B_2 = 0. \quad (2.19)$$

By subjecting the coordinates u and v to a translation when m and p from (2.18) are substituted into the metric, the constant β cancels out. Therefore one always can choose β equal to zero.

With all generality, the constant α can be assumed positive, say, $\alpha = \kappa^2 \neq 0$. (For $\alpha < 0$, the hyperbolic functions below are to be replaced by the trigonometric one, and inversely.) From the reality of m, one has $A_2 = \overline{A}_1$, and $A_1 = |A_1|e^{i\tau}$, where τ is a real constant. Consequently,

$$m = |A_1|(e^{i\kappa(x+\tau/\kappa)} + e^{-i\kappa(x+\tau/\kappa)}),$$

changing the variable x to $x - \tau/\kappa$ one obtains

$$m = 2|A_1|\cos\kappa x, \quad |A_1| \neq 0. \quad (2.20)$$

To fulfill relation (2.19), it is sufficient to choose B_i as

$$B_1 = |A_1|e^{\sigma}, \quad B_2 = |A_1|e^{-\sigma},$$

where σ is a real constant. Substituting these constants B_i into p from (2.18), and transforming y to $y - \sigma/\kappa$, one obtains

$$p = 2|A_1|\cosh \kappa y. \tag{2.21}$$

Integrating the equations for the function H, one arrives at the general solution

$$H = 4\cos(\kappa/2)x \cosh(\kappa/2)y \times\{\mu + \nu \tan(\kappa/2)x \tanh(\kappa/2)y\}, \tag{2.22}$$

where μ and ν are arbitrary real constants. This expression suggests the change of variables

$$x' = \tan(\kappa/2)x, \quad y' = \tanh(\kappa/2)y, \tag{2.23}$$

under which the functions m, p, and H^2 reduce to

$$m = 2|A_1|(1-x'^2)/(1+x'^2),$$
$$p = 2|A_1|(1+y'^2)/(1-y'^2),$$
$$H^2 = (16/(1-y'^2)(1+x'^2))\{\mu + \nu x'y'\}^2. \tag{2.24}$$

Substituting these expressions into the metric (1.20), and executing there the transformation (2.23), accompanied by the redefinition of K and the coordinates u and v according to

$$K = -\frac{|A_1|}{\kappa^2(1+x'^2)^2}K', \quad u \pm 2|A_1|v = 8\kappa\begin{cases} u', \\ -v', \end{cases}$$

dropping primes, one arrives at the metric

$$g = \frac{1}{(\mu+\nu xy)^2}\left\{\frac{K}{x^2+y^2}(du + y^2dv)^2 + \frac{x^2+y^2}{K}dx^2 + 2\,dy(du - x^2dv)\right\} \tag{2.25}$$

with $K(x)$ constrained to fulfill the same equation as for the branch of solutions with vanishing separation constant. Notice also that the structural function H is the same in both cases. Therefore having demonstrated the equivalence of both cases, one concludes the only null orbit D solution with complex expansion is that given by formulas (2.14) and (2.16).

III. NULL ORBIT TYPE D SOLUTIONS WITH VANISHING COMPLEX EXPANSION

For the sake of completeness, we would like to present our previous results[5] dealing with divergenceless solutions from the point of view of the present treatment.

The conditions under which the congruence e^4 is free of complex expansion are

$$m_x = 0 = H_y + \tfrac{1}{2}(p_y/(m-p))H. \tag{3.1}$$

Therefore being that $m = m_0 = \text{const}$, from Eq. (1.13), one infers

$$p = m_0 - p_0 e^{\kappa y}, \tag{3.2}$$

where p_0 and κ are constants different from zero.

The equations for H now reduce to

$$H_{xx} + (\kappa/2)^2 H = 0,$$
$$H_{yy} - (\kappa/2)^2 H = 0, \tag{3.3}$$
$$H_y - (\kappa/2)H = 0,$$

the integration of which yields

$$H = e^{(1/2)\kappa y}\cos(\kappa/2)x(\mu + \nu\tan(\kappa/2)x). \tag{3.4}$$

Subjecting the coordinates in (1.20) to the transformation

$$x' = l\tan(\kappa/2)x, \quad y' = e^{-\kappa y},$$
$$u' = -(1/\kappa)(u + m_0 v), \quad v' = (2p_0 l/\kappa)v,$$
$$(\mu' = \mu l), \tag{3.5}$$

and redefining the structural function K according to

$$K' = (\kappa^2/4l^2p_0)(l^2 + x'^2)^2 K, \tag{3.6}$$

one brings the metric (1.20), dropping primes, to the form

$$g = \frac{1}{(\mu+\nu x)^2}\left\{\frac{K(x)}{l^2+x^2}(dv + 2ly\,du)^2 + \frac{l^2+x^2}{K(x)}dx^2 + 2(l^2+x^2)dy\,du\right\}, \tag{3.7}$$

with $K(x)$ constrained to

$$-4\lambda(l^2+x^2) = (\mu+\nu x)^2 K_{xx} - 6\nu(\mu+\nu x)K_x + 12\nu^2 K. \tag{3.8}$$

Notice that formulas (3.7) and (3.8) are meaningful even in the case of l equal to zero.

The presence of a conformal factor in the metric (3.7) suggests making a homographic transformation of the variable x. In this way, one arrives at the metrical structure equivalent to the previous one, but now with ν equal to zero. Thus the general nondiverging and nontwisting null orbit type D solutions are

$$g = \frac{K(x)}{l^2+x^2}(dv + 2ly\,du)^2 + \frac{l^2+x^2}{K(x)}dx^2 + 2(l^2+x^2)dy\,du,$$
$$K(x) = \alpha + \beta x - \lambda(\tfrac{1}{3}x^4 + 2l^2x^2), \tag{3.9}$$

where α and β are constants related to charges and magnetic mass; $\alpha = -(e^2+g^2) + \lambda l^4$, and $\beta = 2n$. The term λl^4 in α is useful from the viewpoint of limiting transitions. The electromagnetic 2-form ω accompanying (3.9) is

$$\omega = -(e+ig)d\left\{\frac{1}{x+il}(i\,dv - (x-il)y\,du)\right\}, \tag{3.10}$$

and the curvature quantities which characterize the solution (3.9), (3.10) are

$$C^{(3)} = 2/(x+il)^3\{n + \tfrac{4}{3}i\lambda l^3 - (e^2+g^2)/(x-il)\}, \quad R = -4\lambda,$$
$$C_{12} = -(e^2+g^2)/(x^2+l^2). \tag{3.11}$$

As was mentioned, the parameter e and g are interpretable as electric and magnetic charges, respectively, while n and l represent the NUT (magnetic mass) and the rotation parameters.

To obtain directly, from formulas (3.7) and (3.8), a null orbit exceptional electrovac solution [which does not satisfy condition (iii)], one equates l and μ to zero, and accomplishes the transformations $\{v', y'\} = \nu^{-2}\{v, y\}$, $x' = x^{-1}$, $K' = \nu^2 K x^{-4}$ [the (3.8) equation becomes $K'_{x'x'} = -4\lambda$], arriving, dropping primes, at

$$g = K\,dv^2 + K^{-1}dx^2 + 2\,dy\,du,$$
$$\omega = (e+ig)d\{ix\,dv + y\,du\}, \tag{3.12}$$

where K is always reducible to $1 - 2\lambda x^2 = K$. This solution is characterized by $C^{(3)} = -\tfrac{2}{3}\lambda$, $C_{12} = -(e^2+g^2) = -\lambda$,

($C_{12} = \frac{3}{2} C^{(3)}$), and $R = -4\lambda$. The constant e and g are the electric and magnetic charges, respectively.

On the basis of the results[6,7] dealing with all possible exceptional D's with an aligned electromagnetic field along the DP directions, one concludes that solution (3.12) is the only null orbit exceptional type D solution. It belongs to the Bertotti–Robinson[8,9] class of metrics; therefore, it can be named the Bertotti–Robinson null orbit exceptional electrovac solution.

IV. CONCLUSIONS

The results of this paper can be summarized by the following statement.

All null orbit type D solutions of the Einstein–Maxwell equations for an aligned general electromagnetic field along the double DP vectors, one of which is always free of complex expansion, belong to two disjoint classes. First, the class of solutions with the second DP vector possessing complex expansion reduce to a single solution, given by formulas (2.14) and (2.15). The second class is given completely, modulo contractions, by the free of complex expansion null orbit solutions, formulas (3.9) and (3.10).

The only exceptional null orbit solution, formula (3.12), belongs to the Bertotti–Robinson class of solutions with positive λ.

Our present and previous results[10] confirm our conjecture, stated in Ref. 10, that null orbit solutions can be obtained as limiting contractions of non-null orbit solutions.

V. REMARKS ON NON-NULL ORBIT TYPE D ELECTROVAC METRICS ADDED IN THE REFEREEING TIME

At the refereeing period of the present work, we have succeeded in obtaining, applying the same integration procedure developed here, all type D electrovac solutions with λ for an algebraically general electromagnetic field aligned along the DP directions of the Weyl tensor and possessing a group of symmetries with non-null orbits,[11] i.e., the metric structure given by formulas (T1). In this case, the metric with signature + 2 and the 2-form of the electromagnetic field can be given as

$$g = H^{-2}(x,y)\left\{\frac{m(x)-p(y)}{K(x)}\,dx^2 + \frac{K(x)}{m(x)-p(y)}\,(du + p(y)\,dv)^2 + \frac{m(x)-p(y)}{Q(y)}\,dy^2 - \frac{Q(y)}{m(x)-p(y)}\,(du + m(x)\,dv)^2\right\},$$

$$\omega = H^{-2}(\mathscr{E}(x,y) + i\check{\mathscr{B}}(x,y))\{(du + m\,dv)\wedge dy + i(du + p\,dv)\wedge dx\}.$$

The structural functions H, m, and p have to satisfy the system of equations (1.13), (1.16)–(1.19) studied in this report. The functions $\mathscr{E}$ and $\check{\mathscr{B}}$ ought to fulfill Eq. (1.23) of this text. The structural functions $K(x)$ and $Q(y)$ satisfy a generalized scalar curvature equation; the second member of Eq. (1.20) presented here acquires the term $D_y Q$, which is solvable by separation of variables. Note that by executing in the metric above the transformations

$$dv \to dv - \frac{dy}{Q}, \quad du \to du + \frac{p}{Q}\,dy,$$

and sending Q to zero, one arrives just at the null orbit metric (1.15) studied here. A detailed determination of all solutions of the quoted equations for the non-null orbit class of metrics is given in Ref. 11. The most general non-null orbit electrovac type D solution, modulo all possible contractions, is the one obtained early by Plebański and Demiański,[12] which contains as limiting transitions the null orbit solutions determined in the present paper and also all vacuum with λ type D metrics.[13]

According to the referee's report of the present work, Debever, Kamran, and McLenaghan[14] have recently obtained a single expression for the general type D solution of Einstein vacuum and electrovac field equations with λ, with a nonsingular aligned Maxwell field if present, which contains the null orbit solutions as a special case.

ACKNOWLEDGMENT

One of us (H.S.I.) acknowledges the Consejo Nacional de Ciencias y Tecnología for financial assistance through a doctoral fellowship.

[1] R. Debever and R. G. McLenaghan, J. Math. Phys. **8**, 1711 (1981); related results are also in Bull. Cl. Sci. Acad. R. Belg. **66**, 204 (1980).
[2] R. Debever and N. Kamran, Bull Cl. Sci. Acad. R. Belg. **66**, 585 (1980).
[3] J. Leroy, Bull. Cl. Sci. Acad. R. Belg. **64**, 130 (1978).
[4] R. Debever, Bull. Soc. Math. Belg. **XXIII**, 360 (1971).
[5] A. García D. and H. Salazar I. "All free of complex expansion null orbits type D electrovac solutions with λ," Gen. Rel. Grav. **15**, 417 (1983).
[6] J. F. Plebański and S. Hacyan, J. Math. Phys. **20**, 1004 (1979).
[7] A. García D. and J. F. Plebański, J. Math. Phys. **23**, 123 (1982).
[8] B. Bertotti, Phys. Rev. **116**, 1331 (1959).
[9] I. Robinson, Bull. Acad. Pol. Sci. Ser. Math. Phys. **7**, 352 (1959).
[10] A. García D. and J. F. Pelbański, J. Math. Phys. **23**, 1463 (1982).
[11] A. García D. "Electrovac type D solutions with cosmological constant," submitted to J. Math. Phys. (1983).
[12] A. García D. and H. Salazar I. "Type D vacuum solutions with cosmological constant," submitted to Gen. Rel. Grav. (1983).
[13] J. F. Plebański and M. Demiański, Ann. Phys. (N.Y.) **98**, 98 (1976).
[14] R. Debever, N. Kamran, and R. G. McLenaghan, Phys. Lett. A **93**, 399 (1983).

Cluster expansion of the distribution functions for a ground state Fermi system

S. K. Sinha
Department of Physics, L. S. College, Bihar University, Muzaffarpur 842001, India
Y. Singh
Department of Physics, Banaras Hindu University, Varanasi-221005, India

(Received 15 January 1982; accepted for publication 30 September 1982)

The spin-averaged Slater sum of the fermion system is expanded in terms of the square of the ground state wavefunction of a boson system and the "antisymmetry" Ursell function. This expansion is used to obtain the cluster series for the radial distribution function of the fermion system in terms of $(-\mathscr{C}^{(n)}/S)$, where $\mathscr{C}^{(n)}$ is sum of chains of $(-f/S)$ and $(-fh^{B}_{2}/S)$ bonds. The series is further expressed in a more compact form using a function $L^{(n)}$ defined by Eq. (55), and the "modified" FHNC approximation for the radial distribution function is presented.

PACS numbers: 05.30.Fk

I. INTRODUCTION

The aim of this paper is to develop a theory for calculating the radial distribution function (RDF) for the infinite homogeneous ground state fermion system. In the variational method, a trial many-body wavefunction for such a system is chosen as[1,2]

$$\psi_{\mathrm{F}} = \psi_{\mathrm{B}}\Phi, \tag{1}$$

where ψ_{B} is the exact boson ground state wavefunction and Φ is an antisymmetric function describing a state of the N-particle system in the absence of interaction. For a normal Fermi liquid, Φ is taken to be a Slater determinant constructed from products of plane wave orbital and spin functions[3,4]

$$\Phi = a\left[\prod_{i=1}^{N}\exp(i\bar{k}_i\cdot\bar{r}_i)\xi_i(m_i)\right], \tag{2}$$

where a is the antisymmetrizer and ξ_i are spin functions. The RDF is defined in canonical ensemble in terms of ψ_{F} by

$$\rho^2 g^{\mathrm{F}}(1,2)\equiv\rho^2 g_2^{\mathrm{F}}(1,2) = \frac{N(N-1)\int\cdots\int d\bar{r}_3\cdots d\bar{r}_N|\psi_{\mathrm{F}}|^2}{\int\cdots\int d\bar{r}_1\cdots d\bar{r}_N|\psi_{\mathrm{F}}|^2}. \tag{3}$$

Iwamoto and Yamada[5] and others[6–8] have used Eq. (1) to develop a cluster expansion method for calculating the distribution functions. In their methods, the cluster series was developed in terms of the boson distribution functions and was truncated at the third term involving three-body distribution functions. Since the antisymmetry of Eq. (1) was expressed in a straightforward permutations, the truncated series takes into account only low-order exchange cluster expansion. As Zabolitzky[9] has recently shown, such expansion can lead to convergence only at low densities. At higher densities this series has been found to diverge. Though any finite order of permutation may not guarantee convergence of the series, it has, however, been argued that this approach can be made successful by considering large number of exchange clusters.[9,10] The more number of clusters are included, the better is the convergence.

The other method, which has been developed recently using a Bijl–Jastrow form[11] of the wavefunction ψ_{B},

$$\psi_{\mathrm{B}} = \prod_{i<j}\exp[\tfrac{1}{2}u_2(i,j)] = \prod_{i<j}[1+F_{ij}]^{1/2}, \tag{4}$$

involves the expansion of distribution function in terms of diagrams with F- and f-bonds, where $F_{ij}=\exp[u_2(i,j)]-1$ and $f(k_{\mathrm{F}}r_{ij})$ is defined in Eq. (13) below. In this case, a diagram is a collection of circles (vertices) and F- and f-bonds connecting some pairs of circles. The f-bonds always connect circles in such a way that they form a closed loop or ring. There are two types of circles—black and white. The white circles are labelled, but the black circles are unlabelled. The value of a diagram is defined in terms of these functions and an integration over the positions, which can be assigned to each black circle. In what is popularly known as Fermi-hypernetted-chain (FHNC) approximation, the cluster series is summed by neglecting the terms, which correspond to elementary diagrams $E(1,2)$, where $E(1,2)$ diagrams are both 1,2-irreducible and free of bridge points.[12,13] By choosing two different subseries to be summed, two different FHNC methods were obtained: One was developed by Krotscheck and Ristig[14] (KR), and other was due to Fantoni and Rosati[15] (FR). Recently, Zabolitzky[9] has used the FR–FHNC method to calculate the properties of liquid ^{3}He and two model fermion liquids. He has found that this method gives good results for the ground state energy at low densities and short range correlation functions. For high densities and/or long range correlation functions, the results are not so good. Further, he has found that while the FR–FHNC gives reasonable values of $g^{\mathrm{F}}(r)$ for small values of r, the KR–FHNC overestimates it.

By partial summation in the FHNC method, one has tried to take into account the antisymmetry correctly, but the method as such is not expected to give good results, at least at high densities, because of certain approximations involved in the theory. Attempts have, however, been made to improve upon FHNC theory by evaluating the first few elementary diagrams appearing in the expansion and incorporating them as a link in the chains generated by FHNC

 0022-2488/83/102504-08$02.50

procedure. This gives FHNC/4 approximation. The FHNC/4 approximation can, therefore, be regarded as a first step beyond the FHNC of a systematic procedure that eventually sums all diagrams. However, to go beyond the FHNC/4 approximation for Fermi liquid is extremely difficult. Apart from this, the FHNC method has been developed using the Bijl–Jastrow type of wavefunction for ψ_B, which is approximate one. One may improve the result by using the exact wavefunction ψ_B. Thus we believe that a theory, which is developed using the exact boson wavefunction ψ_B and taking the partial summations of the permutation expansion for antisymmetry, will be better than the FHNC or its improved form.

In this paper, we develop a cluster expansion method for evaluating the RDF of the fermion system, in which the series is expressed in terms of the boson distribution functions and the permutation expansions or antisymmetry are partially summed. This approach differs from the FHNC method in selection of the subseries of diagrams to be summed in the permutation expansion for the antisymmetry and by expressing the series in terms of the boson distribution functions. We use the grand canonical ensemble and functional differentiation technique to derive the results.

In Sec. II, we give the basic theory for a fermion system. Section III is devoted to develop expansion at constant fugacity z and obtain expansion of the l-particle Ursell and density distribution function of fermion system in terms of graphs. The series obtained in Sec. III is reduced at constant density ρ by means of topological reduction technique in Sec. IV. Using Kirkwood's superposition approximation,[16] the series is expressed in terms of h_2^B-bonds, where $h_2^B(i,j) = g_2^B(i,j)$-1 is the pair correlation function of the boson system. The series is further reduced in Sec. V in terms of $\mathscr{C}^{(n)}(\bar{r}_i, \bar{r}_j)$, which is defined by Eq. (42). The expression for the RDF is reported there in terms of $\mathscr{C}^{(n)}$. Section VI is devoted to reducing the series in more compact form and to discuss the "modified" FHNC approximation.

II. GENERAL FORMALISM

We define the spin-averaged Slater sum for the fermion system as

$$W_N^F(1,2,...,N) = (1/A)\sum_\sigma |\psi_F|^2, \tag{5}$$

where Σ_σ indicates the summation over all spin (discrete) states and the constant A is the norm of the wavefunction ψ_F. Using (1), we can write (5) in the form

$$W_N^F = W_N^B W_N^A, \tag{6}$$

where

$$W_N^B = |\psi_B|^2 \tag{7}$$

is the square of the ground state wavefunction for the interacting boson system and

$$W_N^A = (1/A)\sum_\sigma |\Phi|^2 = (1/A)\sum_\sigma \left| a\left[\prod_{i=1}^N e^{i\bar{k}_i\cdot\bar{r}_i}\xi_i(m_i)\right]\right|^2. \tag{8}$$

W_N^A may be called as the spin-averaged sum of the noninteracting Fermi gas of N particles. Using relation (8), we can obtain

$$W_1^A(1) = 1, \tag{9a}$$

$$W_2^A(1,2) = 1 + U_2^A(1,2), \tag{9b}$$

$$W_3^A(1,2,3) = 1 + U_2^A(1,2) + U_2^A(1,3) + U_2^A(2,3) + U_3^A(1,2,3), \tag{9c}$$

$$W_N^A(1,2,...,N) = 1 + \sum U_2^A(i,j) + \sum U_3^A(i,j,k) + \sum U_2^A(i,j)U_2^A(k,l) + \sum U_4^A(i,j,k,l) + \cdots, \tag{9d}$$

where

$$U_2^A(1,2) = -(1/S)f^2(k_F r_{12}), \tag{10a}$$

$$U_3^A(1,2,3) = (2/S^2)f(k_F r_{12})f(k_F r_{23})f(k_F r_{31}), \tag{10b}$$

$$U_4^A(1,2,3,4) = -(2/S^3)[f(k_F r_{12})f(k_F r_{23})f(k_F r_{34})f(k_F r_{41}) + f(k_F r_{12})f(k_F r_{24})f(k_F r_{43})f(k_F r_{31}) + f(k_F r_{13})f(k_F r_{32})f(k_F r_{24})f(k_F r_{41})], \tag{10c}$$

for $m > 2$

$$U_m^A(1,2,...,m) = (-S)^{1-m}2\sum f(k_F r_{12})f(k_F r_{23})\cdots f(k_F r_{m1}), \tag{10d}$$

and

$$f(k_F r) = \frac{S}{(2\pi)^3\rho}\int_{k<k_F} d\bar{k}\exp(i\bar{k}\cdot\bar{r}) = 3[\sin(k_F r) - (k_F r)\cos(k_F r)]/(k_F r)^3. \tag{11}$$

Here $k_F = (6\pi^2\rho/S)^{1/3}$ is the Fermi momentum at density ρ and the factor S gives the degeneracy of the system ($S = 2$ for paramagnetic state of ^{3}He or neutron matter; $S = 4$ for nuclear matter).

