Weight vector representatives for SO(2*N*)

S. K. Soni
University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 19 October 1981; accepted for publication 5 February 1982)

Following Mohapatra and Sakita, it is convenient to rewrite the algebra of $2N\gamma$-matrices in terms of N annihilation operators and N creation operators. We find them useful in representing weight vectors in the spinorial $\sigma_N + \sigma_N{}^c$ and the irreducible subcomponents of $\sigma_N \times \sigma_N$ and $\sigma_N \times \sigma_N{}^c$ operators. Though their Dynkin labels are readily accessible, a natural basis for the enumeration of their complete set of weights is given by N positive weights of the vectorial representation. Their subgroup content under $\mathrm{SU}(N) \subset \mathrm{SO}(2N)$ and $\mathrm{SO}(2m) \otimes \mathrm{SO}(2N-2m) \subset \mathrm{SO}(2N)$ is made obvious by using a simple identity. Conjugation and Yukawa couplings are touched on briefly.

PACS numbers: 02.10. + w

1. INTRODUCTION

It has become a part of the folklore to represent the (compact) generators of SO(2N) as $\Sigma_{\mu\nu} = \frac{1}{4}[\gamma_\mu, \gamma_\nu] \equiv \frac{1}{4}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu)$ by introducing $2N$ Hermitian γ-matrices satisfying the Clifford algebra

$$\{\gamma_\mu, \gamma_\nu\} \equiv \gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\delta_{\mu\nu}, \quad \mu, \nu = 1,2,\ldots,2N. \tag{1.1}$$

The Cartan subalgebra is generated by $\Sigma_{2i-1,2i}$ ($i = 1,2,\ldots,N$), which is diagonal in the basis $\frac{1}{2}(\gamma_{2i} + i\gamma_{2i-1}) \equiv b_i$, $\frac{1}{2}(\gamma_{2i} - i\gamma_{2i-1}) \equiv b_i^\dagger$ for the $2N$-dimensional vectorial representation γ_μ ($\mu = 1,2,\ldots,2N$). Following Mohapatra and Sakita[1], we consider the algebra satisfied by the N annihilation operators and N creation operators defined above.

$$\{b_i, b_j\} = \{b_i, b_j^\dagger\} - 2\delta_{ij} = 0, \quad i,j = 1,2,\ldots N. \tag{1.2}$$

They also introduce (SU(N)-invariant) vacuum $|0\rangle$

$$b_i|0\rangle = 0, \quad i = 1,2,\ldots N. \tag{1.3}$$

In terms of all those objects, we represent the shift action for (complex extension of the algebra associated with) SO(2N) and the highest weights of the spinorial representation $\sigma_N + \sigma_N{}^c$ as well as of $\phi_N^{\{k\}}$. Here $\phi_N^{\{k\}}$ denotes the tensor operator representation corresponding to the k-fold-linear antisymmetrized product of γ's, e.g., $\phi_N^{\{1\}}$ is the vectorial and $\phi_N^{\{2\}}$ is the adjoint. Given those weight-vector representatives (w.v.r.) for the highest weights, w.v.r. for the complete set of weights in $\sigma_N + \sigma_N^c$ and $\phi_N^{\{k\}}$ may be generated by repeated shift action. Though the Dynkin labels for weights (in the Dynkin basis) are readily accessible from w.v.r., a more natural basis for weights associated with w.v.r. is the self-dual basis spanned by the N positive weights in the vectorial representation. In the latter basis, the complete set of weights referred to above may be enumerated quite easily. In physical applications, knowledge of w.v.r. is useful for visualizing reflection under group-theoretic conjugation and for computing Clebsch–Gordan coefficients in group-invariant couplings of the $\sigma_N \times \sigma_N$ or the $\sigma_N \times \sigma_N^c$ operator with any one of the irreducible subcomponents contained in it.

We start off in Sec. 2 with the definition of the self-dual basis for SO(2N), reviewing also the Dynkin basis and its dual. Section 3 is devoted to the decompositions of $\sigma_N \times \sigma_N$ and $\sigma_N \times \sigma_N^c$ and the subgroup content of the irreducible representations contained in them. In Secs. 4 and 5 we come to the heart of our discussion. In Sec. 4, those SO(2N) shift operators whose shift action is either along or antiparallel to the simple roots, are represented as bilinears in b_i and $b_i^\dagger$, $i = 1,2,\ldots,N$. In Sec. 5, we give w.v.r. for $\sigma_N + \sigma_N^c$ and irreducible subcomponents in $\sigma_N \times \sigma_N$ and $\sigma_N \times \sigma_N^c$, starting from their representatives with highest weights. Sections 6 and 7, where we only briefly discuss conjugation and Yukawa couplings, respectively, are potentially useful for physical applications.

2. THE SELF-DUAL BASIS FOR SO(2*N*)

The Dynkin basis[2] and its dual are reviewed. The self-dual basis for SO(2N) is defined and related to the Dynkin basis for SO(2N).