In analogy to the grand partition function of classical statistical mechanics,[12] we define a generating functional

$$\Xi^F = \sum_{N=0}^\infty \frac{1}{N!}\int\cdots\int\prod_{i=1}^N z(i)W_N^F(1,2,...,N)\prod_{i=1}^N d\bar{r}_i \tag{12}$$

for the distribution functions of the fermion system. The spin-averaged l-particle distribution function for the fermion system may be defined as

$$n_l^F(1,2,...,l) = \Xi^{F^{-1}}\sum_{N\geq l}^\infty \frac{1}{(N-l)!} \times\int\cdots\int\prod_{i=1}^N z(i)W_N^F(1,2,...,N)\prod_{i=l+1}^N d\bar{r}_i, \tag{13}$$

where

$$z(i) = z\exp[v(i)]$$

z is the fugacity and $v(i)$ is a function related to the potential energy of a particle at $\bar{r}_i$ due to external forces.

The l-particle distribution function can also be obtained

from its relationship with the Ursell function[12]

$$n_l^X(1,2,...,l) = \sum_{\Sigma\alpha p_\alpha = l} \pi\chi_l^X(i_i, i_2, ..., i_p), \tag{14}$$

where the sum of the products is carried out over all possible divisions of l particles with the condition that $\Sigma\alpha\, p_\alpha = l$. The l-particle Ursell function is defined by a functional derivative of $\ln\Xi^X$,[12]

$$\chi_l^X(1, 2, ..., l) = \prod_{1<i<l} z(i)\frac{\partial^l \ln\Xi^X}{\Pi_{1<i<l}\partial z(i)}. \tag{15}$$

Here X stands for F or B system.

In the following section, we use these relations to obtain cluster expansion for the distribution functions.

III. CLUSTER EXPANSION AT CONSTANT FUGACITY

Substituting (6)–(10) in (12), we obtain the graphical expansion of Ξ^F at constant fugacity in terms of the composite graphs, with f-bonds (represented by dashed line with arrow) connecting vertices to form closed loop and χ_m^B-polyhedron (represented by a vertex, solid line, shaded triangle ··· for $m = 1, 2, 3, ...$, respectively). Thus

$$\Xi^F/\Xi^B = 1 + \sum_{m>2} \Gamma_m(z), \tag{16}$$

where $\Gamma_m(z)$ is a composite graph with m black vertices, no white vertices, some f-bonds connecting vertices to form at least one closed loop with χ_α^B-polyhedron ($1\leqslant\alpha\leqslant m$), each loop with p vertices is multiplied by a factor $(-S)^{1-p}$ and each vertex is attached to at most one f-loop, and one χ_α^B-polyhedron. The closed loop may or may not form around χ_α^B. The orthogonality condition is not considered here. It will be considered later when we discuss correlation functions.

Using Lemma 3 of Ref. 13, we obtain the graphical expansion of $\ln\Xi^F$:

$$\ln\Xi^F = \ln\Xi^B + \xi(z), \tag{17}$$

where $\xi(z)$ is sum of all distinct connected composite graphs (CCG) with no white vertices, some f-bonds forming closed loops, with χ_α^B-polyhedron ($\alpha\geqslant 1$), with the same restrictions and multiplying factor as in (16).

Let Γ be a graph of (17). Then it can be shown using the formula

$$z(1)\frac{\partial\chi_{n-1}^X(2,...,n)}{\partial z(1)} = \chi_n^X(1, 2, ..., n) + \sum_{i=2}^{n}\delta(1-i)\chi_{n-1}^X(2, ..., n) \tag{18}$$

that

$$\prod_{i=1}^{n} z(i)\frac{\partial^n\Gamma}{\Pi_{i=1}^n\partial z(i)}$$

= [sum of all distinct graphs obtained from Γ by changing n black vertices into white vertices labeled 1, 2, ..., n]

+ [sum of all distinct graphs obtained by inserting at most n vertices labeled 1, 2, ..., n]. (19)

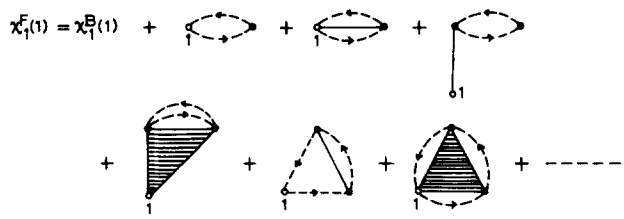


FIG. 1. The first few graphs of $\chi_1^F(1)$: the f-bond is represented by a dashed line with arrow and χ_m^B by a vertex, solid line, shaded triangle... for $m = 1, 2, 3, ...$.

With the aid of (15), (17), and (19), and using the orthogonality condition, we obtain expansion of $\chi_l^F(1,2,...,l)$ at constant z:

$$\chi_l^F(1, 2, ..., l) = \chi_l^B(1, 2, ..., l) + \chi_l^A(1, 2, ..., l), \tag{20}$$

where $\chi_l^A(1,2,...,l)$ is sum of all distinct CCG with l white vertices labeled 1, 2, ..., l, respectively, some or no black vertices, some f-bonds forming closed loops, with χ_α^B-polyhedron ($\alpha\geqslant 1$) of nonoverlapping set of vertices, each loop with p vertices ($p\geqslant 2$) is multiplied by a factor $(-S)^{1-p}$, and each black vertex is attached to an f-loop and χ_α^B-polyhedron ($\alpha\geqslant 2$). The first few graphs for one- and two-particle Ursell functions are shown in Figs. 1 and 2, respectively. Using relation (14), we get

$$n_l^F(1, 2, ..., l) = n_l^B(1, 2, ..., l) + n_l^A(1, 2, ..., l), \tag{21}$$

where $n_l^A(1,2,...,l)$ is sum of all distinct composite graphs (CG) and white vertices labeled 1, 2, ..., l, respectively, some or no black vertices, some f-bonds forming closed loops, with χ_α^B-polyhedron ($\alpha\geqslant 1$), with same restriction and multiplying factor as in (20), and each component contains at least one white vertex.

From (21), evaluating the first few graphs, we may obtain expressions for $n_1^F(1)$ and $n_2^F(1,2)$ at constant fugacity z in

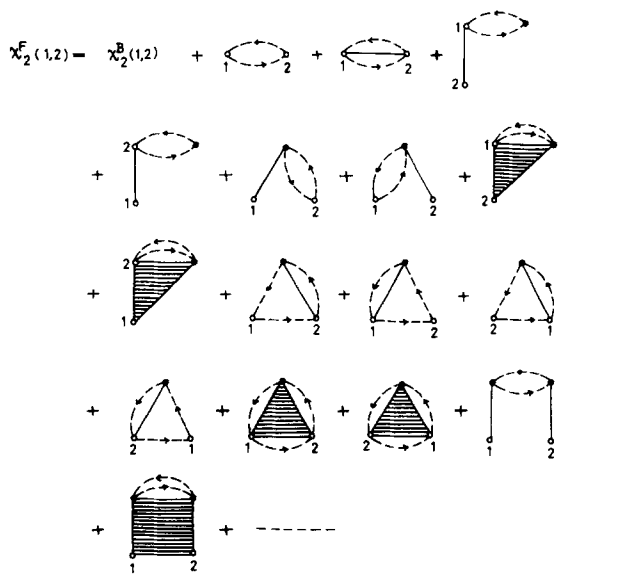


FIG. 2. The first few graphs of $\chi_2^F(1,2)$; the symbols are the same as in Fig. 1.

the first order of f^2:

$$n_1^F(1) = n_1^B(1) - (1/S)\int n_2^B(1,2) f^2(k_F r_{12})\, d\bar{r}_2 - (1/2S)\int [n_2^B(1,2,3) - n_2^B(2,3) n_1^B(1)] \times f^2(k_F r_{23})\, d\bar{r}_2\, d\bar{r}_3 + \cdots \quad (22)$$

and

$$n_2^F(1,2) = n_2^B(1,2)[1 - (1/S) f^2(k_F r_{12})] - (2/S)\int n_3^B(1,2,3) f^2(k_F r_{13})\, d\bar{r}_3 - (1/2S)\int [n_4^B(1,2,3,4) - n_2^B(1,2) n_2^B(3,4)] \times f^2(k_F r_{34})\, d\bar{r}_3\, d\bar{r}_4 + \cdots. \quad (23)$$

IV. EXPANSION OF CORRELATION FUNCTIONS AT CONSTANT DENSITY

In this section, we obtain the expansion of correlation fucntions at constant density ρ. This may be obtained either by Taylor expansion or by topological reduction technique. Here we used topological reduction technique to obtain the result. Instead of using n_l^X and χ_l^X, it is convenient to use the correlation functions, defined by the relations[12]

$$g_l^X(1, 2, ..., l) = \frac{n_l^X(1, 2, ..., l)}{n_1^X(1) n_1^X(2) \cdots n_1^X(l)}, \quad \text{for } l \geqslant 2, \quad (24)$$

$$h_l^X(1, 2, ..., l) = \frac{\chi_l^X(1, 2, ..., l)}{n_1^X(1) n_1^X(2) \cdots n_1^X(l)}, \quad \text{for } l \geqslant 2, \quad (25)$$

$$h_1^X(1) = \ln[\chi_1^X(1)/z(1)] \quad (26)$$

and

$$g_2^X(1, 2) = 1 + h_2^X(1, 2). \quad (27)$$

In order to obtain the cluster expansion of correlation fucntions at constant density ρ, we first express χ_α^B-faces of (20) in terms of h^B-faces by using the relation (25) and then introduce a function s_m^B, defined by

$h_l^B(1, 2, ..., l)$ = [sum of all distinct simply connected simple graphs, consisting of l white 1 vertices labeled 1, 2, ..., l, respectively, and s_m^B-faces $(2 \leqslant m \leqslant l)$], (28)

giving

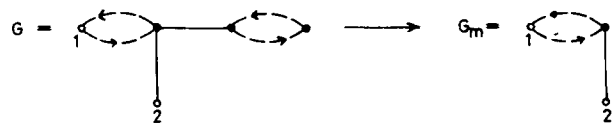
$$h_2^B(1, 2) = s_2^B(1, 2),$$
$$h_3^B(1, 2, 3) = s_2^B(1, 2) s_2^B(2, 3) + s_2^B(1, 2) s_2^B(1, 3) + s_2^B(1, 3) s_2^B(2, 3) + s_3^B(1, 2, 3),$$

and so on.

With the help of (25) and (28), we obtain from (20)

$$\chi_2^F(1, 2) = n_1^F(1) n_1^F(2) h_2^F(1, 2) = n_1^B(1) n_1^B(2) h_2^B(1, 2) + E_2^A(1, 2), \quad (29)$$

where $E_2^A(1,2)$ is sum of CCG with two white n_1^B-vertices labeled 1, 2, respectively, some or no black n_1^B-vertices, some or no s_m^B-faces $(m \geqslant 2)$, some f-bonds forming closed loops, with the same restriction and multiplying factor as in (20). $E_2^A(1,2)$ has been obtained by transforming χ_l^B of the left-hand side of Eq. (20) into s_m^B, using Eqs. (25) and (28). Now it is desirable to transform the n_1^B-expansion of $\chi_2^F(1,2)$ into the n_1^F-expansion. This is done by using topological reduction. Every graph G in the set of Eq. (29) contains a maximal 1-irreducible graph G_m defined as the largest subgraph of G, which contains all the white circles but contains no articulation circles. For example,



It is clear that a set of graphs on the right-hand side of Eq. (29) can be obtained by starting with G_m and replacing each black vertex of G_m by some graph Γ_α of the expansion of $n_1^F(1)$ shown in Fig. 1. The replacement can always be done so that the graph Γ_α is attached to G_m by the white vertex of Γ_α, which is first stripped of its label and blackened. From this (see, for example, Lemma 4 of Ref. 13), we obtain the expression for the pair correlation function $h_2^F(1,2)$ at constant density:

$$\rho^2 h_2^F(1,2) = \rho^2 h_2^B(1,2) + \xi_2^A(1,2), \quad (30)$$

where $\xi_2^A(1,2)$ is sum of CCG with two white ρ vertices labeled 1, 2, respectively, some or no black ρ vertices, some or no s_m^B-faces $(m \geqslant 2)$, some f-bonds forming closed loops, with the same restriction and multiplying factor as in (20), no articulation vertices, and no articulation pair vertices. Here $n_1^F(1) = n_1^F(2) = \cdots = \rho$, and we use the conventional graphical terminology.[12,13]

Equation (30) gives a graph theoretic recipe for the general term in the expansion of the pair correlation function in terms of the exchange function f at the constant density ρ. Evaluating the first few graphs of (30), we obtain the expression for $g_2^F(1,2)$ in the first order of f^2:

$$g_2^F(1,2) = g_2^B(1,2)[1 - (1/S) f^2(k_F r_{12})] - (2/S)\rho\int [g_3^B(1,2,3) - g_2^B(1,2) g_2^B(1,3)] \times f^2(k_F r_{13})\, d\bar{r}_3 - (1/2S)\rho^2\int [g_4^B(1,2,3,4) - 2 g_2^B(1,2) g_3^B(1,3,4) + g_2^B(1,2) g_2^B(3,4)] f^2(k_F r_{34})\, d\bar{r}_3\, d\bar{r}_4 + \cdots. \quad (31)$$

If we evaluate a few different sets of graphs of (30), we get

$$g_2^F(1,2) = g_2^B(1,2)[1 - 1/S f^2(k_F r_{12})] - (2/S)\rho\int [g_3^B(1,2,3) - g_2^B(1,2) g_2^B(1,3)] \times f^2(k_F r_{13})\, d\bar{r}_3 + (2\rho/S^2) f(k_F r_{12})\int g_3^B(1,2,3) \times f(k_F r_{23}) f(k_F r_{31})\, d\bar{r}_3 + \cdots. \quad (32)$$

Evaluating more graphs, one can obtain the higher-order exchange terms. Exact evaluation of the integrals appearing in (31) and (32) is not possible, because the values of correlation functions $g_3^B(1,2,3)$, $g_4^B(1,2,3,4)$, ··· are not known. Such integrals may be evaluated by using superposition approximation.[16] Under superposition approximation, Eq. (32) reduces to an Aviles–Harton–Tolhock (AHT) expression.[7,8]

In order to sum the series of (30), we express $s_m^B(1,2,...,m)$ in terms of $h_2^B(i,j)$ by using the superposition approximation and write

$s_m^B(1,2,...,m) =$ [sum of all distinct simply connected simple graphs, consisting of m white 1 vertices, labeled 1, 2, ..., m, respectively, and h_2^B-bonds, such that each vertex is attached to at least two h_2^B-bonds]. (33)

For example,

$$s_3^B(1,2,3) = h_2^B(1,2)h_2^B(1,3)h_2^B(2,3).$$

The consistency of Eq. (33) with Eq. (28) can be demonstrated by expanding $h_l^B(1,2,...,l)$ in terms of $h_2^B(i,j)$ by using the superposition approximation. When (33) is substituted in (30), we obtain the cluster series for the RDF of the fermion system in terms of h_2^B-bonds:

$$g_2^F(1,2) = g_2^B(1,2) + \mathscr{L}^A(1,2), \tag{34}$$

where $\mathscr{L}^A(1,2)$ is the sum of CCG with 1 vertices labeled 1, 2, respectively, some or no black ρ vertices, some or no h_2^B-bond between any two vertices, at most one h_2^B-bond, some f-bonds forming a closed loop, each loop with p vertices is multiplied by a factor $(-s)^{1-p}$, each black vertex is attached to an f-loop and at least one h_2^B-bond, at least one f-loop, no articulation vertices, and no articulation pair vertices. Some graphs of the RDF are shown in Fig. 3, where the white and black vertices have the vertex function unity and ρ, respectively. Here the h_2^B-bond is represented by a solid line and g_2^B-bond by a curly line.

This method can be extended to obtain the cluster expansions of the higher-body distribution functions, but we do not pursue this further in this paper. In the following sections, we develop a method for summing the series.

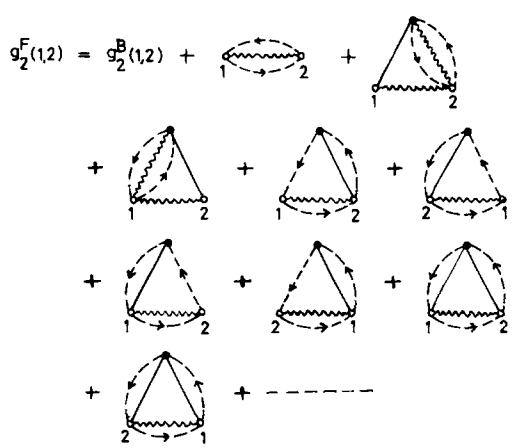


FIG. 3. The first few graphs of $g_2^F(1,2)$: the f-bond is represented by a dashed line with arrow, h_2^B and g_2^B, respectively, by the solid line and the curly line: white and black vertices have the vertex function 1 and ρ, respectively.

V. EFFECTIVE TWO-BODY EXCHANGE FACTOR AND THE SUMMATION OF SERIES

The expression for the RDF of the fermion system can be expressed in terms of a function, which is obtained by summing all the chains (graphs consisting of only one path of f-bonds superimposed or not on h_2^B-bonds joining the white vertices 1 and 2) appearing in (34). Such types of graphs with two or more f-bonds superimposed on an h_2^B-bond can be constructed by inserting f-bonds in all possible ways at the end or in the middle of the chains. If $\mathscr{C}^0(\bar{r}_1,\bar{r}_2)$ be the sum of the terms corresponding to all such graphs, then $\mathscr{C}^0(\bar{r}_1,\bar{r}_2)$ can be redefined as

$$\mathscr{C}^0(\bar{r}_1,\bar{r}_2) = f(k_F r_{12})[1 - h_2^B(\bar{r}_1,\bar{r}_2)] + B(\bar{r}_1,\bar{r}_2), \tag{35}$$

where

$(\rho^2/S^2)B(\bar{r}_1,\bar{r}_2) =$ [sum of chains of one or more f-bonds connecting two white $(-\rho/S)$ vertices superposing h_2^B-bonds in all possible ways such that each black vertex is connected with at least one h_2^B-bond, each black vertex has a vertex function $(-\rho/S)$]. (36)

We can write the series as

$$(\rho^2/S^2)B(\bar{r}_1,\bar{r}_2) = (\rho^2/S^2)\sum_{m=1}^{\infty} B^{(m)}(\bar{r}_1,\bar{r}_2) \tag{37}$$

where $B^{(m)}(\bar{r}_1,\bar{r}_2)$ is the sum of all such chains with m f-bonds superposed on h_2^B-bonds (i.e., m, "$f h_2^B$-" bonds). These graphs with $m = 1$ and 2 are shown in Figs. 4(a) and 4(b), respectively. These chains can be expressed in a more compact form, introducing a function defined by

$$F(\bar{r}_1,\bar{r}_2) = -(\rho/S)\delta(\bar{r}_1 - \bar{r}_2) + (-\rho/S)^2 f(k_F r_{12}), \tag{38}$$

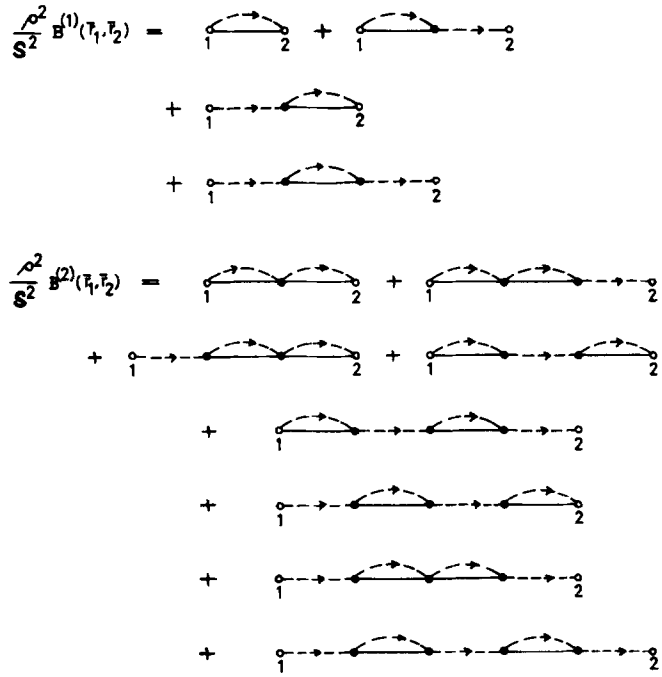


FIG. 4. Simple chains representing (a) $(\rho^2/S^2)B^{(1)}(\bar{r}_1,\bar{r}_2)$, (b) $(\rho^2/S^2)B^{(2)}(\bar{r}_1,\bar{r}_2)$; the symbols are the same as in Fig. 3.

where δ is the Dirac delta function. $F(\bar{r}_1, \bar{r}_2)$ is represented graphically by a large circle surrounded by two white or black circles corresponding to the coordinates $\bar{r}_i$ and $\bar{r}_j$. Then we can express $B^{(m)}$ for $n = 1$ and 2:

$(\rho^2/S^2)B^{(1)}(\bar{r}_1, \bar{r}_2) =$ 

$$= \int F(\bar{r}_1, \bar{x}_1) f(k_F x_{12}) \times h_2^B(\bar{x}_1, \bar{x}_2) F(\bar{x}_2, \bar{r}_2)\, d\bar{x}_1 d\bar{x}_2, \tag{39a}$$

$(\rho^2/S^2)B^{(2)}(\bar{r}_1, \bar{r}_2) =$ 

$$= \int F(\bar{r}_1, \bar{x}_1) f(k_F x_{12}) h_2^B(\bar{x}_1, \bar{x}_2) F(\bar{x}_2, \bar{x}_3) \times f(k_F x_{34}) h_2^B(\bar{x}_3, \bar{x}_4) F(\bar{x}_4, \bar{r}_2) \times d\bar{x}_1 d\bar{x}_2 d\bar{x}_3 d\bar{x}_4. \tag{39b}$$

With the help of above definitions, Eq. (37) can be summed using a Fourier transform giving

$$(\rho^2/S^2)B(r) = \frac{1}{(2\pi)^3} \int d\bar{k}\, e^{i\bar{k}\cdot\bar{r}} \left[\frac{F^2(k)\phi(k)}{1 - F(k)\phi(k)}\right], \tag{40}$$

where $F(k)$ and $\phi(k)$ are the Fourier transforms of $F(r)$ and $\phi(r) \equiv f(k_F r) h_2^B(r)$, i.e.,

$$F(\bar{k}) = \int d\bar{r}\, e^{i\bar{k}\cdot\bar{r}} F(r) = [1 + \tilde{f}](-\rho/S), \tag{41a}$$

$$\phi(k) = \int d\bar{r}\, e^{-i\bar{k}\cdot\bar{r}} f(k_F r) h_2^B(r), \tag{41b}$$

$$\tilde{f} = (-\rho/S) \int d\bar{r}\, e^{-i\bar{k}\cdot\bar{r}} f(k_F r) = -\theta(k_F - k). \tag{41c}$$

θ is the unit step function. From (35) and (40), we obtain

$$\mathscr{C}^0(\bar{r}_1, \bar{r}_2) = \frac{1}{(2\pi)^3(-\rho/S)} \int d\bar{k} \times \left[\frac{\tilde{f}(1 + \tilde{\beta}) + \tilde{\beta}^2(1 + \tilde{f})}{1 - \tilde{\beta}(1 + \tilde{f})}\right] e^{-i\bar{k}\cdot\bar{r}_{12}}, \tag{42}$$

where $\tilde{\beta} = (-\rho/S)\phi(k)$.

The link $\mathscr{C}^0$ is used to express the RDF of the fermion system. This is obtained by replacing f-bond by $\mathscr{C}^0$-bond in (34). We can repeat the process of summing the chain graphs, except that the basic link in each chain is not f but $\mathscr{C}^{(n-1)}$ to obtain the new link $\mathscr{C}^{(n)}$. Thus $\mathscr{C}^{(n)}(\bar{r}_1, \bar{r}_2)$ is expressed by (42) using $\mathscr{C}^{(n-1)}$ in place of f.

Evaluating graphs in terms of $\mathscr{C}^{(n)}$, (34) can be written as

$$g_2^{F(n)}(1, 2) = g_2^B(1, 2)\left[g_2^{A(n)}(1, 2) + \sum_{m=3}^{\infty} \Delta g_m^{A(n)}(1, 2)\right]. \tag{43}$$

The subscript m indicates the m-body contribution to the RDF. In the first order, the RDF for the fermion system is approximated to

$$g_2^{F(n)}(r) = g_2^B(r) g_2^{A(n)}(r), \tag{44}$$

where

$$g_2^{A(n)}(r) = [1 - (1/S)\mathscr{C}^{(n)^2}(r)] \tag{45}$$

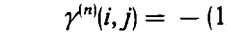
and $g_2^B(r)$ is the RDF arising from ψ_B. The analytic expression of the three-body contribution $\Delta g_3^{A(n)}(r)$ may be obtained from (34);

$$\Delta g_3^{A(n)}(1, 2) = -(2\rho/S) \int g_2^B(1,3) h_2^B(2,3) \mathscr{C}^{(n)^2}(1,3)\, d\bar{r}_3 + (2\rho/S)\mathscr{C}^{(n)}(1,2) \int [g_2^B(1,3) g_2^B(2,3) - 1] \mathscr{C}^{(n)}(2,3)\mathscr{C}^{(n)}(3,1)\, d\bar{r}_3. \tag{46}$$

Evaluating further classes of graphs, one can obtain expressions for $\Delta g_4^{A(n)}(r)$, $\Delta g_5^{A(n)}(r)$, ···. If $\mathscr{C}^{(n)}(r)$ is approximated to $f(k_F r)$, Eq. (43) gives the Iwamoto–Yamada[5] (IY) expansion for the RDF of the Fermi system.

VI. "MODIFIED" FHNC APPROXIMATION

We now define

$\gamma^{(n)}(i,j) = -(1/S)\mathscr{C}^{(n)^2}(\bar{r}_i, \bar{r}_j) \equiv$  . (47)

Since the two-particle exchange factor is expressed by $(-1/S)f^2(k_F r_{ij})$, one may call the function $\gamma^{(n)}(i,j)$ the *effective two-particle exchange factor.*

Using (47), (43) can be written as

$$g_2^F(1,2) = \hat{g}_2^F(1,2) + X(1,2). \tag{48}$$

Here $\hat{g}_2^F(1,2)$ is the RDF of the fermion system expressed in terms of $\gamma^{(n)}$:

$$\hat{g}_2^F(1,2) = g_2^B(1,2) + \hat{\mathscr{L}}^A(1,2), \tag{49}$$

where $\hat{\mathscr{L}}^A(1,2)$ is the sum of CCG with two white 1 vertices labeled 1, 2, respectively, some or no black ρ vertices, some $\gamma^{(n)}$-bonds, some or no h_2^B-bonds, each black vertex, which is attached to an $\gamma^{(n)}$-bond (superimposed or not an h_2^B-bond) is attached to an h_2^B-bond, no vertex is attached to more than one $\gamma^{(n)}$-bond, no articulation vertices and no articulation pair vertices. While $X(1,2)$ belongs to $\mathscr{L}^A(1,2)$ with three or more $\mathscr{C}^{(n)}$-bonds forming a loop. Some graphs of this type are shown in Fig. 5.

We now sum up all the "chain" terms, appearing in (49), namely, those associated with the graphs, which consist of only one path of lines joining the white vertices 1 and 2. The graphs of this type consist of alternate h_2^B-bond and $\gamma^{(n)}$-bond: the $\gamma^{(n)}$-bond may be superimposed or not on an h_2^B-bond. These chain graphs may consist of

(a) h_2^B-bonds at the two extremities,

(b) One h_2^B-bond and one $\gamma^{(n)}$-bond superimposed on g_2^B-bond as extreme lines,

FIG. 5. Some graphs of the $X(1,2)$ type.

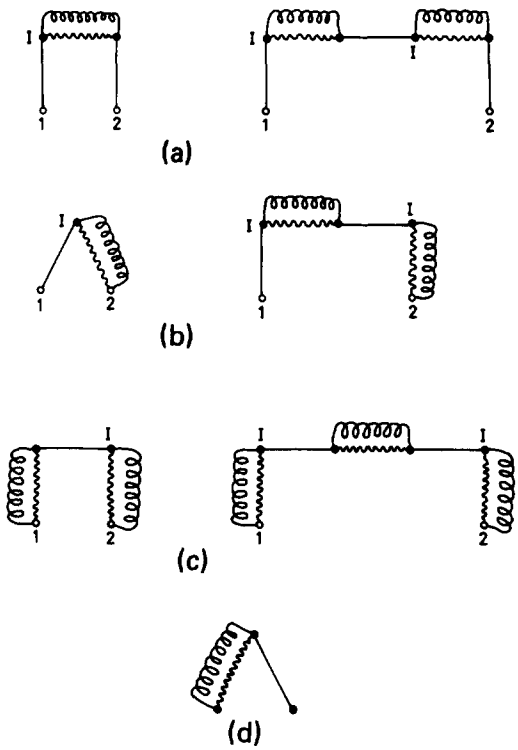


FIG. 6. (a) Chains with h_2^B-bonds at the two extremities; (b) chains with one h_2^B-bond and one $\gamma^{(n)}h_2^B$-bond ($\gamma^{(n)}$-bond superimposed on g_2^B-bond) at the two extremities; (c) chains with $\gamma^{(n)}g^{(B)}$-bonds at the two extremities; (d) propagator.

(c) $\gamma^{(n)}$-bonds superimposed on g_2^{B}-bonds (i.e., $\gamma^{(n)}g_2^{\mathrm{B}}$-bonds) at the two extremities.

These graphs can be constructed by repeatedly inserting the "propagator" consisting of graphs of Fig. 6(d) in place of the black points [labeled I in Figs. 6(a,b,c)]. The few graphs of these series are shown in Fig. 6(a,b,c).

The contributions of the terms corresponding to the chain graphs of these three types are

$$G_{hh}^{(n)}(1,2)=\frac{1}{(2\pi)^3\rho}\int d\bar{k}\left[\frac{\tilde{h}^2\tilde{\alpha}}{1-\tilde{h}\tilde{\alpha}}\right]e^{-i\bar{k}\cdot\bar{r}_{12}}, \tag{50}$$

$$G_{h\gamma}^{(n)}(1,2)=\frac{1}{(2\pi)^3\rho}\int d\bar{k}\left[\frac{\tilde{h}\tilde{\alpha}}{1-\tilde{h}\tilde{\alpha}}\right]e^{-i\bar{k}\cdot\bar{r}_{12}}, \tag{51}$$

$$G_{\gamma\gamma}^{(n)}(1,2)=\frac{1}{(2\pi)^3\rho}\int d\bar{k}\left[\frac{\tilde{h}\tilde{\alpha}^2}{1-\tilde{h}\tilde{\alpha}}\right]e^{-i\bar{k}\cdot\bar{r}_{12}}, \tag{52}$$

where

$$\tilde{h}=\rho\int d\bar{r}h_2^{\mathrm{B}}(r)e^{i\bar{k}\cdot\bar{r}}, \tag{53a}$$

$$\tilde{\alpha}=\rho\int d\bar{r}\gamma^{(n)}(r)[1+h_2^{\mathrm{B}}(r)]e^{i\bar{k}\cdot\bar{r}}. \tag{53b}$$

In the chain approximation, the RDF of the Fermi system is obtained from (49) as

$$g_2^{\mathrm{F}}(1,2)=g_2^{\mathrm{B}}(1,2)\{[1+\gamma^{(n)}(1,2)][1+G_{hh}^{(n)}(1,2)]+2G_{h\gamma}^{(n)}(1,2)+G_{\gamma\gamma}^{(n)}(1,2)\}. \tag{54}$$

Having $G_{hh}^{(n)}$, $G_{h\gamma}^{(n)}$, and $G_{\gamma\gamma}^{(n)}$, we can construct composite chain graphs by connecting the white vertices 1 and 2 by two or more single chains, as in Fig. 7. The contribution of

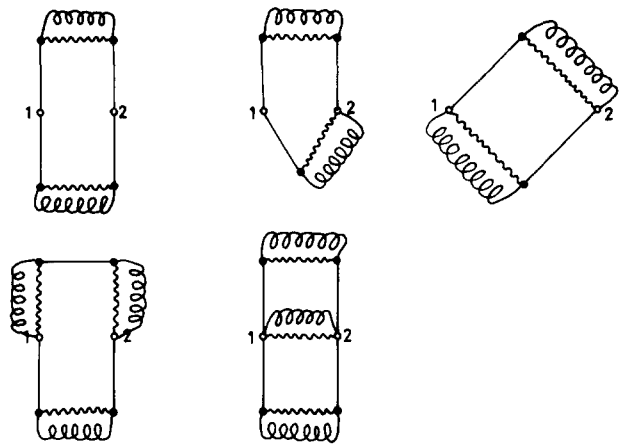


FIG. 7. Composite chain graphs.

any composite chain graphs is a product of contributions from single chain.

We now sum all graphs in which the white vertices 1 and 2 are connected by any number of chains of the types shown in Fig. 7. The sum of all these graphs is given by

$$L^{(n)}(1,2)=L_1^{(n)}(1,2)+L_2^{(n)}(1,2)+L_3^{(n)}(1,2)+L_4^{(n)}(1,2), \tag{55}$$

where

$$L_1^{(n)}(1,2)=\gamma^{(n)}(1,2)+[1+\gamma^{(n)}(1,2)]G_{hh}^{(n)}(1,2)+2G_{h\gamma}^{(n)}(1,2)+G_{\gamma\gamma}^{(n)}(1,2) \tag{56}$$

$$L_2^{(n)}(1,2)=[1+\gamma^{(n)}(1,2)]\{\exp[G_{hh}^{(n)}(1,2)]-G_{hh}^{(n)}(1,2)-1\}, \tag{57}$$

$$L_3^{(n)}(1,2)=[2G_{h\gamma}^{(n)}(1,2)+G_{\gamma\gamma}^{(n)}(1,2)]\times\{\exp[G_{hh}^{(n)}(1,2)]-1\}, \tag{58}$$

$$L_4^{(n)}(1,2)=G_{h\gamma}^{(n)2}(1,2)\exp[G_{hh}^{(n)}(1,2)]. \tag{59}$$

Thus, we find

$$L^{(n)}(1,2)=[1+\gamma^{(n)}(1,2)+G_{\gamma\gamma}^{(n)}(1,2)+2G_{h\gamma}^{(n)}(1,2)+G_{h\gamma}^{(n)2}(1,2)]\exp[G_{hh}^{(n)}(1,2)]-1. \tag{60}$$

We next construct chains from more complicated link $L^{(n)}(i,j)$. The new set of single chain graphs is obtained by replacing $\gamma^{(n)}$ by $L^{(n)}$ in Fig. 6. So the contributions of these single chain graphs can be given by (50)–(52) provided $\gamma^{(n)}$ is replaced by $L^{(n)}$ in (53b). The composite chain graphs are made from the new single chain graphs. Putting $G_{hh}^{(n)}$, $G_{h\gamma}^{(n)}$, and $G_{\gamma\gamma}^{(n)}$ thus obtained in (60) gives the link $L^{(n)}$ to be used in constructing still more complicated chains.

In this approximation, which we may term the "modi-

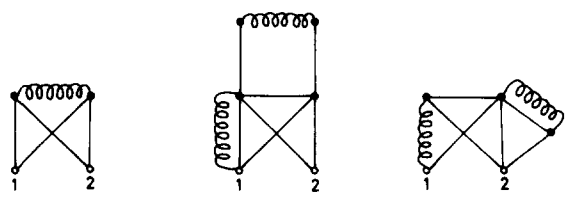


FIG. 8. Elementary graphs.

fied" FHNC approximation, the graphs with open structure such as chain graphs are included in the summation. The graphs with a very compact structure, such as those shown in Fig. 8, are left out from the summation. The other class of graphs, which have not been included here, are those belonging to the category $X(1,2)$ [see Eq. (48)]. It is, however, possible at least in principle, to include some of these graphs. In a future publication we plan to explore this and to evaluate the RDF for some fermion systems.

[1]C. E. Campbell, in *Progress in Liquid Physics*, edited by C. A. Croxton (Wiley, New York, 1978), Chap. 6.
[2]D. Ceperley, G. V. Chester, and M. H. Kalos, Phys. Rev. B **16**, 3081 (1977).
[3]E. Feenberg, *Theory of Quantum Fluids* (Academic, New York, 1969).
[4]M. D. Miller, Phys. Rev. B **14**, 3937 (1976).
[5]F. Iwamoto and M. Yamada, Progr. Theor. Phys. (Kyoto) **17**, 543 (1957); **18**, 345 (1957).
[6]F. Y. Wu and E. Feenberg, Phys. Rev. **128**, 943 (1962).
[7]J. B. Aviles, Ann. Phys. (N.Y.) **5**, 251 (1958).
[8]C. D. Harton and H. A. Tolhock, Physica **24**, 721, 875, 896 (1958).
[9]J. G. Zabolitzky, Phys. Lett. B **64**, 233 (1976); Phys. Rev. A **16**, 1258 (1977).
[10]B. H. Brandow, Phys. Lett. B **61**, 117 (1976).
[11]A. Bijl, Physica **I**, 869 (1940); R. Jastrow, Phys. Rev. **98**, 1479 (1955).
[12]G. Stell, in *The Equilibrium Theory of Classical Fluids*, edited by H. L. Frisch and J. L. Lebowitz (Benjamin, New York, 1964).
[13]T. Morita and K. Hiroike, Progr. Theor. Phys. (Jpn.) **25**, 537 (1961).
[14]E. Krotscheck and M. L. Ristig, Phys. Lett. A **48**, 17 (1974); Nucl. Phys. A **242**, 389 (1975).
[15]S. Fantoni and S. Rosati, Nuovo Cimento A **10**, 145 (1972); **20**, 179 (1974); **25**, 593 (1975); Lett. Nuovo Cimento **10**, 545 (1974).
[16]J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).

Derivation of the generalized Langevin equation by a method of recurrence relations

M. Howard Lee
Department of Physics, University of Georgia, Athens, Georgia 30602

(Received 15 March 1983; accepted for publication 17 June 1983)

The generalized Langevin equation was first derived by Mori using the Gram-Schmidt orthogonalization process. This equation can also be derived by a method of recurrence relations. For a physical space commonly used in statistical mechanics, the recurrence relations are simple and they lead directly to the Langevin equation. The Langevin equation is shown to be composed of one homogeneous and one inhomogeneous equation.

PACS numbers: 05.40. + j, 05.60. + w

I. INTRODUCTION

The generalized Langevin equation (GLE) is an exact equation of motion. It was first derived and formally solved by Mori using a projection operator technique[1] which turns out to be a formal version of the Gram–Schmidt orthogonalization process. Because the Gram–Schmidt process is a general method of orthogonalization, it is unwieldy when applied to some specific infinite-dimensional Hilbert spaces. Thus the work of Mori, although very remarkable an achievement, seems highly formal. It was recently shown[2] that since the GLE is equivalent to the Liouville equation (hence the Heisenberg equation of motion), one can obtain formal solutions for time evolution by solving the Liouville equation directly instead of the GLE.

The alternative way of obtaining the formal solutions was spurred by the recognition that the Hilbert space in which statistical problems are considered is a realization of abstract Hilbert space. For such a realized space there is a simplifying orthogonalization process via recurrence relations.[3] It was found that the recurrence relations represent formal solutions for time evolution in this space. Also, they contain other details not previously seen by the approach of the GLE.

Although the GLE and Liouville equation are equivalent, the former is still physically the more interesting. Because of its close and deep connection to the classical Langevin equation,[4] the GLE can provide considerable insight into dynamical problems. Hence it is advantageous sometimes to view a physical problem of interest through the GLE. For this reason it seems useful to derive the GLE via recurrence relations.

II. RECURRENCE RELATIONS

Let $A = A(t=0)$ be a dynamical variable in $\mathscr{S}$. For simplicity we assume A to be Hermitian. The time evolution of A is governed by the Liouville equation

$$dA(t)/dt = iLA(t), \tag{1}$$

where $LA = [H,A]$ and H is the Hamiltonian of the system. We further assume that the space $\mathscr{S}$ is a realization of abstract Hilbert space, defined by the inner product,

$$(X,Y) = \int_0^\beta d\lambda \, \langle X(\lambda) Y^\dagger \rangle - \langle X \rangle \langle Y^\dagger \rangle, \tag{2}$$

where $X, Y \subset \mathscr{S}$, β is the inverse temperature $\beta = 1/kT$, $X(\lambda) = \exp(\lambda H) X \exp(-\lambda H)$, and $\langle XY \rangle$ is an ensemble average defined by $\langle XY \rangle = \mathrm{Tr}(e^{-\beta H} XY)/\mathrm{Tr}(e^{-\beta H})$. It was shown that the solution for $A(t)$ in $\mathscr{S}$ may then be given as

$$A(t) = \sum_{\nu=0} a_\nu(t) f_\nu, \tag{3}$$

where $\{a_\nu(t)\}$ are a set of time-dependent real functions.[2] Also $\{f_\nu\}$ are a set of *orthogonalized* basis vectors which span $\mathscr{S}$ and satisfy the recurrence relation (RR),

$$f_{\nu+1} = \dot{f}_\nu + \Delta_\nu f_{\nu-1}, \quad \nu \geqslant 0, \tag{4}$$

where $\dot{f}_\nu = iLf_\nu$, $\Delta_\nu = (f_\nu, f_\nu)/(f_{\nu-1}, f_{\nu-1})$, $f_{-1} \equiv 0$, and $\Delta_0 \equiv 1$.[3] The above three-term RR will be referred to as the RRI. It has one arbitrary initial choice among $\{f_\nu\}$. Given this choice, the rest is allowed no more freedom by the RRI. If we choose $f_0 = A$, the boundary condition requires that $a_0(0) = 1$ and $a_\nu(0) = 0$, $\nu \geqslant 1$. Then $a_0(t)$ represents the relaxation function of linear response theory.[1] Now since (3) must satisfy (1), (4) yields a recurrence relation for $\{a_\nu(t)\}$:

$$\Delta_{\nu+1} a_{\nu+1}(t) = -\dot{a}_\nu(t) + a_{\nu-1}(t), \quad \nu \geqslant 0, \tag{5}$$

where $\dot{a}_\nu(t) = da_\nu(t)/dt$ and $a_{-1}(t) \equiv 0$. The above RR will be referred to as the RRII. It will be seen that the RRII leads directly to the GLE.

III. THE GENERALIZED LANGEVIN EQUATION

We apply the Laplace Transform $\mathscr{T}$ on the RRII to obtain

$$1 = z a_0(z) + \Delta_1 a_1(z), \tag{6a}$$

$$a_{\nu-1}(z) = z a_\nu(z) + \Delta_{\nu+1} a_{\nu+1}(z), \quad \nu \geqslant 1, \tag{6b}$$

where $a_\nu(z) = \mathscr{T}[a_\nu(t)]$. To derive the GLE from the above two equations, we introduce two quantities ϕ and b_ν, defined entirely in terms of a_ν:

$$\phi(z) = \Delta_1 a_1(z)/a_0(z), \tag{7}$$

$$b_\nu(z) = a_\nu(z)/a_0(z), \quad \nu \geqslant 1. \tag{8}$$

Then, from (6a) and (7),

$$1/a_0(z) = z + \phi(z), \tag{9}$$

and from (8) and (9),

$$b_\nu(z) = (z + \phi(z)) a_\nu(z). \tag{10}$$

By applying the inverse Laplace Transform $\mathscr{T}^{-1}$ on (9), we

 0022-2488/83/102512-03$02.50

obtain, recalling $a_0(t=0)=1$,

$$\dot{a}_0(t)+\int_0^t dt'\,\phi(t-t')a_0(t')=0, \tag{11}$$

where $\phi(t)=\mathcal{T}^{-1}[\phi(z)]$. Hence similarly, we have, recalling $a_\nu(t=0)=0$ for $\nu\geqslant 1$,

$$\dot{a}_\nu(t)+\int_0^t dt'\,\phi(t-t')a_\nu(t')=b_\nu(t), \tag{12}$$

where $b_\nu(t)=\mathcal{T}^{-1}[b_\nu(z)]$. We now multiply (11) by f_0 and (12) by f_ν, $\nu\geqslant 1$, and then combine the two equations. With the definition (3), we obtain

$$\dot{A}(t)+\int_0^t dt'\,\phi(t-t')A(t')=f[t], \tag{13}$$

where we define

$$f[t]=f_1[t]=\sum_{\nu=1} b_\nu(t)f_\nu. \tag{14}$$

If $f[t]$ indeed is the random force, the above expression (13) is precisely the GLE.[1] It is interesting to note that the GLE consists of two types of equations, one homogeneous and the other inhomogeneous, reflecting (6a) and (6b). In analogy to the role of $a_0(t)$ with respect to $A(t)$, the memory function ϕ may be termed the relaxation function of the random force.

Also, from (8) we obtain

$$a_\nu(t)=\int_0^t dt'\,a_0(t-t')b_\nu(t'),\quad \nu\geqslant 1. \tag{15}$$

Hence it follows that

$$A(t)=a_0(t)A+\int_0^t dt'\,a_0(t-t')f[t'], \tag{16}$$

where $A=f_0$. The second term of (16), often called the nonsecular part, is orthogonal to the first term, the secular part, for $t>0$ owing to (14).[5]

IV. THE RANDOM FORCE

The necessary condition for $f[t]$ being the random force is that $(f[t],A)=0$ for $t\geqslant 0$, where $A=f_0$. The sufficient condition is that $f[t]$ lies in the largest linear manifold of $\mathcal{S}$. Let $\mathcal{S}_\nu$ be the νth linear manifold of $\mathcal{S}\equiv\mathcal{S}_0$, spanned by $\{f_\nu,f_{\nu+1},\ldots\}$. Also we define $f_\nu[t]=\Sigma_{\nu'=\nu}a_{\nu'}^{(\nu)}(t)f_{\nu'}$ so that $f_\nu[t]\subset\mathcal{S}_\nu$. For example,

$$f_0[t]=\sum_{\nu=0} a_\nu(t)f_\nu=A(t),$$

$$f_1[t]=\sum_{\nu=1} b_\nu(t)f_\nu=f[t],$$

$$f_2[t]=\sum_{\nu=2} c_\nu(t)f_\nu,$$

etc., where we have put $a_\nu^{(0)}=a_\nu$, $\nu\geqslant 0$; $a_\nu^{(1)}=b_\nu$, $\nu\geqslant 1$; $a_\nu^{(2)}=c_\nu$, $\nu\geqslant 2$; etc. Then, (i) $\mathcal{S}_0\supset\mathcal{S}_1\supset\mathcal{S}_2\cdots$ but $\mathcal{S}_0\not\subset\mathcal{S}_1\not\subset\mathcal{S}_2\cdots$, (ii) $f_\nu\perp\mathcal{S}_\mu$ if $\nu<\mu$, (iii) $f_\nu[t]\subset\mathcal{S}_\mu$ if and only if $\nu\geqslant\mu$. Now, considering $f_1[t]$, we have $(f_1[t],f_0)=0$ and $f_0\perp\mathcal{S}_1$. Also $f_1[t]\not\subset\mathcal{S}_\nu$ if $\nu>1$, but $f_1[t]\subset\mathcal{S}_1$, which is the largest linear manifold of $\mathcal{S}$, i.e., $\mathcal{S}_1\perp f_0$ only. By extending this argument, it is easy to generalize that $f_\nu[t]$ is the random force for $f_{\nu-1}[t]$, $\nu\geqslant 1$.

One can arrive at the same conclusion somewhat less formally by the following route: From (8), we have

$$b_1(z)=a_1(z)/a_0(z), \tag{17a}$$

$$b_\nu(z)/b_{\nu-1}(z)=a_\nu(z)/a_{\nu-1}(z),\quad \nu\geqslant 2. \tag{17b}$$

Hence with (6b),

$$b_1(z)=[z+\Delta_2 b_2(z)/b_1(z)]^{-1}, \tag{18a}$$

$$b_\nu(z)/b_{\nu-1}(z)=[z+\Delta_{\nu+1}b_{\nu+1}(z)/b_\nu(z)]^{-1},\quad \nu\geqslant 2. \tag{18b}$$

Combining the two, we obtain

$$b_1(z)=1/z+\Delta_2/z+\Delta_3/z+\cdots \quad \text{(continued fraction)}. \tag{19}$$

Thus

$$b_1=a_0(\{\Delta_\nu\}\to\{\Delta_{\nu+1}\}),\quad \nu\geqslant 1. \tag{20}$$

We can generalize it as

$$b_\nu=a_{\nu-1}((\Delta_\nu,\Delta_{\nu+1},\cdots)\to(\Delta_{\nu+1},\Delta_{\nu+2},\ldots)),\quad \nu\geqslant 1. \tag{21}$$

Now, since $\{\Delta_\nu\}$ represents a set of lengths of the vectors spanning a linear manifold of $\mathcal{S}$, $f_1[t]$ is in a linear manifold spanned by all but f_0. Hence it is orthogonal to f_0 and belongs to the largest linear manifold of $\mathcal{S}$.

Finally, one can obtain the GLE for a dynamical variable in any linear manifold of $\mathcal{S}$. Since $f_\nu[t]\subset\mathcal{S}_\nu$, $f_{\nu+1}[t]\subset\mathcal{S}_{\nu+1}$, and $f_\nu\perp\mathcal{S}_{\nu+1}$, we have, for $\nu\geqslant 0$,

$$\dot{f}_\nu[t]+\int_0^t dt'\,\phi_{\nu+1}(t-t')f_\nu[t']=f_{\nu+1}[t], \tag{22}$$

where

$$\phi_\nu(t)=(f_\nu,f_\nu[t])/(f_{\nu-1},f_{\nu-1}),\quad \nu\geqslant 1, \tag{23}$$

with $\phi_1=\phi$. The solution for $f_\nu[t]$ is obtained by generalizing (16):

$$f_\nu[t]=a_\nu^{(\nu)}(t)f_\nu+\int_0^t dt'\,a_\nu^{(\nu)}(t-t')f_{\nu+1}[t'], \tag{24}$$

where $a_\nu^{(\nu)}(t)=\mathcal{T}^{-1}[a_\nu^{(\nu)}(z)]$, $\nu'\geqslant\nu$, and

$a_\nu^{(0)}(z)=a_\nu(z)$, $\nu\geqslant 0$; $a_\nu^{(1)}(z)=b_\nu(z)=a_\nu(z)/a_0(z)$, $\nu\geqslant 1$; $a_\nu^{(2)}(z)=c_\nu(z)=b_\nu(z)/b_1(a)=a_\nu(z)/a_1(z)$, $\nu\geqslant 2$; etc.

V. DISCUSSION

We have seen that the RRII leads directly to the GLE. The two, however, are not entirely equivalent. The RRII gives a relationship between the components of the dynamical variable and its random force. Hence this relationship is "microscopic." The GLE is a gross or total relationship between the two physical quantities. The microscopic detail contained in the RRII, in fact, shows that the GLE is composed of two types of equations. It also gives a precise meaning to the random force.

The RRII represents a realization of an abstract Hilbert space. Hence it contains information about a particular space in which a given physical problem is considered. For example, its structure denotes what functions are admissible or inadmissible as solutions for $a_\nu(t)$. In addition, given $a_0(t)$, the RRII yields all other components of $A(t)$. [See the Appendix for an illustration.]

The inequivalence between the RRII and the GLE stems from the very nature of orthogonalization. The GLE is

obtained by the Gram–Schmidt process. Hence it is general and not specific. The approach via the RR is specific and not general. The advantage of one approach over the other ultimately rests on whether one's space is realized or remains abstract.

ACKNOWLEDGMENTS

I wish to thank Professor Hans De Raedt of the Universitaire Instelling Antwerpen, Belgium for drawing my attention to Ref. 5. This work was supported in part by the U. S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AS09-77ER01023.

APPENDIX A: FORCE CORRELATION FUNCTION

We consider the time integral of the force correlation function[6]

$$I=\int_0^\infty dt\,(\dot{A},\dot{A}\,(t))/(\dot{A},\dot{A}\,), \tag{A1}$$

where $\dot{A}=\dot{A}\,(t=0)$. From (3) and (5), we have $\dot{A}=f_1$ since $a_\nu(0)=0$ if $\nu\geqslant 1$ and $\dot{a}_1(0)=a_0(0)=1$. Hence

$$I=\int_0^\infty dt\,(f_1,f_1(t))/(f_1,f_1)=\int_0^\infty dt\,\dot{a}_1(t)$$
$$=a_1(t=\infty)=-\dot{a}_0(t=\infty)/\Delta_1. \tag{A2}$$

The last step follows from the RRII. Thus if the slope of $a_0(t)$ vanishes at $t=\infty$, $I=0$.

We consider a few examples: (1) If $a_0(t)=\exp(-\chi t^2)$, where χ is some positive constant, then $I=0$. This is realized in the spin van der Waals model[7] with $A=S_x$, where S_x is the x-component of the total spin operator. (2) If $a_0(t)=J_0(\mu t)$, where μ is a constant, $\dot{J}_0(t)=-J_1(t)$ goes to zero as $t\to\infty$ (although very slowly) and $I=0$. This is realized in the ideal electron gas in two dimensions at $T=0$.[8] Here J_0 and J_1 are the Bessel functions of orders 0 and 1, respectively. (3) If $a_0(t)=\cos\omega_0 t$, $I\neq 0$. This is realized in the high-frequency limit of a variety of cooperative models including the $3d\,XY$ model[9] and also in noncooperative models such as single-spin models.[10]

We next consider the time integral of the random current

$$J=\int_0^\infty dt\,(f_1,f_1[t\,])/(f_1,f_1). \tag{A3}$$

From the definition of the random force (see Sec. IV) and the orthogonality of $\{f_\nu\}$, we have

$$J=\int_0^\infty dt\,b_1(t). \tag{A4}$$

Also using (13), the above can be expressed as

$$J=I+\int_0^\infty dt\int_0^t dt'\,\phi\,(t-t')a_1(t'). \tag{A5}$$

Clearly the two quantities J and I are very different. Even if $I=0$, J need not vanish. The expression (A3) may be recognized as the zero-frequency limit of the Kubo conductivity.[11]

[1] H. Mori, Prog. Theor. Phys. **33**, 423 (1965); **34**, 399 (1965). Also see S. Nordholm and R. Zwanzig, J. Stat. Phys. **13**, 347 (1975).
[2] M. H. Lee, Phys. Rev. B **26**, 2547 (1982).
[3] M. H. Lee, Phys. Rev. Lett. **49**, 1072 (1982).
[4] R. F. Fox, Phys. Rep. **48**, 180 (1978).
[5] The RRI was apparently first derived by M. Dupuis [Prog. Theor. Phys. **37**, 502 (1967)]. Dupuis, however, *presupposed* that $a_0(z)$ is a continued fraction, known from Mori (Ref. 1). Then utilizing the connection between continued fractions and series expansions established by the theory of the problem of moments, he obtained a recurrence relation for $\{f_\nu\}$, which becomes the RRI upon realizing an abstract Hilbert space. This derivation is circuitous. It depends on the assumption that $a_0(z)$ is a continued fraction. In Ref. 3, we have shown that the RRI is a simple property of an abstract Hilbert space realized by (2). That $a_0(z)$ is a continued fraction, i.e., the RRII, is a consequence, not the cause, of the RRI.
[6] R. Kubo, Rep. Prog. Phys. **27**, 255 (1966). See esp. Eq. 8.10.
[7] R. Dekeyser and M. H. Lee, Phys. Rev. B **19**, 265 (1979). Also see Ref. 2.
[8] M. H. Lee and J. Hong, Phys. Rev. Lett. **48**, 634 (1982).
[9] M. H. Lee, Phys. Rev. B **8**, 3290 (1973).
[10] M. H. Lee, Can. J. Phys. **61**, 428 (1983).
[11] See. Eq. 11.41 of Ref. 6. Also see Ref. 8.

A group of coordinate transformations which preserve the metric of Weinhold

Peter Salamon
Department of Mathematical Sciences, San Diego State University, San Diego, California 92182

Ed Ihrig
Department of Mathematics, Arizona State University, Tempe, Arizona 85281

R. Stephen Berry
Department of Chemistry, University of Chicago, Chicago, Illinois 60637

(Received 10 February 1982; accepted for publication 22 April 1983)

Recently Weinhold described a natural metric on the state space of an equilibrium thermodynamic system. We describe the coordinate transformations which preserve both the first law of thermodynamics and this metric.

PACS numbers: 05.70. − a, 02.20. + b

I. INTRODUCTION

We consider an n degree of freedom thermodynamic system whose extensive variables are $x_1,...,x_n$, and conjugate intensities are $\partial E/\partial x_i = y_i$, where E is the internal energy. Weinhold[1] pointed out that the second derivative matrix

$$\left[\frac{\partial^2 E}{\partial x_i\,\partial x_j}\right] = [\eta_{ij}] = \eta, \tag{1}$$

being symmetric and positive definite, may be used to define a metric structure on the set of equilibrium states of a thermodynamic system. Distances measured by η have been interpreted as changes of velocities characteristic of the type of path.[2,3] He examined the group of coordinate transformations in the state space of $X = (x_1,...,x_n)$ which preserved η, and found that this group was isomorphic to Gl(n), the group of all invertible linear transformations. Hermann[4] pointed out that a fuller view of the mathematical structure of equilibrium thermodynamics may be found in *phase-energy space* $(X,Y,E) = (x_1,...,x_n, y_1,...,y_n, E)$. Below, we solve the problem of finding the group of coordinate transformations in phase-energy space which preserve η.

Following Hermann, we identify the state space of an n-degree of freedom thermodynamic system with a surface of maximal dimension in phase-energy space which is a solution of the Pfaffian equation

$$\omega = dE - \sum_{i=1}^{n} y_i\,dx_i = 0 \tag{2}$$

expressing the first law of thermodynamics, where ω is the differential form defined by (2). The general theory[3] states that this will be an n-dimensional surface, which may be coordinatized by $x_1,...,x_n$. When E is restricted to such a surface, it becomes a function of $x_1,...,x_n$ alone. The most natural condition is the invariance of the first law. We will call coordinates (U,V,P) admissible provided

$$dE - \sum_{i=1}^{n} y_i\,dx_i = 0 \quad \text{iff} \quad dP - \sum_{i=1}^{n} v_i\,du_i = 0. \tag{3}$$

This condition is equivalent to

$$dE - \sum_{i=1}^{n} y_i\,dx_i = \alpha\left(dP - \sum_{i=1}^{n} v_i\,du_i\right) \tag{4}$$

for some function α. An obvious stronger requirement is (4) with $\alpha = 1$. If this is satisfied, we say that the form of the first law is preserved. Such transformations are called contact transformations,[4] and they form an infinite-dimensional group.

We will also require the invariance of the Weinhold metric

$$\eta = \sum_{i=1}^{n} dy_i\,dx_i, \tag{5}$$

where we used the same symbol η for this differential 2-form, since when η is restricted to a maximal surface, its matrix relative to the coordinates $(x_1,...,x_n)$ is given by (1). We can see this as follows. In such a maximal surface $dE = \Sigma_{i=1}^{n} \times(\partial E/\partial x_i)dx_i$. Since $\omega = 0$ on the surface we have $dE = \Sigma_{i=1}^{n} y_i\,dx_i$. Since dx_i are independent on a maximal surface $y_i = \partial E/\partial x_i$. Thus

$$\eta = \sum_{i=1}^{n} dy_i\,dx_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{\partial^2 E}{\partial x_i\,\partial x_j}dx_j\right)dx_i, \tag{6}$$

which gives the result. There are two important things to note here. The first is that $\partial^2 E/\partial x_i\,\partial x_j$ is only defined on a maximal surface and not on phase-energy space (E and x_i are independent coordinates on phase-energy space and thus $\partial^2 E/\partial x_i\,\partial x_j = 0$ in this setting). Since there are many maximal surfaces going through one point, it is not even clear that the $\partial^2 E/\partial x_i\,\partial x_j$ defined using one surface will be consistent with the metric defined using another surface. However, all these metrics are consistent because they agree with η which is defined on all of phase-energy space. We assume that η in (5) is the fundamental object, but caution the reader that η is *not* the only metric whose restriction gives $\partial^2 E/\partial x_i\,\partial x_j$ on each maximal surface. A quadratic form η_1 will have this restriction on each maximal surface if and only if it is of the form

$$\eta_1 = \eta + \omega\theta, \tag{7}$$

where θ is an arbitrary 1-form. Thus there are $(2n + 1)$ free functions in the general metric that extends $\partial^2 E/\partial x_i\,\partial x_j$ to phase-energy space. Among these metrics, the choice with $\theta = 0$ seems the most natural. The only thing that is clear about the problem with general θ is that its solution is way

 0022-2488/83/102515-06$02.50

beyond the techniques presented in this paper. Note that since θ has been fixed, the requirement of the invariance of η is a stronger requirement than the invariance of $\partial^2 E/\partial x_i\,\partial x_j$ alone.

The invariance of η amounts to the condition

$$\sum_{i=1}^{n} dy_i\,dx_i = \sum_{i=1}^{n} dv_i\,du_i. \tag{8}$$

We find two types of transformations that leave ω and η invariant. The first consists of transformations that are analogous to translations. They are of the form

$$F(X,Y,E) = (X + \mathbf{b}, Y + \mathbf{d}, E + \mathbf{d}\cdot X + \mathbf{e}), \tag{9}$$

where $\mathbf{b}, \mathbf{d} \in R^n$ and $\mathbf{e} \in R$. These pseudotranslations do not commute like real Euclidean translations, but they do form a group isomorphic to the n-dimensional Heisenberg group (H_n), a group[6,7] which is very close in structure to R^n. The second type of transformations are analogous to rotations. They fix the origin and are linear. Their matrices are of the form

$$\begin{pmatrix} A & 0 & 0 \\ 0 & (A^t)^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{10}$$

where A is any invertible $n \times n$ matrix. Thus these transformations form a group that is isomorphic to Gl(n). Just as the Euclidean group is the semidirect product of the rotations with the translations, we find the group of all transformations that fix ω and η is isomorphic to a semidirect product of Gl(n) with H_n. Thus the structure of this group is similar to the structure of the transformation group on R^n that leaves the standard Euclidean metric invariant.

In Sec. III we described those transformations which multiply ω and η by some constant factor. This allows more transformations. It is natural to allow the multiplication of ω by an arbitrary function as we described in the beginning of this chapter. However, a long and tedious computation using the ideas of Sec. II may be used to show that if a transformation fixes η and ω by a factor α, then α must be a constant. The suitability of letting η change by a factor is not so obvious. However, if the factor is constant, it may be interpreted as being only a change in energy scale. Thus these transformations are also given in Sec. III.

II. ABSOLUTE INVARIANCE

Letting X,Y,U, and V be n-component real vectors, we wish to find the group G of all coordinate transformations $f: R^{2n+1} \to R^{2n+1}$ such that

$$f(X,Y,E) = (U,V,P),$$

while

$$dE - \sum_{i=1}^{n} y_i\,dx_i = dP - \sum_{i=1}^{n} v_i\,du_i, \tag{11a}$$

$$\sum_{i=1}^{n} dy_i\,dx_i = \sum_{i=1}^{n} dv_i\,du_i, \tag{11b}$$

and [as a consequence of (11a)],

$$\sum_{i=1}^{n} dx_i \wedge dy_i = \sum_{i=1}^{n} du_i \wedge dv_i. \tag{11c}$$

It proves convenient to separate the problem into two parts by examining the subgroup $J = \{f \in G: f(0) = 0\}$ of transformations which fix the origin and

$$H = \{f \in G: f(X,Y,E) = (X + \mathbf{b},\ Y + \mathbf{d},\ E + \mathbf{d}\cdot X + \mathbf{e}),\ \mathbf{b}, \mathbf{d} \in R^n, \mathbf{e} \in R\} \tag{12}$$

of "translations." Elements of H involve arbitrary translations in the X and Y variables, but the invariance of (11a) requires the additional term $\mathbf{d}\cdot X$ in P. The fact that J and H are subgroups of G is easily verified.

Lemma 1: An arbitrary $f \in G$ may be written as the product of $j \in J$ and $h \in H$.

Proof: Consider $f \in G$ and suppose $f(0) = (\mathbf{b}_0, \mathbf{d}_0, \mathbf{e}_0)$. Then choosing h such that $h(X,Y,E) = (X - \mathbf{b}_0,\ Y - \mathbf{d}_0,\ E - \mathbf{d}_0\cdot X - \mathbf{e}_0 + \mathbf{b}_0\mathbf{d}_0)$, we find $h \circ f$ fixes the origin and hence equals j for some $j \in J$. But then $f = h^{-1} \circ j$, where

$$h^{-1}(X,Y,E) = (X + \mathbf{b}_0,\ Y + \mathbf{d}_0,\ E + \mathbf{d}_0\cdot X + \mathbf{e}_0) \in H. \tag{13}$$

Note further that the only element of H which fixes the origin is the identity. We see below that H is normal. These facts are enough to guarantee that G is the semidirect product of J and H.

Theorem 1: H is isomorphic to the $(2n+1)$-dimensional Heisenberg group[6,7] H(n).

Proof: Clear by the correspondence

$$H^{\text{def}} = T_{(\mathbf{b},\mathbf{d},\mathbf{e})} \sim \begin{pmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot \\ b_1 & 1 & 0 & \cdot & \cdot & \cdot \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_n & 0 & \cdot & \cdot & 1 & 0 \\ e & d_1 & \cdot & \cdot & d_n & 1 \end{pmatrix}. \tag{14}$$

Theorem 2: The subgroup J of elements of G which fix the origin is isomorphic to

$$O(n,n) \cap Sp(n) \cong Gl(n).$$

The key here is to note that by (11a) the $(2n+1)$th component P of $j(X,Y,E)$ is determined once the $2n$ components U and V are specified, while in terms of these $2n$ components the restriction of j must preserve a nondegenerate quadratic form $\tilde{\eta}$ of type (n,n) and a symplectic form $\widetilde{d\omega}$. Hence j must be in both O(n,n) and Sp(n). For a proof of the second isomorphism indicated in the theorem as

$$O(n,n) \cap Sp(n) \cong Gl(n). \tag{15}$$

see Helgason.[8]

Proof of Theorem 2: Define $\pi: R^{2n+1} \to R^{2n}$ and $I_{\mathbf{e}}: R^{2n} \to R^{2n+1}$ for each $\mathbf{e} \in R$ by

$$\pi(X,Y,E) = (X,Y) \tag{16}$$

and

$$I_{\mathbf{e}}(X,Y) = (X,Y,\mathbf{e}). \tag{17}$$

Then $\tilde{\eta} = \pi^*(\eta)$ and $\widetilde{d\omega} = \pi^*(d\omega)$ define nondegenerate quadratic and symplectic forms on R^{2n}. For $j \in J$ and $\mathbf{e} \in T$, define

$$j_{\mathbf{e}} = \pi \circ j \circ I_{\mathbf{e}} \tag{18}$$

which maps $R^{2n} \to R^{2n}$ and leaves $\tilde{\eta}$ and $\widetilde{d\omega}$ invariant. Note that $j_{\mathbf{e}}$ is the action of j on the X and Y variables with E held constant equal to $\mathbf{e}$. Since $j_{\mathbf{e}}$ preserves $\tilde{\eta}$ and fixes the origin,

it is in $O(n,n)$ and is in particular linear.[9] Since j_e also preserves $\widetilde{d\omega}$, it must also be in $Sp(n)$. Thus

$$j_e(X,Y) = (AX,(A^t)^{-1}Y) \tag{19}$$

for some $A \in Gl(n)$, where $A = A(e)$ may depend on e. Then

$$j(X,Y,E) = (AX,(A^t)^{-1}Y,P(X,Y,E)), \tag{20}$$

i.e., J fixes hyperplanes of constant E. To see that A is independent of E we let $A = [a_{ij}]$ and $A^{-1} = [b_{ij}]$ and use the invariance of η to get

$$\begin{aligned}\sum_{i=1}^{n} dx_i\, dy_i &= \sum_{i=1}^{n} d\left(\sum_{j=1}^{n} a_{ij}x_j\right) d\left(\sum_{k=1}^{n} b_{ki}y_k\right)\\ &= \sum_{i=1}^{n} dx_i\, dy_i + \sum_{i,j,k=1}^{n} x_j b_{ki}\frac{da_{ij}}{dE}dE\, dy_k\\ &\quad + \sum_{i,j,k=1}^{n} y_k a_{ij}\frac{db_{ki}}{dE}\, dE\, dx_j\\ &\quad + \sum_{i,j,k=1}^{n} x_j y_k \frac{da_{ij}}{dE}\frac{db_{ki}}{dE}dE\, dE\end{aligned} \tag{21}$$

which, by the linear dependence of the differential forms in the expansion, implies the equality of corresponding coefficients. In particular,

$$\sum_{i,j=1}^{n} x_j b_{ki}\frac{da_{ij}}{dE} = 0, \quad k = 1,\ldots,n \tag{22}$$

for arbitrary $(x_1,\ldots,x_n)$. Thus

$$\sum_{i=1}^{n} b_{ki}\frac{da_{ij}}{dE} = 0, \quad k,j = 1,\ldots,n \tag{23}$$

and, since $[b_{ki}]$ is invertible,

$$\frac{da_{ij}}{dE} = 0, \quad i,j = 1,\ldots,n. \tag{24}$$

Substituting (24) into (11a) gives

$$dE = dP \tag{25}$$

and, since $j(0) = 0$,

$$E = P. \tag{26}$$

The correspondence between

$$J^{\text{def}} = L_A = \begin{pmatrix} A & 0 & 0\\ 0 & (A^t)^{-1} & 0\\ 0 & 0 & 1\end{pmatrix} \in J \tag{27}$$

and

$A \in Gl(n) \cong O(n,n) \cap Sp(n)$ is now clear.

Lemma 2: H is normal in G.

Proof: For $j = L_A \in J$ and $h = T_{(\mathbf{b},\mathbf{d},\mathbf{e})} \in H$, we compute

$$\begin{aligned}j\circ h\circ j^{-1}(X,Y,E) &= j\circ h\,(A^{-1}X, A^t Y, E)\\ &= j(A^{-1}X + \mathbf{b}, A^t Y + \mathbf{d}, E\\ &\quad + \mathbf{d}\cdot A^{-1}X + \mathbf{e})\\ &= (X + A\mathbf{b}, Y + (A^t)^{-1}\mathbf{d},\\ &\quad E + (A^t)^{-1}\mathbf{d}\cdot X + \mathbf{e})\\ &= h'(X,Y,E),\end{aligned} \tag{28}$$

where $h' = T_{(A\mathbf{b},(A^t)^{-1}\mathbf{d},\mathbf{e})} \in H$. This can be written as

$$L_A T_{(\mathbf{b},\mathbf{d},\mathbf{e})} L_A^{-1} = T_{L_A(\mathbf{b},\mathbf{d},\mathbf{e})}. \tag{29}$$

The above results prove the following theorem.

Theorem 3: $G \cong J \otimes_\alpha H \cong Gl(n) \otimes_\alpha H(n)$.

III. RELATIVE INVARIANCE

We now allow η and ω to change by constant scale factors. This extra degreee of freedom adds two generators. Without this freedom, η fixes our unit of energy, though diagonal matrices in $Gl(n)$ allow scaling of the other quantities for changes of units. Changes of energy units give one generator, while a complete Legendre transform on all variables gives the other generator. Accordingly, we seek the group $\hat{G}$ of all coordinate transformations $f(X,Y,E) = (U,V,P)$ such that

$$dE - \sum_{i=1}^{n} y_i\, dx_i = \left(dP - \sum_{i=1} v_i\, du_i\right)\alpha \tag{30a}$$

and

$$\sum_{i=1}^{n} dx_i\, dy_i = \left(\sum_{i=1}^{n} du_i\, dv_i\right)\beta \tag{30b}$$

for some nonzero $\alpha,\beta \in R$. Again, by taking exterior derivatives in (30a), we get

$$\sum_{i=1}^{n} dx_i \wedge dy_i = \left(\sum_{i=1}^{n} du_i \wedge dv_i\right)\alpha. \tag{30c}$$

Note that $G \subseteq \hat{G}$ consists of those elements of $\hat{G}$ with $\alpha = \beta = 1$. Note that the correspondence $f^s \to (\alpha,\beta)$, sending $f \in \hat{G}$ to the pair of scale factors of f, defines a homomorphism of $\hat{G}$ to $R^* \otimes R^*$ with kernel G showing that G is a normal subgroup of $\hat{G}$.

Lemma 3: For all $f \in \hat{G}$, $|\alpha| = |\beta|$.

Proof: For $f \in \hat{G}$, again define

$$f_e = \pi f I_e \tag{31}$$

and note that

$$f_e^*(\tilde{\eta}) = (1/\beta)\tilde{\eta} \tag{32}$$

and

$$f_e^*(\widetilde{d\omega}) = (1/\alpha)\widetilde{d\omega}. \tag{33}$$

Recalling that a quadratic form Q transforms under f_e to $f_e^*(Q) = Df_e\, Q\, (Df_e)^t$, where Df_e is the Jacobian matrix of f_e, we get, by looking at (32) and (33) in any coordinates and taking determinants,

$$(\det Df_e)^2(\det\tilde{\eta}) = \det((1/\beta)\tilde{\eta}) = (1/\beta^{2n})(\det\tilde{\eta}), \tag{34}$$

$$(\det Df_e)^2(\det\widetilde{d\omega}) = \det((1/\alpha)\widetilde{d\omega}) = (1/\alpha^{2n})(\det\widetilde{d\omega}). \tag{35}$$

Since $\det(\tilde{\eta})$ and $\det(\widetilde{d\omega})$ are nonzero, this gives

$$\alpha^{2n} = 1/(\det Df_e)^2 = \beta^{2n}, \tag{36}$$

i.e., $|\alpha| = |\beta|$ as desired.

We now find the generators of $\hat{G}$ which are not in G. For $a \in R^*$, a nonzero real number, let

$$U_a(X,Y,E) = (X,aY,aE). \tag{37}$$

Note that $U_a \in \hat{G}$ and $(\alpha)U_a = \beta(U_a) = 1/a$. Let U be the one-parameter subgroup of $\hat{G}$ generated by the U_a, $a \in R^*$. Since the only element in U with $\alpha = \beta = 1$ is the identity,

$U \cap G$ contains only the identity element. U and G generate a subgroup

$$M = \{f \in \hat{G} : \alpha(f) = \beta(f)\}. \tag{38}$$

U and G are clearly contained in M. On the other hand, for $f_0 \in M$ with $\alpha(f_0) = \beta(f_0) = \alpha_0$, $U_{\alpha_0} \cdot f_0 \in G$ since

$$\alpha(U_{\alpha_0} \cdot f_0) = \alpha(U_{\alpha_0})\alpha(f_0) = 1, \tag{39a}$$

$$\beta(U_{\alpha_0} \cdot f_0) = \beta(U_{\alpha_0})\beta(f_0) = 1. \tag{39b}$$

The commutation relations

$$U_a L_A = L_A U_a, \tag{40}$$

$$U_a T_{(\mathbf{b},\mathbf{d},\mathbf{e})} = T_{(\mathbf{b},a\mathbf{d},a\mathbf{e})} U_a \tag{41}$$

follow easily by applying each side to (X,Y,E). We have established that

$$M \cong U \otimes_\beta G \cong R^* \otimes_\beta (\mathrm{Gl}(n) \otimes_\alpha \mathrm{H}(n)). \tag{42}$$

The commutation relations (29), (40), and (41) show that U_a and L_A commute, while the translations are normal in M. Thus we may also write

$$M = (R^* \otimes \mathrm{Gl}(n)) \otimes_\epsilon \mathrm{H}(n). \tag{43}$$

Note that this expression involves a direct product. The fact that M is normal in $\hat{G}$ follows by composing the homorphism $s: \hat{G} \to R^* \otimes R^*$ with the homomorphism

$$t(\alpha,\beta) = (\mathrm{sgn}(\alpha), \mathrm{sgn}(\beta)) \tag{44}$$

of $R^* \otimes R^*$ onto $Z_2 \otimes Z_2$ and noting that M is the kernel of $t \circ s$.

For our final generator, we define

$$\tau(X,Y,E) = (Y,X,X \cdot Y - E). \tag{45}$$

Note that $\tau^2 = 1$ and that τ is a Legendre transformation[10] exchanging all conjugate variables.

$$\tau^*(\omega) = d(X \cdot Y - E) - X \cdot dY = -dE + Y \cdot dx = -\omega, \tag{46}$$

$$\tau^*(\eta) = \eta. \tag{47}$$

Thus $\tau \in \hat{G}$ with $\alpha(\tau) = -1$ and $\beta(\tau) = 1$. τ generates a two element subgroup whose intersection with M is the identity since $\tau \notin M$. If f is an arbitrary element of $\hat{G}$, then either $f \in M$ or $\alpha(f) = -\beta(f)$ by Lemma 3. Then $\tau f \in M$, since

$$\alpha(\tau f) = \alpha(\tau)\alpha(f) = -\alpha(f) = \beta(f) = \beta(\tau)\beta(f) = \beta(\tau f). \tag{48}$$

We have proved the following theorem.

Theorem 4:

$$\hat{G} \cong Z_2 \otimes_\gamma (R^* \otimes_\beta (\mathrm{Gl}(n) \otimes_\alpha \mathrm{H}(n))) \cong Z_2 \otimes_\gamma ((R^* \otimes \mathrm{Gl}(n)) \otimes_\epsilon \mathrm{H}(n)).$$

The fact that the last product is only semidirect follows from the remaining commutation relations:

$$T_{(\mathbf{b},\mathbf{d},\mathbf{e})} \tau = \tau T_{\tau(\mathbf{b},\mathbf{d},\mathbf{e})}, \tag{49}$$

$$L_A \tau = \tau L_{(A^t)^{-1}}, \tag{50}$$

$$U_a \tau = \tau U_a L_{aI}. \tag{51}$$

The connected component of the identity N which is a normal subgroup of $\hat{G}$ is easily obtained from Theorem 4 by taking products of the connected components of the factors.

$$N \cong R \otimes_\beta (\mathrm{Gl}^+(n) \otimes_\alpha \mathrm{H}(n)) \cong ((R \otimes \mathrm{Gl}^+(n)) \otimes_\epsilon \mathrm{H}(n)), \tag{52}$$

where we have used $(R, +)$ instead of the isomorphic group $(R^+, \cdot)$. The generators τ, U_{-1}, and L_A together with N generate $\hat{G}$, where $A \in \mathrm{Gl}(n)$, $\det(A) = -1$, and $A^2 = I$. Since

$$\tau U_{-1} \tau U_{-1} = L_{-I}, \tag{53}$$

however, these generators do not generate a subgroup disjoint from N except for the identity unless $U_{-I} \notin N$. For n odd, this is the case since then $-I \notin \mathrm{Gl}^+(n)$ and we get $\hat{G} = D_4 \otimes_\lambda N$, where D_4 is the diheral group $Z_2 \otimes_\mu Z_4$ of symmetries of the square.[11] For n even, $-I \in \mathrm{Gl}^+(n)$ and no decomposition of the above form is possible. We have, in this case only, that

$$\hat{G}/N = Z_2 \otimes Z_2 \otimes Z_2. \tag{54}$$

IV. INTERPRETATIONS AND CONCLUSIONS

We have found that the group of coordinate transformations in the phase space of a thermodynamic system having

$$\omega = dE - Y\,dX \tag{55}$$

and

$$\eta = dX \cdot dY \tag{56}$$

as relative invariants is $\hat{G} \cong Z_2 \otimes_\gamma (R^* \otimes_\beta (\mathrm{Gl}(n) \otimes_\alpha \mathrm{H}(n)))$ with generators $T_{(b,d,e)}$, L_A, U_a, and τ, and commutation relations

$$L_A T_{(\mathbf{b},\mathbf{d},\mathbf{e})} = T_{L_A(\mathbf{b},\mathbf{d},\mathbf{e})} L_A, \tag{57}$$

$$U_a L_A = L_A U_a, \tag{58}$$

$$U_a T_{(\mathbf{b},\mathbf{d},\mathbf{e})} = T_{(\mathbf{b},a\mathbf{d},a\mathbf{e})} U_a, \tag{59}$$

$$T_{(\mathbf{b},\mathbf{d},\mathbf{e})} \tau = \tau T_{(\mathbf{d},\mathbf{b},\mathbf{d}\cdot\mathbf{b}-\mathbf{e})}, \tag{60}$$

$$L_A \tau = \tau L_A, \tag{61}$$

$$U_a \tau = \tau U_a L_{aI}. \tag{62}$$

If we require that ω and η be absolute invariants, the appropriate group shrinks to $G \cong \mathrm{Gl}(n) \otimes_\alpha \mathrm{H}(n)$ with generators L_A and $T_{(\mathbf{b},\mathbf{d},\mathbf{e})}$, and commutation relation (57).

For ω, relative invariance is physically a more reasonable requirement than absolute invariance, since $\omega = 0$ and $\alpha\omega = 0$ define the same solution surface. On the other hand, absolute invariance seems more reasonable for η, though relative invariance may be interpreted as a change of units of energy (see below). If we require absolute invariance of η and relative invariance of ω, then by Lemma 3, $\alpha = \pm 1$. The group becomes $\tilde{G} = Z_2 \otimes_\gamma (\mathrm{Gl}(n) \otimes_\alpha \mathrm{H}(n))$. $\tilde{G}$ is generated by τ, L_A, and $T_{(\mathbf{b},\mathbf{d},\mathbf{e})}$, i.e., the generators of $\hat{G}$ excluding U_a.

If we require the invariance of the origin, i.e., $f(0) = 0$, the generators $T_{(\mathbf{b},\mathbf{d},\mathbf{e})}$ are eliminated, and we are left with $Z_2 \otimes_\alpha (R^* \otimes_\beta \mathrm{Gl}(n))$ for relative invariance—relations (58), (61), and (62)—or with $\mathrm{Gl}(n)$ for absolute invariance.

We advance the following interpretations for the generators:

τ: τ corresponds to a classical Legendre transformation. As discussed in Sec. II, its interpretation[10,12] is to describe the state of a system using states of its environment, i.e., using Y instead of X.

U_a: U_a multiplies both E and Y by the constant factor a. This is exactly the effect of a change in the unit of energy.

L_a: Applying L_a may prove convenient when dealing with a chemical system. We can choose the coordinates AX to involve reaction coordinates and mole numbers of independent components. Using AX and the corresponding intensities $(A^t)^{-1}Y$ (affinities) can simplify analysis.[13]

$T_{(\mathbf{b},\mathbf{d},\mathbf{e})}$: This generator is difficult to interpret. The potential $P = E + \mathbf{d}X + \mathbf{e}$, which results from the action of $T_{(\mathbf{b},\mathbf{d},\mathbf{e})}$ on (X,Y,E), gives one clue. Note first that

$$T_{(\mathbf{b},\mathbf{d},\mathbf{e})} = T_{(\mathbf{b},0,\mathbf{e})}\, T_{(0,\mathbf{d},0)}. \tag{63}$$

For translations of the form $T_{(0,\mathbf{d},0)}$, if we interpret the Y variables as intensities representing the environment,[12] then replacing Y by $Y + \mathbf{d}$, i.e., placing the same system into an environment with intensities $Y + \mathbf{d}$, gives rise to the extra internal energy $\mathbf{d}\cdot X$. The translation $T_{(\mathbf{b},0,\mathbf{e})}$ corresponds then to changing the zero of our intensities. Note that $\mathbf{b}$ does not show up in the potential so an interpretation analogous to the one for $T_{(0,\mathbf{d},0)}$ is not possible. One is tempted to rule out transformations of the form $T_{(\mathbf{b},0,\mathbf{e})}$ by requiring that the origin in the subspace (X,E) remain invariant. However, since

$$\tau T_{(0,d,0)}\tau = T_{(d,0,0)}, \tag{64}$$

allowing translations in Y requires that we allow translations in X. Since $T_{(0,\mathbf{d},0)}$ moves the zero of E, we cannot require the invariance of $X = 0$ or $E = 0$ without requiring $Y = 0$. It is possible to rule out translation altogether by requiring the invariance of the origin in (X,Y,E). As mentioned above, this gives the group $G^* \cong Z_2 \otimes_\gamma (R^* \otimes Gl(n))$.

Note that none of the group generators, and hence no element of $\hat{G}$, mix the X and Y variables. More precisely, if $f \in \hat{G}$ sends (X,Y,E) to $(U,V,P) = f(X,Y,E)$, then there are exactly n variables (U or V) which depend only on X and exactly n variables (V or U) which depend only on Y. In general, P may depend on all $(2n + 1)$ variables (X,Y,E). If X and Y are initially extensive and intensive, there will again be $(n + 1)$ extensive and n intensive variables after the action of any $f \in \hat{G}$. Even when no such initial division into extensive and intensive variables can be made, e.g., when considering surface effects, the interpretation of the two sets of variables as parameters of state and parameters of environment remains valid.[12] Our conclusion then is that $f \in G$ does not mix parameters of the system and parameters of the environment except in the potential function P whose extrema determine the coexisting states of the system with the environment.

Finally, we note that $\hat{G}$ does not contain most of the classical Legendre transforms: namely, those which exchange only some of the conjugate pairs of variables. For example,

$$g(x_1,x_2,y_1,y_2,E) = (y_1,x_2,x_1,y_2,x_1y_1 - E) \tag{65}$$

preserves η but not ω since

$$d(x_1y_1 - E) - x_1\, dy_1 - x_2\, dx_2 \neq (dE - y_1\, dx_1 - x_2\, dx_2)\alpha \tag{66}$$

for any $\alpha \in R^*$. We can regain the invariance of ω by using instead

$$g(x_1,x_2,y_1,y_2,E) = (y_1,x_2,-x_1,y_2,E - x_1y_1), \tag{67}$$

which is more familiar from standard treatments.[10] In this case, however, η is not invariant since

$$dx_1\, dy_1 + dx_2\, dy_2 \neq (-dx_1\, dy_1 + dx_2\, dy_2)\beta \tag{68}$$

for any $\beta \in R^*$.

Suppose, however, that we are interested only in processes for which one of the x_i (respectively y_i) remains constant. Along the corresponding subsurface of phase-energy space, we can take the differential form dx_i (dy_i) to be zero and ask for the invariance of ω and η with such zeros dropped out. If one chooses the right potential, then it is possible to eliminate the pair of variables x_i, y_i from consideration.

Case A: $x_i = \text{const}$. In this case, x_i and y_i both drop out of ω and η on setting $dx_i = 0$. The problem thereby reduces to a problem with one less degree of freedom. This is implicit in the standard neglect of degress of freedom (e.g., magnetic) which "don't participate in a given process."

Case B: $y_i = \text{const}$. In this case, x_i and y_i do not drop out of ω. However, if we apply $\tau(X,Y,E) = (Y,X,X\cdot Y - E)$, we get Case A with the potential $X\cdot Y - E$. We can then throw away x_i and y_i, since they disappear from $\tau^*(\omega)$ and $\tau^*(\eta)$. Let $\hat{\tau}$ be the Legrendre involution on the space $(\hat{X},\hat{Y},P)$ with $\hat{\tau}(\hat{X},\hat{Y},P) = (\hat{Y},\hat{X},\hat{X}\cdot\hat{Y} - P)$, where $\hat{X} = (x,\ldots,x_{i-1}, x_{i+1},\ldots,x_n)$, $\hat{Y} = (y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_n)$. Applying $\hat{\tau}$, we regain the familiar form:

$$\hat{\tau}(\hat{Y},\hat{X},X\cdot Y - E) = (\hat{X},\hat{Y},E - x_iy_i); \tag{69}$$

the partial Legrende transform of E on the ith variable. The fact that thermodynamic analyses standardly employ such potentials for processes in which some $y_i = \text{const}$ fits in nicely with the above formalism.

In conclusion, we contrast the above group to previous group theoretic investigations in equilibrium thermodynamics. Koenig[14] and others,[15–17] while considering transformations associated with the Born diagram[10] have discussed contact transformations which also satisfy

$$E + G = H + A, \tag{70}$$

where E,G,H,A are the four classical thermodynamic potentials. The resulting group is generated by permutations of the variables induced by classical Legrende transformations. In light of work on generalizations of the thermodynamic Legrende transformation,[4,12] it appears that the invariance of (70) may be too strict a requirement.

Tisza,[18] observing that the matrix $\partial^2 E/\partial x_i\, \partial x_j$ is related to the stability of the system, studied coordinate transformations that leave the determinant and all principal minors of this matrix invariant.

As the first referee pointed out, it would be desirable to see a group theoretic investigation which required the invariance of the first two laws. Weinhold interprets the positive definiteness of η for stable systems to be the state space version of the second law. Thus all our transformations leave the second law invariant (τ changes the sign to negative definite). Preliminary evidence[2,3] seems to indicate, however, that not just the sign, but the magnitude of η, may be physically significant and will perhaps yield a strengthened form of the second law.

ACKNOWLEDGMENTS

We gratefully acknowledge helpful conversations with B. Andresen, H. Bray, D. Soda, and A. Swimmer.

[1]F. Weinhold, J. Chem. Phys. **63**, 2479, 2484, 2488, 2496 (1975); **65**, 559 (1976).
[2]P. Salamon, B. Andresen, P. D. Gait, and R. S. Berry, J. Chem. Phys. **73**, 1001 (1980).
[3]P. Salamon, and R. S. Berry, "Weinhold Length and Dissipated Availability," Phys. Rev. Lett. (to be published).
[4]R. Hermann, *Geometry, Physics, and Systems* (Marcel Dekker, New York, 1973).
[5]R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds* (Macmillan, New York, 1968), p. 272.
[6]R. Howe, Bull. Am. Math. Soc. **3**, 779 (1980).
[7]K. B. Wolf, in *Group Theory and Its Applications*, Vol. III, edited by E. M. Loebl (Academic, New York, 1981), p. 189.
[8]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[9]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1963), p. 238.
[10]See, for example, H. B. Callen, *Thermodynamics* (Wiley, New York, 1960).
[11]This copy of D_4 in $\hat{G}$ bears no relation to the transformations of Koenig[14] which derive from partial Legendre transforms corresponding to the symmetries of the Born square.
[12]P. Salamon, Ph.D. thesis (University of Chicago, 1978); P. Salamon, B. Andresen, and R. S. Berry, Phys. Rev. A **15**, 2094 (1977).
[13]I. Prigogine and R. Defay, *Chemical Thermodynamics* (Longmans, London, 1954).
[14]F. Koenig, J. Chem. Phys. **3**, 39 (1935); **56**, 4556 (1972).
[15]J. A. Prins, J. Chem. Phys. **16**, 65 (1948).
[16]F. Buckley, J. Res. Natl. Bur. Stand. **33**, 213 (1944).
[17]W. D. Hayes, Quart. Appl. Math. **4**, 227 (1945).
[18]L. Tisza, *Generalized Thermodynamics* (M.I.T., Cambridge, MA, 1966).

Solution of the Dirac equation for the general even power potential with application to particle spectroscopy

L. K. Sharma, Sheilly Mehta, and V. P. Iyer
Department of Applied Physics, Government Engineering College, Jabalpur-482 011, India

(Received 13 May 1982; accepted for publication 24 September 1982)

High-energy asymptotic expansions of eigenenergies, wavefunctions, and Regge trajectories of the generalized even power potential $V(r) = -g^2\Sigma_{j=0}^{\infty} N_{2j} r^{2j}$ in the Dirac equation are obtained. These general expansions are then used to obtain eigenenergy expansions and Regge trajectories for the anharmonic oscillator and Gauss potentials. Finally, the relevance of these investigations to the spectroscopy of heavy quark composites is discussed. In particular, we study the charmonium spectroscopy for an harmonic oscillator.

PACS numbers: 11.10.Qr, 14.80.Dq

I. INTRODUCTION

The discovery[1] of the ψ family of particles marks an important turning point in the development of particle physics. Recent development[2] strongly suggests the existence of the charmed quark proposed by Glashow and collaborators,[3] and that the ψ particles are charm–anticharm quark bound states. Since the states of the ψ spectroscopy are very narrow, their masses are accurately determined, and a lot of their properties are known. The ψ spectroscopy, therefore, provides an ideal testing ground for any comprehensive model of hadron dynamics.

Recently nonrelativistic quantum mechanics as embodied in the Schrödinger equation with a linear potential[4,5] has been successfully employed by high energy physicists, in the discovery of ψ (3 GeV/c^2) and Υ (10 GeV/c^2) families of heavy neutral mesons. Because of this success it is important to go beyond the nonrelativistic approach, which must be regarded as a first approximation for a complicated hadronic system.

The complete treatment should incorporate both relativistic and quantum effects and, in addition, requires a full understanding of the underlying dynamics of the quarks. However, the difficulties encountered in bound-state problems in relativistic field theory suggested that the complete solution of this problem is rather remote at the present time. Nevertheless, one might hope to gain some insight by examining crude approximations which include some of the effects. For studying the hadron spectrum, relativistic equations for a linear potential model in the frame work of Dirac and Klein–Gordon type wave equations have been investigated.[6,7] In certain quark-confining potentials,[8–12] the Dirac equation has also been solved with scalar potential functions. In the MIT bag model, free relativistic quarks are confined in a hadron bag of finite radial dimension by the conditions that the quark density vanishes on the bag surface and the quark pressure balances the natural pressure of the confined bag. Thus, by construction, the quarks can never "leak out" of the bag. From a purely phenomenological viewpoint,[13] the meson spectra using a harmonic-oscillator potential has been calculated in the Klein–Gordon equation. A semirelativistic quark model for mesons has also been using a square well potential.[14]

The general even power finds wide applications in potential theory since such well-known potentials as the harmonic oscillator, Gauss potential, and anharmonic oscillator potentials (with even anharmonicities) may be derived from this as particular cases. The general even power potential has already been studied in the Schrödinger framework[15] and also in the Klein–Gordon framework.[16] Motivated by the desire to complete the chain of investigating this potential for both the nonrelativistic and relativistic particles, the present paper deals with the derivation of the high eigenenergy and Regge trajectory expansions for a general even power potential in the Dirac equation. This has been done in Sec. II of this paper. In the present investigations, the perturbation procedure explained in Ref. 17 has been used. Applications of these eigenenergy expansions to the Gauss and anharmonic oscillator potentials have been given in Sec. III. In this section meson spectra of charmed quarks is also investigated using the harmonic oscillator potential. In Sec. IV, some concluding remarks have been given.

II. ASYMPTOTIC EIGENSOLUTIONS FOR THE GENERAL EVEN POWER POTENTIAL

The two coupled equations for the radial parts of the Dirac equation can be written as

$$\begin{aligned} &\frac{d}{dr}(rf) - \frac{k}{r}(rf) + [E - m - V(r)](rg) = 0, \\ &\frac{d}{dr}(rg) + \frac{k}{r}(rg) - [E + m - V(r)](rf) = 0, \end{aligned} \tag{2.1}$$

with parameter k defined as

$$k = \mp(j + \tfrac{1}{2}) \quad \text{for } l = j \pm \tfrac{1}{2}. \tag{2.2}$$

Here l is the orbital angular momentum designation in nonrelativistic nomenclature.

Further setting

$$\begin{aligned} &E^2 = m^2 + \kappa^2, \quad K = i\kappa, \\ &z = -2Kr, \quad \dot{X} = \frac{dx}{dz}, \end{aligned} \tag{2.3}$$

and

$$\begin{aligned} rg(r) &= e^{Kr} r^{k+1} X(r), \\ rf(r) &= e^{Kr} r^{k} X(r), \end{aligned} \tag{2.4}$$

we obtain for $rg(r)$

 0022-2488/83/102521-07$02.50

$$Z\ddot{X}+\left[2(k+1)-\frac{1}{(E+m-V)}\cdot\frac{z}{2K}\frac{d}{dr}V(r)-z\right]\dot{X}$$
$$-\left[(k+1)-(V^2-2EV)\frac{z}{4K^2}+\frac{1}{(E+m-V)}\right.$$
$$\left.\times\left(-\frac{z}{4K}+\frac{2k+1}{2K}\right)\frac{dV}{dr}\right]X=0. \tag{2.5}$$

We now consider the potential of the form

$$V(r)=-g^2\sum_{j=0}^{\infty}N_{2j}r^{2j}, \tag{2.6}$$

where the coefficients N_{2j} can be negative or positive.

Substitution of (2.6) in (2.5) yields

$$Z\ddot{X}+\{2(k+1)-Z\}\dot{X}-(k+1)X$$
$$=\sum_{j=0}^{\infty}\frac{1}{(2K)^{2j+1}}\left\{\left[\frac{2K}{E+m}z^{2j}(2jN_{2j})(-1)^{2j-1}+\frac{2K}{(E+m)^2}\sum_{p=0}^{\infty}(2jN_{2p}N_{2j})(-1)^{2p+2j-1}\cdot\frac{z^{2(p+j)}}{(2K)^{2p}}\right]\dot{X}\right.$$
$$-\left[(Mj-2EN_{2j})\frac{z^{2j+1}}{2K}-\frac{2KjN_{2j}}{(E+m)}z^{2j}-\frac{4jK(2k+1)N_{2j}}{(E+m)}z^{2j-1}\right.$$
$$\left.\left.-\frac{1}{(E+m)^2}\sum_{p=0}^{\infty}N_{2p}N_{2j}\frac{z^{2(p+j)}}{(2K)^{p-1}}\left(1-\frac{2j(2k+1)}{z(2K)^p}\right)\right]X\right\}. \tag{2.7}$$

Here,

$$M_j=\sum_{j=0}^{\infty}N_{2(j-i)}N_{2i} \tag{2.8}$$

and have taken

$$\frac{1}{(E+m-V)}\approx\frac{1}{(E+m)}\left(1+\frac{V}{E+m}\right) \tag{2.9}$$

as a first approximation. In the limit $|K|\to\infty$, (2.7) may be approximated by

$$Z\ddot{X}^{(1)}+[2(k+1)-z]\dot{X}^{(1)}-(k+1)X^{(1)}=0. \tag{2.10}$$

A solution of Eq. (2.10) is

$$X^{(1)}=\phi(a,b;z), \tag{2.11}$$

where ϕ is a confluent hypergeometric function and

$$a=k+1,\quad b=2(k+1). \tag{2.12}$$

This solution will be normalizable bound-state wavefunction if

$$a=-n\quad\text{for } n=0,1,2,\ldots. \tag{2.13}$$

For the complete solution we set

$$k+1=a+\Delta(K)/2K=-n+\Delta(K)/2K. \tag{2.14}$$

The quantity in Eq. (2.14) remains to be determined; substituting (2.14) in (2.7), we have an equation which can be written

$$\mathscr{D}_nX=\Delta(K)h+\sum_{j=0}^{\infty}\frac{1}{(2K)^{2j+1}}\left\{\left[\frac{2K}{(E+m)}2jN_{2j}(-1)^{2j-1}z^{2j}+\frac{2K}{(E+m)^2}\sum_{p=0}^{\infty}2jN_{2p}N_{2j}(-1)^{2p+2j-1}\frac{z^{2(p+j)}}{(2K)^{2p}}\right]\dot{X}\right.$$
$$-\left[(M_j-2EN_{2j})\frac{z^{2j+1}}{2K}-\frac{2KjN_{2j}}{(E+m)}z^{2j}-\frac{4jK(2k+1)N_{2j}}{(E+m)}z^{2j-1}\right.$$
$$\left.\left.-\frac{1}{(E+m)^2}\sum_{p=0}^{\infty}N_{2p}N_{2j}\frac{z^{2(p+j)}}{(2K)^{p-1}}\left(1-\frac{2j(2k+1)}{z(2K)^p}\right)\right]X\right\}, \tag{2.15}$$

where

$$\mathscr{D}_n=z\frac{d^2}{dz^2}+(b-z)\frac{d}{dz}+n \tag{2.16}$$

and

$$h=1/2K. \tag{2.17}$$

Equation (2.15) is now in a form suitable for the application of perturbation method. As a first approximation to X we have (apart from an overall normalization constant)

$$X=X^{(1)}=\phi_n(z). \tag{2.18}$$

The first approximation then leaves uncompensated terms amounting to

$$R^{(1)} = \Delta h + \sum_{j=0}^{\infty} \frac{1}{(2K)^{2j+1}} \left\{ \left[\frac{2K}{(E+m)} 2jN_{2j}(-1)^{2j-1} z^{2j} + \frac{2K}{(E+m)^2} \sum_{p=0}^{\infty} 2jN_{2p}N_{2j}(-1)^{2p+2j-1} \frac{z^{2(p+j)}}{(2K)^{2p}} \right] \dot{\phi}_n \right.$$
$$- \left[(M_j - 2EN_{2j}) \frac{z^{2j+1}}{2K} - \frac{2KjN_{2j}}{(E+m)} z^{2j} - \frac{4jK(2k+1)N_{2j}}{(E+m)} z^{2j-1} \right.$$
$$\left. \left. - \frac{1}{(E+m)^2} \sum_{p=0}^{\infty} N_{2p}N_{2j} \frac{z^{2(p+j)}}{(2K)^{p-1}} \left(1 - \frac{2j(2k+1)}{z(2K)^p} \right) \right] \phi_n \right\}. \tag{2.19}$$

For convenience we set $\phi_n(z) = \phi(a,b;z) = \phi(a)$ and write the recurrence relation for $\phi(a)$ in the form

$$z\phi(a) = (a,a+1)\phi(a+1) + (a,a)\phi(a) + (a,a-1)\phi(a-1), \tag{2.20}$$

where

$$(a,a+1) = a = -n = k+1-\Delta h,$$
$$(a,a) = b - 2a = 2(k+n+1) = 2\Delta h, \tag{2.21}$$
$$(a,a-1) = a - b = -(2k+2+n) = n - 2\Delta h.$$

By repeated application of (2.20) we obtain the following general relation:

$$z^m\phi(a) = \sum_{j=-m}^{m} S_m(a, a+j)\phi(a+j), \tag{2.22}$$

where the coefficients $S_m(a, a+\gamma)$ satisfy the recurrence relation

$$S_m(a, a+\gamma) = S_{m-1}(a, a+\gamma-1)(a+\gamma-1, a+\gamma) + S_{m-1}(a, a+\gamma)(a+\gamma, a+\gamma) + S_{m-1}(a, a+\gamma+1)(a+\gamma+1, a+\gamma), \tag{2.23}$$

with

$$S_0(a, a) = 1,$$
$$\text{all } S_0(a, a+i) = 0 \quad \text{for } i \neq 0 \tag{2.24}$$
$$\text{and } S_m(a, a+\gamma) = 0 \quad \text{for } |\gamma| > m.$$

By means of term-by-term differentiation of confluent hypergeometric function $\phi(a)$ yields

$$\dot{\phi}(a, b; z) = \frac{a}{b}\phi(a+1, b+1; z);$$

this relation when combined with the relations between contiguous functions gives the following expression:

$$z\dot{\phi}(a, b; z) = a[\phi(a+1, b; z) - \phi(a, b; z)]. \tag{2.25}$$

The expression for $R^{(1)}$ with the help of (2.22) and (2.25) can be written

$$R^{(1)} = \sum_{j=0}^{\infty} -h^{2j+1}\{[a, a+2j+4]_{2j+1}\phi(a+2j+4) + [a, a+2j+3]_{2j+1}\phi(a+2j+3) + \cdots + [a, a-2j-3]_{2j+1}\phi(a-2j-3) + [a, a-2j-4]_{2j+1}\phi(a-2j-4)\}, \tag{2.26}$$

where

$$[a, a]_1 = \Delta + D_j h S_1(a, a),$$
$$[a, a \pm (2j+4)]_{2j+1} = h^3[J_j S_{2j+4}(a, a \pm 2j \pm 4) - C_j S_{2j+3}(a+1, a \pm 2j \pm 4)],$$
$$[a, a \pm (2j+3)]_{2j+1} = -h^3\{C_j[S_{2j+3}(a+1, a \pm 2j \pm 3) - S_{2j+3}(a, a \pm 2j \pm 3)] - J_j S_{2j+4}(a, a \pm 2j \pm 3) + P_j S_{2j+3}(a, a \pm 2j \pm 3)\},$$
$$[a, a \pm (2j+2)]_{2j+1} = -h\{h^2[C_j S_{2j+3}(a+1, a \pm 2j \pm 2) - C_j S_{2j+3}(a, a \pm 2j \pm 2) - J_j S_{2j+4}(a, a \pm 2j \pm 2) + P_j S_{2j+3}(a, a \pm 2j \pm 2)] - I_j S_{2j+2}(a, a \pm 2j \pm 2)\},$$
$$[a, a \pm (2j+1)]_{2j+1} = -h\{h^2[C_j S_{2j+3}(a+1, a \pm 2j \pm 1) - C_j S_{2j+3}(a, a \pm 2j \pm 1) - J_j S_{2j+4}(a, a \pm 2j \pm 1) + P_j S_{2j+3}(a, a \pm 2j \pm 1)] + B_j S_{2j+1}(a+1, a \pm 2j \pm 1) - B_j S_{2j+1}(a, a \pm 2j \pm 1) - D_j S_{2j+1}(a, a \pm 2j \pm 1) - I_j S_{2j+2}(a, a \pm 2j \pm 1) + L_j S_{2j+1}(a, a \pm 2j \pm 1)\},$$
$$[a, a \pm 2j]_{2j+1} = \frac{1}{h}[A_j S_{2j-1}(a+1, a \pm 2j) + (F_j + H_j)S_{2j}(a, a \pm 2j)] - h\{B_j S_{2j+1}(a+1, a+2j) - (B_j + D_j - L_j)S_{2j+1}(a, a \pm 2j) - I_j S_{2j+2}(a, a \pm 2j) + h^2[C_j S_{2j+3}(a+1, a \pm 2j) - (C_j - P_j)S_{2j+3}(a, a \pm 2j) - J_j S_{2j+4}(a, a \pm 2j)]\},$$
$$[a, a \pm (2j-1)]_{2j+1} = \frac{1}{h}[A_j S_{2j-1}(a+1, a \pm 2j \mp 1) - (A_j - G_j - K'_j)S_{2j-1}(a, a \pm 2j \mp 1) + (F_j + H_j)S_{2j}(a, a \pm 2j \mp 1)] + h[B_j S_{2j+1}(a+1, a \pm 2j \mp 1) - (B_j - D_j - L_j)S_{2j+1}(a, a \pm 2j \mp 1) + I_j S_{2j+2}(a, a \pm 2j \mp 1)] + h^3[C_j S_{2j+3}(a+1, a \pm 2j \mp 1) - (C_j - P_j)S_{2j+3}(a, a \pm 2j \mp 1) + J_j S_{2j+4}(a, a \pm 2j \mp 1)],$$
$$\vdots$$
$$[a, a]_{2j+1} = \frac{1}{h}[A_j S_{2j-1}(a+1, a) - (A_j - G_j - K'_j)S_{2j-1}(a, a) + (F_j + H_j)S_{2j}(a, a)] + h[B_j S_{2j+1}(a+1, a) - (B_j - D_j - L_j)S_{2j+1}(a, a) + I_j S_{2j+2}(a, a)] + h^3[C_j S_{2j+3}(a+1, a) - (C_j - P_j)S_{2j+3}(a, a) + J_j S_{2j+4}(a, a)]. \tag{2.27}$$

The constants A_j, B_j, etc. have the following meaning:

$$
\begin{aligned}
A_j &= g^2(2jnN_{2j})/(E+m) - g^4(2jN_0N_{2j})/(E+m)^2, \\
B_j &= g^4(2jnN_2N_{2j})/(E+m)^2, \\
C_j &= g^4(2jnN_4N_{2j})/(E+m)^2, \\
D_j &= g^2(2EN_{2j}) - g^4M_j, \\
F_j &= g^2(jN_{2j})/E + m, \\
G_j &= g^2(2k+1)(2jN_{2j})/(E+m), \\
H_j &= g^4(jN_0N_{2j})/(E+m)^2, \\
I_j &= g^4(jN_2N_{2j})/(E+m)^2, \\
J_j &= g^4(jN_4N_{2j})/(E+m)^2, \\
K_j' &= g^4(2k+1)(2jN_0N_{2j})/(E+m)^2, \\
L_j &= -g^4(2k+1)(2jN_2N_{2j})/(E+m)^2, \\
P_j &= -g^4(2k+1)(2jN_4N_{2j})/(E+m)^2.
\end{aligned}
\tag{2.28}
$$

We now observe that $\mathscr{D}_n\phi(a+q) = q\phi(a+q)$ so that a term $\mu\phi(a+q)$ may be removed by adding to $X^{(1)}$ the contribution $(\mu/q)\phi(a+q)$ except, of course, when $q=0$.

Hence the next contribution to $X^{(1)}$ becomes

$$
X^{(2)} = \sum_{j=0}^{\infty} h^{2j+1}\left[\frac{[a, a\pm 2j\pm 4]_{2j+1}}{\pm(2j+4)}\phi(a\pm 2j\pm 4) + \frac{[a, a\pm 2j\pm 3]_{2j+1}}{\pm(2j+3)}\phi(a\pm 2j\pm 3) + \cdots\right]. \tag{2.29}
$$

This contribution leaves uncompensated a sum of terms $R^{(2)}$, which again leads to $X^{(3)}$. Repeating this process successively and adding these contributions to $X^{(1)}$ yields

$$
X = X^{(1)} + X^{(2)} + X^{(3)} + \cdots. \tag{2.30}
$$

However, (2.30) will be a solution of our equation only if the sum of all terms containing $\phi(a)$ in $R^{(1)}$, $R^{(2)}$, ··· left compensated so far is set equal to zero. Thus

$$
\begin{aligned}
O = {} & h\,[a,a]_1 + h^2\left\{\frac{[a,a+1]_1[a+1,a]_1}{1} + \frac{[a,a-1]_1[a-1,a]_1}{-1}\right\} \\
& + h^3\left\{[a,a]_3 + \frac{[a,a+1]_1[a+1,a+1]_1[a+1,a]_1}{1} + \frac{[a,a-1]_1[a-1;a-1]_1[a-1,a]_1}{1}\right\} + O(h^4).
\end{aligned}
\tag{2.31}
$$

The expansion (2.30) is then an eigensolution and (2.31) the appropriate secular equation which enables us to calculate Δ and hence the eigenenergy for large K or small h. Explicit calculation of terms upto $O(h^5)$ yields the following expression for the eigenenergies:

$$
\begin{aligned}
n+k+1 = \Delta h = {} & \frac{1}{1 + P_1'h^2 + P_2'h^3 + P_3'h^4 + P_4'h^5} \\
& \times\left[(P_5'h^2 + P_6'h^3 + P_7'h^4 + P_8'h^5) + (\Delta h)^2(P_9'h^2 + P_{10}'h^3 + P_{11}'h^4 + P_{12}'h^5)\right. \\
& \left. + (\Delta h)^3(P_{13}'h^2 + P_{14}'h^4 + P_{15}'h^5) + (\Delta h)^4(P_{16}'h^2 + P_{17}'h^4 + P_{18}'h^5)\right],
\end{aligned}
\tag{2.32}
$$

$$
\begin{aligned}
P_1' &= [-(4+4n)A_1 + 2(G_1 + K_1') + (2+4n)(F_1 + H_1) \\
&\quad + 2D_0 + (1+6n)D_0^2], \\
P_2' &= [(4+8n)A_1D_0 - (10 - 36n - 3n^2)(F_1 + H_1)D_0 \\
&\quad - 4(G_1 + K_1')D_0 + 4nD_0^2], \\
P_3' &= [-(28 - 12n - 24n^2)B_1 + (4 - 12n - 12n^2)(L_1 + D_1 + G_2 + K_2') \\
&\quad + (12 - 4n + 60n^2 - 120n^3)(I_1 + F_2 + H_2) - (28 + 36n - 24n^2)A_2 \\
&\quad + (4 - 8n - 4n^2)D_0^3 - (8 + 8n + 56n^2 + 8n^4)D_0^2 - (8n - 16n^2)], \\
P_4' &= [-(176 + 76n - 316n^2 + 480n^3 - 224n^4)I_1D_0 \\
&\quad + (108n + 792n^2)(L_1 + D_1)D_0 + (154 - 16n - 40n^2)B_1D_0], \\
P_5' &= [(1+5n)(A_1 + F_1 + H_1) - nD_0^2], \\
P_6' &= [-(2 + 4n + 3n^2)A_1D_0 + (20n - 50n^2)(F_1 + H_1)D_0 \\
&\quad - (4 - 4n)(G_1 + K_1')D_0 - 4n^2D_0^2],
\end{aligned}
$$

$$
\begin{aligned}
P'_7 &= [(6-12n-6n^2-3n^3)B_1-(6n+52n^2+36n^3-176n^4)(I_1+F_2+H_2)\\
&\quad +(6+3n-6n^2-3n^3)A_2+(8n+8n^2+32n^3)D_0^3+8n^2D_0^2],\\
P'_8 &= [(224+176n+20n^2-20n^3+352n^4-224n^5)I_1D_0\\
&\quad -(24n+36n^2+336n^3)(L_1+D_1)D_0-(24+40n+90n^2+88n^3)B_1D_0],\\
P'_9 &= [4(F_1+H_1)+(2+6n)D_0^2],\\
P'_{10} &= [-4A_1D_0-(12-4n)(F_1+H_1)D_0],\\
P'_{11} &= [(24-24n)B_1+(12+24n)(L_1+D_1+G_2+K'_2)\\
&\quad +(44-72n+72n^2)(I_1+F_2+H_2)+(24-24n)A_2\\
&\quad +(4+8n+4n^2+10n^3)D_0^3+(16n+16n^3)D_0^2+8],\\
P'_{12} &= [-(352-88n-336n^2)I_1D_0-(72+476n)(L_1+D_1)D_0\\
&\quad -(136-76n)B_1D_0+(6+6n+18n^2)D_0^3-(3n-3n^2-18n^3)D_0^4],\\
P'_{13} &= [-(2+2n+4n^2)D_0^2],\\
P'_{14} &= [-32B_1+8(L_1+D_1+G_2+K'_2)+(48+64n)(I_1+F_2+H_2)\\
&\quad -32A_2+(4-8n+18n^2)D_0^3],\\
P'_{15} &= [-(384+640n)I_1D_0+176B_1D_0],\\
P'_{16} &= [(2+4n)D_0^2],\\
P'_{17} &= [16[I_1+F_2+H_2]+(4+8n)D_0^3],\\
P'_{18} &= [128I_1D_0].
\end{aligned}
\tag{2.33}
$$

Expanding the denominator of (2.32) in powers of h and iterating for (Δh), we get

$$
\kappa = -n-1+[h^2(P'_5)+h^3(P'_6)+h^4(P'_7-P'_1P'_5)+h^5(P'_8-P'_1P'_6-P'_2P'_5)]+O(h^6). \tag{2.34}
$$

Similarly the eigenenergy for the wavefunction (rf) can be evaluated by setting in (2.34)

$$
n = n'-1, \quad n' = 0, 1, 2, \ldots . \tag{2.35}
$$

It may also be noted that for the wavefunction (rf) the value of b is

$$
b = 2k. \tag{2.36}
$$

III. APPLICATIONS OF THE GENERAL EIGENENERGY EXPANSION

We now apply the eigenenergy expansion (2.34) to the two cases.

A. Gauss potential

The Gauss potential is given as

$$
V(r) = -g^2e^{-\alpha^2r^2}, \tag{3.1}
$$

so that

$$
N_{2j} = (-1)^j\alpha^{2j}/j!. \tag{3.2}
$$

Hence the eigenenergy expansion is obtained directly from (2.34) wherein the terms P_j terms are given by (2.33). Various

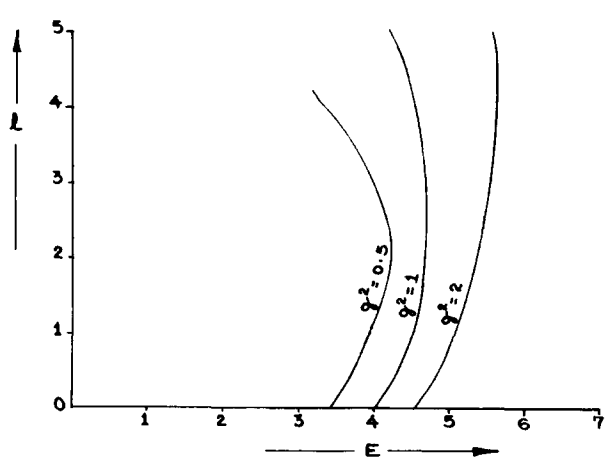


FIG. 1. Ground state Regge trajectories for the Gauss potential using parameters $m = 1$, $\alpha^2 = 1$, and different values of coupling constant g^2.

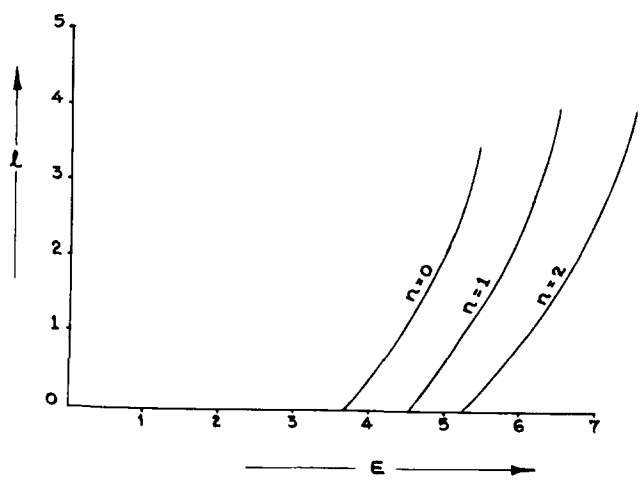


FIG. 2. Regge trajectories for anharmonic oscillator for different values of n. Other parameters are $N_0 = -0.05$ GeV, $N_2 = 0.0005$ GeV^3, $N_4 = 0.0005$ GeV^5, and $g = 1$.

TABLE I. Predicted masses in GeV for mesons with charmed quark pairs with the oscillator potential of parameter $N_0 = -0.05$ GeV, $N_2 = 0.0005$ GeV^3, $g = 1$, and $m = 1.5$ GeV.

n	l	Meson mass (cc^-) GeV Theory	Expt.
0	S	3.096	3.097 ± 0.002
1	S	3.69	3.686 ± 0.003
2	S	4.12	4.03 ± 0.01
0	P	3.46	3.446
1	P	4.00	3.876 99
2	P	4.35	4.147 3

N_{2j} coefficients occurring in (2.28) are defined by (3.2) for this case.

The ground state Regge trajectories for the Gauss potential have been shown in Fig. 1.

B. Anharmonic oscillator

Next we consider the potential

$$V(r) = -g^2(N_0 + N_2 r^2 + N_4 r^4). \tag{3.3}$$

The eigenenergy expansion is given by (2.34) where the terms A_j, B_j, etc., are obtained by putting $j = 0, 1, 2$.

The Regge trajectories for this potential are shown in Fig. 2.

C. Charmonium spectroscopy

We study the potential

$$V(r) = -g^2(N_0 + N_2 r^2). \tag{3.4}$$

Here in the eigenenergy expansion obtained from (2.34) the terms A_j, B_j, etc., are obtained by putting $j = 0, 1$.

In fact, for obtaining the Regge trajectories for the above potentials and for calculating the meson masses, Eq. (2.32) (which is a fourth power equation in Δh) has been used.

The meson masses are given in Table I and Regge trajectories for the same are shown in Fig. 3.

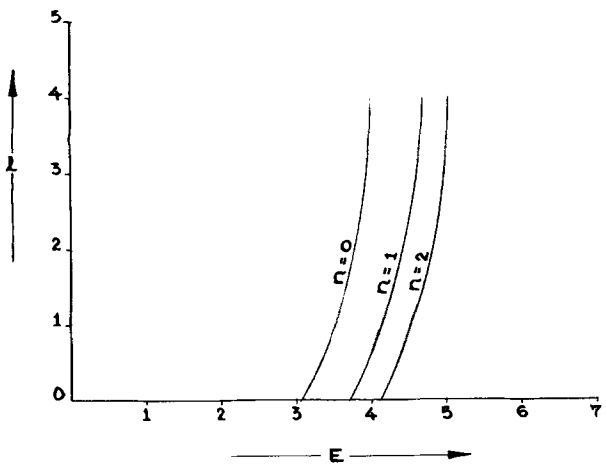


FIG. 3. Regge trajectories for the harmonic oscillator potential in the Dirac equation with parameters $N_0 = -0.05$ GeV, $N_2 = 0.0005$ GeV^3, $g = 1$, and $m = 1.5$ GeV.

TABLE II. Bound state energies for different values of reduced mass (μ) with $N_0 = -0.002$ GeV, $N_2 = 0.0005$ GeV^3, and $g = 1$.

μ	E
0.0	1.0206
0.2	1.33
0.4	1.63
0.6	1.84
0.8	1.99
1.0	2.958
1.2	3.31
1.4	3.89
1.6	4.986
1.8	5.52
2.0	6.98

IV. DISCUSSION

It is well known that in the nonrelativistic framework the linear potential yields the most rapidly rising trajectories. Similarly, oscillator potential in this frame work also yields the linearly rising trajectory. The same sort of behavior was observed for the Regge trajectory in the Klein–Gordon frame work.[16] The Regge trajectories obtained in Fig. 3 also appear to be nearly parallel; the behavior of these trajectories, therefore, is not altered for the case of the Dirac equation. The meson masses with charmed quark–antiquark have also been calculated using this potential (see Table I), and these results are in very good agreement with those obtained experimentally. Although numerical methods have been extensively used in this type of work, it has been thought worthwhile to obtain theoretical expressions for the eigenenergies and wavefunctions using a perturbation theory also. In Table II we have calculated bound state energies for different values of reduced mass μ. From this table we conclude that it is possible to generate bound state mass spectra of a system of both light and heavy quarks in an effective harmonic oscillator potential in the Dirac equation. [For any value of μ the eigenenergies are found to be pure real; therefore, it is possible for the harmonic oscillator potential model to explain relativistic quark confinement.] It may be relevant to point out here that a similar observation[18] has also been recently made.

The Regge trajectories for the anharmonic oscillator with r^4 anharmonicity (see Fig. 2) are also nearly linear in the range of E considered. The ground state Regge trajectories for the Gauss potential (see Fig. 1), though not linear, have nearly equal slopes.

ACKNOWLEDGMENT

One of us (S.M.) acknowledges with thanks the University of Jabalpur, India, for the award of a research fellowship.

[1] J. J. Aubert *et al.*, Phys. Rev. Lett. **33**, 1404 (1974); J. E. Augustin *et al.*, *ibid.* **33**, 1406 (1974).

[2] G. Goldhaber *et al.*, Phys. Rev. Lett. **37**, 255 (2976); I. Peruzzi *et al.*, *ibid.* **37**, 569 (1976).

[3]J. D. Bjorken and S. L. Glashow, Phys. Lett. **11**, 255 (1964); S. L. Glashow, J. Iliopoulos, and L. Malani, Phys. Rev. D **2**, 1285 (1970).
[4]E. Eichten *et al.*, Phys. Rev. Lett. **34**, 369 (1975); B. J. Harrington, S. Y. Park, and A. Yildz, *ibid.* **34**, 168 (1975); K. Jhung, K. Chung, and R. S. Willey, Phys. Rev. D **12**, 1999 (1975).
[5]J. F. Gunion and R. S. Willey, Phys. Rev. D **12**, 174 (1975).
[6]J. F. Gunion and L. F. Li, Phys. Rev. D **12**, 3583 (1975).
[7]J. S. Kang and H. J. Schnitzer, Phys. Rev. D **12**, 841 (1975).
[8]C. L. Critchfield, Phys. Rev. D. **12**, 923 (1975).
[9]C. L. Critchfield, J. Math. Phys. **17**, 261 (1976).
[10]D. W. Rein, Nuovo Cimento **38**, 19 (1977).
[11]B. Ram and R. Halasa, Lett. Nuovo Cimento **26**, 551 (1979).
[12]V. P. Iyer and L. K. Sharma, Phys. Lett. B **102**, 154 (1981).
[13]B. Ram and R. Halasa, Phys. Rev. D **19**, 3467 (1979).
[14]P. Horwitz, Phys. Rev. **161**, 1415 (1967).
[15]L. K. Sharma, J. Choubey, and H. J. W. Müller-Kirsten, J. Math. Phys. **21**, 1533 (1980).
[16]L. K. Sharma and V. P. Iyer, J. Math. Phys. **23**, 1185 (1982).
[17]H. J. W. Müller-Kirsten, Phys. Rev. D **22**, 1952 (1980), and other publications by the same author given under reference in this paper.
[18]N. Barik and B. K. Barik, Pramāṇa **17**, 489 (1981).

On topological boundary characteristics in nonabelian gauge theory

C. Cronström
Department of Theoretical Physics, University of Helsinki, Siltavuorenpenger 20 B, Helsinki 17, Finland

J. Mickelsson
Department of Mathematics, University of Jyväskylä, Seminaarinkatu 15, Jyväskylä 10, Finland, and Department of Theoretical Physics, University of Helsinki, Siltavuorenpenger 20 B, Helsinki 17, Finland

(Received 14 March 1983; accepted for publication 15 April 1983)

We study a topological classification of gauge potentials based on an examination of the Chern–Simons surface term $C(U)$ at appropriate boundary components of the space-time manifold when the potential approaches a pure gauge $dU\,U^{-1}$ at the boundary. We derive an explicit local formula for a 2-form $H(U)$ such that $C(U) = dH(U)$.

PACS numbers: 11.15. − q, 02.40. + m, 02.20.Sv

I. BASIC NOTATIONS AND CONVENTIONS

We consider a pure Yang–Mills theory[1] in Minkowski space. The gauge group G is assumed to be semisimple and compact, for simplicity. In an Appendix, our main result is proved without redundant assumptions. The Yang–Mills Lagrangian is (summation convention)

$$L = -\tfrac{1}{4}G^{a}_{\mu\nu}G^{\mu\nu}_{a}, \tag{1}$$

where the gauge field $G_{\mu\nu}$ is expressed in terms of the gauge potential A_μ as follows:

$$G^{a}_{\mu\nu} = \partial_\nu A^{a}_{\mu} - \partial_\mu A^{a}_{\nu} + g f^{a}_{bc} A^{b}_{\mu} A^{c}_{\nu}. \tag{2}$$

The quantity g in Eq. (2) is an arbitrary (real) coupling constant, and the quantities f^{c}_{ab} are the structure constants of the Lie algebra $\mathcal{g}$ of the gauge group

$$[T_a, T_b] = i f^{c}_{ab} T_c. \tag{3}$$

The Lie algebra (or color) indices $(a, b, c, \ldots)$ are lowered (raised) with the aid of the Killing metric $g_{ab}(g^{ab})$, where

$$g_{ab} = -f^{d}_{ac} f^{c}_{bd} \tag{4}$$

and

$$g^{ab}g_{bc} = \delta^{a}_{c}. \tag{5}$$

Our Minkowski metric is $g_{\mu\nu}$ $(\mu, \nu = 0, 1, 2, 3)$ with $(g_{\mu\nu}) = \mathrm{diag}\,(1, -1, -1, -1)$.

It is convenient to use a matrix representation for the gauge field $G_{\mu\nu}$ and potential A_μ, respectively. Thus let $\{\hat{T}_a\}$ $(a = 1,\ldots,\dim \mathcal{g})$ denote any suitable set of Hermitian matrix representatives of the Lie algebra generators (3). Then we write

$$A_\mu = A^{a}_{\mu}\hat{T}_a, \tag{6}$$

from which follows

$$G_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu - ig[A_\mu, A_\nu] \tag{7}$$

with the aid of the definition (2) and the commutation relations (3). Finally, using the totally antisymmetric density $\epsilon^{\alpha\beta\gamma\delta}$ (convention: $\epsilon^{0123} = +1$), we define the dual gauge field $G^{*\mu\nu}$ as follows:

$$G^{*\mu\nu} = \tfrac{1}{2}\epsilon^{\mu\nu\alpha\beta}G_{\alpha\beta}. \tag{8}$$

II. TOPOLOGICAL DENSITIES AND CURRENTS

As is well known,[2–4] a basic topological density $\nu(x; A)$ (the second Chern class) can be formed out of the gauge field $G_{\mu\nu}$ and its dual:

$$\nu(x; A) = -(g^2/32\pi^2)G_{\mu\nu}G^{*\mu\nu}. \tag{9}$$

The definition (7) of the gauge field in terms of a potential A_μ implies the Bianchi identities

$$\nabla_\alpha(A)G_{\beta\gamma} + \nabla_\beta(A)G_{\gamma\alpha} + \nabla_\gamma(A)G_{\alpha\beta} = 0, \tag{10}$$

where $\nabla_\alpha(A)$ denotes the covariant derivative

$$\nabla_\alpha(A) \equiv \partial_\alpha + ig[A_\alpha, \ ]. \tag{11}$$

The Bianchi identities (10) expressed in terms of the dual gauge field are as follows:

$$\nabla_\mu(A)G^{*\mu\nu} = 0. \tag{12}$$

The identities (12) guarantee that the topological density $\nu(x; A)$ given by Eq. (9) is insensitive to *local* deformations of the potential A_μ:

$$\frac{\delta\nu(x; A)}{\delta A^{a}_{\mu}(y)} = \frac{g^2}{8\pi^2}\delta^{(4)}(x - y)(\nabla_\nu(A)G^{*\mu\nu}(x))_a = 0. \tag{13}$$

Thus the topological density $\nu(x; A)$ depends only on the large scale properties of the potential A_μ. An explicit realization of this fact is the representation of $\nu(x; A)$ as a divergence of a topological current K^μ,

$$\nu(x; A) = \partial_\mu K^{\mu}(A), \tag{14}$$

where

$$K^{\mu}(A) = \frac{g^2}{32\pi^2}\epsilon^{\mu\alpha\beta\gamma}A^{a}_{\alpha}\left(G_{\beta\gamma a}(A) - \frac{1}{3}g f_{bca}A^{b}_{\beta}A^{c}_{\gamma}\right). \tag{15}$$

The representation (14) is, as such, well known in the physical literature.[2–4] Equation (15) is a particular case of a more general formula given in Ref. 5.

The main purpose of this paper is to establish a further topological classification of gauge-field configurations, using Eqs. (14) and (15) as a starting point.

To this end, we consider gauge transformations of the

 0022-2488/83/102528-04$02.50

current (15). The potential A_μ transforms as follows under gauge transformations U:

$$A_\mu \xrightarrow{U} A_\mu^U = UA_\mu U^{-1} + (i/g)(\partial_\mu U)U^{-1}. \tag{16}$$

Then from definition (15) follows,

$$K^\mu(A) \xrightarrow{U} K^\mu(A^U)$$

$$= K^\mu(A) + \frac{ig}{16\pi^2}\epsilon^{\mu\alpha\beta\gamma}\partial_\gamma(U^{-1}\partial_\alpha U, A_\beta) + \frac{1}{96\pi^2}\epsilon^{\mu\alpha\beta\gamma}(U^{-1}\partial_\alpha U, [U^{-1}\partial_\beta U, U^{-1}\partial_\gamma U]), \tag{17}$$

where (A, B) denotes the Killing inner product (4),

$$(A, B) \equiv g_{ab}A^a B^b = A_a B^a. \tag{18}$$

The divergence of the topological current K^μ is gauge invariant. Thus the divergence of the U-dependent terms of the right-hand side of Eq. (17) ought to vanish identically. The last term on the right-hand side of Eq. (17), i.e., the quantity

$$C^\mu(U) \equiv \frac{1}{96\pi^2}\epsilon^{\mu\alpha\beta\gamma}(U^{-1}\partial_\alpha U, [U^{-1}\partial_\beta U, U^{-1}\partial_\gamma U]) \tag{19}$$

is not manifestly divergence-free, however. In what follows, an equivalent formula will be given for the quantity $C^\mu(U)$, which manifestly demonstrates the vanishing of its divergence. The formula in question, which will be given shortly, also demonstrates that $C^\mu(U)$ has a topological significance, as a characteristic of the gauge transformation U. This fact is not unknown in the mathematical literature; quantities related to our $C^\mu(U)$ are known as the Chern–Simons secondary characteristics.[5]

We demonstrate first that $C^\mu(U)$ indeed is a kind of topological current. Consider variations δ_U of $C^\mu(U)$. It is not difficult to verify that

$$\delta_U C^\mu(U) = (1/32\pi^2)\epsilon^{\mu\alpha\beta\gamma}\partial_\alpha \times(\delta U\cdot U^{-1}, [\partial_\beta U\cdot U^{-1}, \partial_\gamma U\cdot U^{-1}]). \tag{20}$$

This means that the variation of the integral of $C^\mu(U)$ over a three-dimensional submanifold V in Minkowski space can be written as an integral over the boundary ∂V of V. Thus in this sense, $C^\mu(U)$ is insensitive to local deformations of U, just as the topological density $\nu(x; A)$ defined previously was insensitive to local deformations of A. Hence the quantity $C^\mu(U)$ can be considered as a topological current for gauge transformations U. Said in other words, for fixed boundary conditions of U on ∂V, the integral of C^μ characterizes the relative homotopy class (with respect to ∂V) of the map U.

Equation (20) [and of course the fact that $\partial_\mu C^\mu(U) = 0$] indicates that $C^\mu(U)$ ought to have a representation of the following form:

$$C^\mu(U) = \partial_\alpha H^{\mu\alpha}(U), \tag{21}$$

where $H^{\mu\alpha}$ is antisymmetric under the interchange of μ and α, that is, H is a 2-form. If U is defined in a star shaped neighborhood D of V in Minkowski space, the existence of H in D follows from the Poincaré lemma. We shall, in the next section, show that there exists a simple local expression for H. In general, the existence of H is guaranteed if the third cohomology group of V vanishes. This happens, for example, when $V = S^2 \times I$ ($I \subset \mathbb{R}$ an interval) which is a natural space to study in the Lagrangian formulation of Yang–Mills theories. In this case it is namely reasonable to assume that the potential A_μ approaches an exact gauge $U^{-1}\partial_\mu U$ at spatial infinity (represented by the sphere S^2) in a given time interval I.

III. REPRESENTATION OF $C^\mu(U)$ AS A TENSOR DIVERGENCE

We shall now give a detailed discussion of the representation (21) announced in the previous section.

To begin with, a few restrictive assumptions are stated.

We consider a suitable three-dimensional surface V in Minkowski space M_4. By a gauge transformation $U(x)$, $x \in V$, we mean a $\mathbb{C}^2$-mapping from V into the gauge group G, U: $V \to G$. In fact, in order to write differential forms in a convenient coordinate form, we assume that $U(x)$ is defined in a small open neighborhood of V. More specifically, G is here identified with a particular (unitary) matrix representation, corresponding to the Lie algebra representation $\{\hat{T}_a\}$ discussed previously. Further, we assume that U has an exponential (Lie algebra) representation,

$$U(x) = e^{iZ(x)}, \quad x \in V, \tag{22}$$

where Z: $V \to \mathcal{g}$ is a $\mathbb{C}^2$-mapping into the Lie algebra $\mathcal{g}$, which here again is identified with the matrix representation discussed above. A necessary condition for the existence of the lifting Z of U is that U as a map from V to G is homotopically trivial (homotopic to a constant map). The Hermitian matrices $Z(x)$ can be written as vectors in the linear vector space spanned by the $\hat{T}_a$'s,

$$Z(x) = Z^a(x)\hat{T}_a, \quad Z^a \text{ real}. \tag{23}$$

The basic quantities occurring in the definition (19) of the topological current $C^\mu(U)$ are the following:

$$W_\alpha \equiv U^{-1}\partial_\alpha U. \tag{24}$$

The quantities iW_α belong to $\mathcal{g}$, i.e., can be written as in Eq. (23). Now we wish to express the quantities W_α defined by Eq. (24) in terms of the $\mathcal{g}$-valued quantities $Z(x)$ occurring in Eq. (22). This is accomplished with the aid of the Magnus formula[6]:

$$W_\alpha = ((1 - e^{-i\,\mathrm{ad}\,Z})/\mathrm{ad}\,Z)\partial_\alpha Z, \tag{25}$$

where $\mathrm{ad}\,Z(Y) \equiv [Z, Y]$ for $Z, Y \in \mathcal{g}$.

The operator function in Eq. (25) is defined as a power series:

$$\frac{e^x - 1}{x} = \sum_{n=0}^{\infty} \frac{x^n}{(n+1)!}. \tag{26}$$

After these preliminaries, we are ready to state our main result.

The topological current $C^\mu(U)$ defined in Eq. (19), viz.,

$$C^\mu(U) = (1/96\pi^2)\epsilon^{\mu\alpha\beta\gamma}(W_\alpha, [W_\beta, W_\gamma]), \tag{27}$$

has the following equivalent representation:

$$C^\mu(U) = \partial_\alpha H^{\mu\alpha}(Z), \tag{28}$$

where

$$H^{\mu\alpha}(Z) = (1/48\pi^2)\epsilon^{\mu\alpha\beta\gamma}(h(-i\,\mathrm{ad}\,Z)\partial_\beta Z, \partial_\gamma Z). \tag{29}$$

The linear operator $h(-i\,\mathrm{ad}\,Z)$ in Eq. (29) is defined by its power series

$$h(x)=\sum_{k=0}^{\infty}\frac{x^{2k+1}}{(2k+3)!}=\frac{\sinh x-x}{x^2}. \tag{30}$$

In order to verify the validity of the representation (28), one needs the following result:

$$\begin{aligned}\partial_\alpha(\mathrm{ad}\,Z)^m\partial_\beta Z&=(\mathrm{ad}\,Z)^m\partial_\alpha\partial_\beta Z\\&\quad+\sum_{k=0}^{m-1}(\mathrm{ad}\,Z)^k[\partial_\alpha Z,(\mathrm{ad}\,Z)^{m-k-1}\partial_\beta Z],\\&\quad m=1,2,\ldots,\end{aligned} \tag{31}$$

the validity of which can be established by induction on m. One further needs the following symmetry property:

$$([Z,A],B)=-(A,[Z,B]), \tag{32}$$

valid for any Lie albegra $\mathcal{g}$.

The actual derivation of the result (28) and (29) is given in the Appendix using compact notation, i.e., differential forms.

To the best of our knowledge, the general representation (28), (29) has not hitherto been given in the literature. There is one special case of Eq. (29), however, which has been introduced and used extensively by Jackiw[7] for the purpose of topological classifications, namely, the case of $G=\mathrm{SU}(2)$. In this case we use the fundamental representation $\hat{T}_a=\frac{1}{2}\sigma_a$ $(a=1,2,3)$, where the σ_a's are the Pauli matrices. The SU(2) Lie algebra is

$$[T_a,T_b]=iT_c\quad(a,b,c\ \text{cyclic}), \tag{33}$$

so that

$$f_{abc}=2\epsilon_{abc}, \tag{34}$$

where ϵ_{abc} is the totally antisymmetric permutation symbol. The Killing metric g_{ab} is the following:

$$g_{ab}=\begin{cases}2 & a=b,\\ 0 & a\neq b.\end{cases} \tag{35}$$

For

$$Z=Z^a\tfrac{1}{2}\sigma_a, \tag{36}$$

one establishes easily that

$$(\mathrm{ad}\,Z)^{2k+1}\partial_\alpha Z=(\tfrac{1}{2}\|Z\|^2)^k\,\mathrm{ad}\,Z\partial_\alpha Z, \tag{37}$$

where

$$\|Z\|^2\equiv(Z,Z)=g_{ab}Z^aZ^b. \tag{38}$$

Using Eq. (37), one obtains from Eq. (29),

$$\begin{aligned}H^{\mu\alpha}(Z)=\frac{1}{48\pi^2}\epsilon^{\mu\alpha\beta\gamma}&\left(\frac{\|Z\|/\sqrt{2}-\sin(\|Z\|/\sqrt{2})}{(\|Z\|/\sqrt{2})^3}\right)\\&\times f_{abc}Z^a(\partial_\beta Z^b)(\partial_\gamma Z^c)\end{aligned} \tag{39}$$

which, allowing for differences in notation, is exactly the formula given by Jackiw. Our general formula thus generalizes Jackiw's SU(2) formula to any semisimple and compact Lie group (in fact any finite-dimensional group; see the Appendix).

Finally a brief comment on the use of our general result (28) and (29). We stated previously that the quantity $C^\mu(U)$ is a topological characteristic of the corresponding gauge transform U. Thus in studying the class of gauge potentials A, which approach a pure gauge at (space) infinity, one finds that these potentials fall into homotopy classes corresponding to the (relative) homotopy classes of the pure gauges, i.e., gauge transformations U. All this has [for SU(2)] been explained fairly explicitly by Jackiw, quoted previously. Our general representation enables one to carry through similar explicit arguments in general.

APPENDIX: DERIVATION OF $C^\mu=\partial_\alpha H^{\mu\alpha}$

Let $\mathcal{g}$ be the Lie algebra of a Lie group G. Consider a $\mathbb{C}^2$-function $U: D\to G$, where D is a domain in $\mathbb{R}^4$. We assume that U can be written in the form $U(x)=\exp Z(x)$, where $Z: D\to\mathcal{g}$ is a $\mathbb{C}^2$-mapping. We denote $A=dU\,U^{-1}$. According to the Magnus rule

$$A=f(\mathrm{ad}\,Z)dZ, \tag{A1}$$

where

$$f(x):=\sum_{n=0}^{\infty}\frac{1}{(n+1)!}x^n=\frac{e^x-1}{x}. \tag{A2}$$

Consider the 3-form

$$C=(A,[A,A]), \tag{A3}$$

where $(\cdot,\cdot)$ is the Killing form on $\mathcal{g}$. We claim that $C=dH$, where

$$\begin{aligned}&H=2(dZ,h(\mathrm{ad}\,Z)dZ),\\&h(x):=(1/x^2)\cdot(\sinh x-x).\end{aligned} \tag{A4}$$

Since $\partial_\nu A_\mu-\partial_\mu A_\nu-[A_\nu,A_\mu]=0$, we have

$$C=(A,dA). \tag{A5}$$

According to (A1),

$$\begin{aligned}\partial_\nu A_\mu&=\partial_\nu\sum\frac{1}{(n+1)!}(\mathrm{ad}\,Z)^n\partial_\mu Z\\&=\sum_{n=1}^{\infty}\sum_{k=0}^{n-1}\frac{1}{(n+1)!}(\mathrm{ad}\,Z)^k\\&\quad\times[\partial_\nu Z,(\mathrm{ad}\,Z)^{n-k-1}\partial_\mu Z]+f(\mathrm{ad}\,Z)\partial_\nu\partial_\mu Z.\end{aligned} \tag{A6}$$

Combining (A5) and (A6),

$$C=2\left(A,\sum_{n=1}^{\infty}\sum_{k=0}^{n-1}\frac{1}{(n+1)!}(\mathrm{ad}\,Z)^k[dZ,(\mathrm{ad}\,Z)^{n-k-1}dZ]\right). \tag{A7}$$

Using the property (32) of the Killing form, we get

$$\begin{aligned}C&=2\left(\sum_{l=0}^{\infty}\frac{1}{(l+1)!}(\mathrm{ad}\,Z)^l dZ,\ldots\right)\\&=2\left(dZ,\sum_{l=0}^{\infty}\sum_{n=1}^{\infty}\sum_{k=0}^{n-1}\frac{1}{(l+1)!}\frac{1}{(n+1)!}(-1)^l(\mathrm{ad}\,Z)^{k+l}\right.\\&\quad\left.\times[dZ,(\mathrm{ad}\,Z)^{n-k-1}dZ]\right).\end{aligned} \tag{A8}$$

Since

$$\begin{aligned}&(dZ,(\mathrm{ad}\,Z)^p[dZ,(\mathrm{ad}\,Z)^q dZ])\\&\quad=(-1)^p((\mathrm{ad}\,Z)^p dZ,[dZ,(\mathrm{ad}\,Z)^q dZ])\\&\quad=-(-1)^p([dZ,(\mathrm{ad}\,Z)^p dZ],(\mathrm{ad}\,Z)^q dZ)\\&\quad=-(-1)^{p+q}(dZ,(\mathrm{ad}\,Z)^q[dZ,(\mathrm{ad}\,Z)^p dZ]),\end{aligned}$$

we can write the right-hand side of (A8) as follows:

$$\begin{aligned}
&\left(dZ, \sum_{l,n,k} \frac{1}{(l+1)!}\frac{1}{(n+1)!}(-1)^l(\mathrm{ad}\,Z)^{k+l}[dZ,(\mathrm{ad}\,Z)^{n-k-1}dZ]\right)\\
&\quad+\left(dZ, \sum_{l,n,k} \frac{1}{(l+1)!}\frac{1}{(n+1)!}(-1)^l(-1)^{n+l-1}(\mathrm{ad}\,Z)^{n-k-1}[dZ,(\mathrm{ad}\,Z)^{k+l}dZ]\right)\\
&\quad=\left(dZ, \sum_{m=1}^{\infty}\sum_{j=0}^{m-1}\sum_{l=0}^{j}\frac{1}{(l+1)!}\frac{1}{(m-l+1)!}(-1)^l(\mathrm{ad}\,Z)^j[dZ,(\mathrm{ad}\,Z)^{m-j-1}dZ]\right)\\
&\quad+\left(dZ, \sum_{m=1}^{\infty}\sum_{j=0}^{m-1}\sum_{l=0}^{m-j-1}\frac{1}{(l+1)!}\frac{1}{(m-l+1)!}(-1)^{m-l-1}(\mathrm{ad}\,Z)^j[dZ,(\mathrm{ad}\,Z)^{m-j-1}dZ]\right).
\end{aligned}\tag{A9}$$

Using the fact that

$$\begin{aligned}
&\sum_{l=0}^{j}\frac{1}{(l+1)!}\frac{1}{(m-l+1)!}(-1)^l+\sum_{l=0}^{m-j-1}\frac{1}{(l+1)!}\frac{1}{(m-l+1)!}(-1)^{m-l-1}\\
&\quad=\frac{1}{(m+2)!}\left(\sum_{l=0}^{j}(-1)^l\binom{m+2}{l+1}+\sum_{l=0}^{m-j-1}(-1)^{m-l-1}\binom{m+2}{l+1}\right)\\
&\quad=\frac{1}{(m+2)!}\left((-1)^j\binom{m+1}{j+1}+1+(-1)^{j-1}\binom{m+1}{m-j}-(-1)^m\right)\\
&\quad=\frac{1-(-1)^m}{(m+2)!},
\end{aligned}\tag{A10}$$

where $\binom{n}{k}=n!/k!(n-k)!$, we get

$$\begin{aligned}
C&=\left(dZ, \sum_{m=1}^{\infty}\sum_{j=0}^{m-1}\frac{1-(-1)^m}{(m+2)!}(\mathrm{ad}\,Z)^j[dZ,(\mathrm{ad}\,Z)^{m-j-1}dZ]\right)\\
&=\left(dZ, d\sum_{m=0}^{\infty}\frac{1-(-1)^m}{(m+2)!}(\mathrm{ad}\,Z)^m dZ\right)\\
&=2\left(dZ, d(h(\mathrm{ad}\,Z)dZ)\right)\\
&=2d\left(dZ, h(\mathrm{ad}\,Z)dZ\right)\\
&=dH.
\end{aligned}\tag{A11}$$

[1]C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
[2]A. A. Belavin *et al.*, Phys. Lett. B **59**, 85 (1975); see also R. Jackiw and C. Rebbi, Phys. Rev. Lett. **37**, 172 (1976).
[3]W. A. Bardeen, Nucl. Phys. B **75**, 246 (1974).
[4]R. J. Crewther, Acta Phys. Austriaca, Suppl. **19**, 47 (1978).
[5]S. Chern and J. Simons, Ann. Math. **99**, 48 (1974).
[6]W. Magnus, Commun. Pure Appl. Math. **7**, 649 (1954).
[7]R. Jackiw, "Gauge Theories in Three Dimensions," Lectures at the Arctic School of Physics, Äkäslompolo, Finland, August 1982, and references quoted therein. [*Springer Lecture Notes in Physics*, Vol. 181, edited by R. Raitio and J. Lindfors (Springer-Verlag, Berlin, 1982).]

On the hydrodynamic self-similar cosmological models

L. K. Chi
Computer Science Department, U.S. Naval Academy, Annapolis, Maryland 21402

(Received 26 October 1982; accepted for publication 8 April 1983)

The hydrodynamic self-similar cosmological models are considered. A new solution is presented. Of all the self-similar solutions, only the Newtonian analogy of the Einstein–deSitter model represents an expanding universe.

PACS numbers: 95.30.Lz, 98.90. − k

Self-similarity has long been used in hydrodynamics to solve degenerate problems or to find asymptotic solutions.[1] Recently, Henriksen and Wesson[2] have discussed some Newtonian and relativistic self-similar cosmological models. We first briefly sketch Henriksen and Wesson's approach of the hydrodynamic models and find a new self-similar solution. Then we make some interesting observations of the self-similar models.

The governing equations for the spherical symmetric, isentropic fluid flow in an internal gravitational field are the continuity equation, the momentum equation, the energy equation, and the gravitational equation:

$$\begin{aligned}
&\frac{\partial \rho}{\partial t}+v\frac{\partial \rho}{\partial r}=-\frac{\rho}{r}\frac{\partial}{\partial r}(r^2 v),\\
&\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial r}=g-\frac{1}{\rho}\frac{\partial p}{\partial r},\\
&\frac{\partial}{\partial t}(p\rho^{-\nu})+v\frac{\partial}{\partial r}(p\rho^{-\nu})=0,\\
&\frac{\partial g}{\partial t}+v\frac{\partial g}{r}=-\frac{2vg}{r},
\end{aligned} \tag{1}$$

where ρ, p, and v are, respectively, the local density, pressure, and velocity. g is the local gravitational acceleration.

If we assume that the Newtonian gravitational constant G is the only constant of nature and there is no characteristic velocity, we can introduce the dimensionless qualities η, Q, ϕ, and V as follows:

$$\begin{aligned}
&\rho=(\lambda/r^{\theta})\eta(\xi), \quad v=(r/t)V(\xi),\\
&g=(r/t^2)\phi(\xi), \quad p=(\lambda/r^{\theta-2}t^2)Q(\xi),
\end{aligned} \tag{2}$$

where

$$\xi=G\lambda t^2/r^{\theta} \tag{3}$$

and θ is a constant of order unity. The parameter λ is introduced so that ξ is dimensionless. In terms of the dimensionless variables in (2) and (3), Eqs. (1) become

$$\begin{aligned}
&(2-\theta V)\xi\eta'=(\theta-3)\eta V+\theta\xi\eta V',\\
&(2-\theta V)\xi V'=\phi+V-V^2+\frac{(\theta-2)Q}{\eta}+\frac{\theta\xi Q'}{\eta},\\
&(2-\theta V)\xi\phi'=(2-3V)\phi,\\
&(2-\theta V)\xi Q'=(2+\theta V)Q+(2-\theta V)\frac{\xi\nu Q\eta'}{\eta}\\
&\qquad\qquad -(2+\nu\theta)QV,
\end{aligned} \tag{4}$$

where primes denote differentiation with respect to ξ.

Two solutions of Eq. (4) can easily be found. If we assume that $V=1$ and $\theta\neq 2$, then $\nu=\frac{4}{3}$ and

$$\begin{aligned}
&\eta=\eta_0\xi^{(\theta-3)/(\theta-2)}, \quad \phi=\phi_0\xi^{-1/(2-\theta)},\\
&Q=Q_0\xi^{(\theta-4)/(2-\theta)},
\end{aligned} \tag{5}$$

where the constants η_0, ϕ_0, and Q_0 are related by $Q_0=(2-\theta)\phi_0\eta_0/4$. If we assume that $V=\frac{2}{3}$ and $\theta\neq 3$, then $\nu=\frac{4}{3}$ and

$$\eta=\eta_0/\xi, \quad \phi=\phi_0, \quad Q=Q_0/\xi, \tag{6}$$

where the constants η_0, ϕ_0, and Q_0 are related by $Q_0=\eta_0(2+9\phi_0)/18$. When $\phi_0=-\frac{2}{9}$, Q_0 is zero and solution (6) reduces to

$$\eta=\eta_0/\xi, \quad \phi=-\tfrac{2}{9}, \quad Q=0. \tag{7}$$

Solutions (5), (6), and (7) were described in Ref. 2. Instead of assuming $V=$ constant, we note that for $\theta=3$, $V\neq$ constant and $Q=0$, Eqs. (4) are greatly simplified:

$$\begin{aligned}
&(2-3V)\xi\eta'=3\xi\eta V',\\
&(2-3V)\xi V'=\phi+V-V^2,\\
&\xi\phi'=\phi.
\end{aligned} \tag{8}$$

It can be shown that Eqs. (8) are satisfied by the following solutions:

$$V=\sqrt{\xi}, \quad \phi=-\tfrac{1}{2}\xi, \quad \eta=\eta_1/(2-3\sqrt{\xi}), \tag{9}$$

where η_1 is a constant.

The expression for η indicates that solution (9) has a singularity at $\xi=\frac{4}{9}$. In fact, $\xi=\frac{4}{9}$ implies $V=\frac{2}{3}$. When $\theta=3$ and $V=\frac{2}{3}$, the system of Eqs. (4) is undetermined. It seems, therefore, the singularity is due to the introduction of self-similarity in Eqs. (1).

In dimensional form solution (9) becomes

$$\begin{aligned}
&v=\sqrt{\frac{\lambda G}{r}}, \quad \rho=\frac{\eta_1\lambda}{2r^3-3tr\sqrt{G\lambda r}},\\
&g=-\frac{G\lambda}{2r^2}, \quad P=0,
\end{aligned} \tag{10}$$

where η_1 is a constant. Solution (10) is well defined when $t=0$ and is, therefore, not consistent with the big-bang theory.

We now make some observations of the self-similar solutions. Using Eqs. (2) and (3), we can rewrite solution (7) in the following dimensional form:

$$v = 2r/3t, \quad \rho = \eta_0/Gt^2, \quad g = -2r/9t, \quad p = 0. \quad (11)$$

The constant η_0 can be determined by the additional requirement that the function $R(t)$, defined by

$$v = \frac{r}{R}\frac{dR}{dt}, \quad (12)$$

satisfies the equation

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi G\rho}{3}R^2. \quad (13)$$

Then $\eta_0 = \frac{1}{6}\pi$. Eq. (13) is simply the Friedmann differential equation with zero cosmological constant and zero curvature.[3] It is readily shown from Eqs. (11) and (12) that $R = R_0 t^{2/3}$ and $\rho R^3 =$ constant. If fact, solution (11) with $\eta_0 = \frac{1}{6}\pi$, is just the Newtonian analog of the zero-curvature Einstein–deSitter universe.

If we introduce the function H defined by $v = Hr$ or $H = (1/R)(dR/dt)$, then $H = \frac{2}{3}t$, and H satisfies

$$\frac{dH}{dt} + H^2 = -\frac{4\pi G}{3}\rho. \quad (14)$$

Solution (11) is seen as identical with the homogeneous and isotropic solution described by Zeldovich[4] with the local density equal to critical density. It is easily seen that of all the self-similar solutions discussed, only solution (11) satisfies both Eqs. (13) and (14).

It is also interesting to note that solution (11) can be obtained by assuming two constants of nature, namely, the gravitational constant G and a characteristic velocity, as shown in Ref. 2. Here, we assume that G is the only constant of nature.

Of all the self-similar solutions discussed here, only the Newtonian analog of the Einstein–deSitter universe describes an expanding, isotropic, and homogeneous universe. As shown in Ref. 4, the hydrodynamic Eqs. (1) admit three types of expanding universes: forever-expanding, expansion followed by compression, and the zero-curvature Einstein–deSitter analog. The additional requirement of self-similarity selects the Einstein–deSitter model.

ACKNOWLEDGMENT

Valuable comments by the referee are acknowledged.

[1] G. I. Barenblatt and Y. B. Zeldovich, "Self-similar solutions as intermediate asymptotics," Annu. Rev. Fluid Mech. 285 (1971).
[2] R. N. Henriksen and P. S. Wesson, "Self-similar space-times," Astrophys. Space Sci. **53**, 429 (1978).
[3] W. Rindler, *Essential Relativity Special, General and Cosmological* (Van Nostrand Reinhold, New York, 1969).
[4] Y. B. Zeldovich, "Hydrodynamics of the universe," Annu. Rev. Fluid Mech. 215 (1977).