Let α_i ($i = 1,2,\ldots,n =$ rank) denote the simple roots. The real linear vector space (l.v.s.)$\Sigma_{i=1}^n a_i\alpha_i$, a_i real, defines the idempotent whose real dimension n equals the complex dimension n of the Cartan subalgebra, the minimal l.v.s. generated by the idempotent over the field of complex numbers. The complete set of weights of a linear representation is a finite system of vectors in the idempotent, whose most important property is that the Cartan scalar product (x, y) defines a Euclidean metric on it. An arbitrary root, which is a nonzero weight in the adjoint representation, is expressible as a linear combination of the simple roots such that the nonzero coefficients are all positive (negative) integers, when the root is positive (negative).

Let $\alpha_i' \equiv 2\alpha_i/(\alpha_i, \alpha_i)$ denote the normalized simple root. An arbitrary vector X in the idempotent may be expanded as

$$X = \sum_{i=1}^{n} X_i\alpha_i'. \tag{2.1}$$

The X_i are called covariant components of X. The basis $\{\alpha_i'\}$ is dual to the basis $\{\Pi_i\}$ defined by the reciprocity relation

$$(\Pi_i, \alpha_j') = \delta_{ij}, \quad i,j = 1,2,\ldots,n. \tag{2.2}$$

$\{\alpha_i'\}$ may be regarded as the basis of a direct lattice; then $\{\Pi_i\}$ is the basis for its reciprocal lattice. The latter, called the Dynkin basis, defines the contravariant components of X through

$$X = \sum_{i=1}^{n} X^i\Pi_i. \tag{2.3}$$

0022-2488/83/040729-07$02.50

An important result is that the contravariant components A^i of the weight A of a representation, called its Dynkin labels, are all integral. Furthermore, Dynkin labels for the highest weight of an irreducible representation, denoted by $\lambda_1, \lambda_2, ..., \lambda_n$, are all nonnegative integers. These n-tuples are in one-to-one correspondence with irreducible representations of a semisimple algebra.

From the definitions (2.1)–(2.3) a number of useful results[2] follow:

$$X^i = (X, \alpha_i') = \sum_{j=1}^{n} h^{ij} X_j,$$

$$X_i = (X, \Pi_i) = \sum_{j=1}^{n} h_{ij} X^j,$$

$$h^{ij} = (\alpha_i', \alpha_j') \tag{2.4}$$

and h_{ij} denotes the inverse of the matrix h^{ij}, both of which are symmetric $n \times n$ matrices. The scalar product of a pair of arbitrary vectors in the idempotent is[2]

$$(X, Y) = \sum_{i=1}^{n} X^i Y_i = \sum_{i=1}^{n} X_i Y^i. \tag{2.5}$$

From the reciprocity relation and the definition $\lambda_j = \delta_{ij}$ ($j = 1,2,...,n$) for the ith fundamental representation, we may identify Π_i as the highest weight of that representation. In other words, the Dynkin basis is spanned by the n highest weights of the n fundamental representations.

Let us specialize to SO($2N$). In the self-dual basis, an arbitrary vector X in the idempotent (in particular, an arbitrary weight in a linear representation) is expressed as a linear combination of the N positive weights of the vectorial representation, denoted by $e_1, e_2, ..., e_N$ and ordered as $e_1 > e_2 > \cdots > e_{N-1} > e_N$,

$$X = \sum_{i=1}^{N} X(i) e_i. \tag{2.6}$$

This basis $\{e_i\}$ is dual to itself since

$$(e_i, e_j) = \delta_{ij}, \quad i, j = 1,2,...,N. \tag{2.7}$$

For a pair of arbitrary vectors X and Y in the idempotent,

$$(X, Y) = \sum_{i=1}^{N} X(i) Y(i). \tag{2.8}$$

The basis vectors in the self-dual and Dynkin bases are related as follows:

$$e_i = \sum_{j=1}^{N} \Pi_j A_{ji},$$

$$\Pi_i = \sum_{j=1}^{N} e_j (A^{-1})_{ji}, \tag{2.9}$$

$$A = \begin{pmatrix} 1 & -1 & 0 & & & & \\ 0 & 1 & -1 & & & \bigcirc & \\ 0 & 0 & 1 & & & & \\ & & & \ddots & & & \\ & & & & 1 & -1 & 0 \\ & \bigcirc & & & 0 & 1 & -1 \\ & & & & 0 & +1 & 1 \end{pmatrix},$$

$$A^{-1} = \begin{pmatrix} 1 & 1 & 1 & & 1 & \frac{1}{2} & \frac{1}{2} \\ & 1 & 1 & & 1 & \frac{1}{2} & \frac{1}{2} \\ & & 1 & & 1 & \frac{1}{2} & \frac{1}{2} \\ & & & \ddots & & & \\ & & & & 1 & \frac{1}{2} & \frac{1}{2} \\ & \bigcirc & & & & \frac{1}{2} & \frac{1}{2} \\ & & & & & -\frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

3. SUBGROUP CONTENT UNDER SU(N) AND SO($2m$) ⊗ SO($2n$) ⊂ SO($2N$)

A simple identity may be used to find the subgroup content of the irreducible subcomponents contained in $\sigma_N \times \sigma_N$ and $\sigma_N \times \sigma_N^c$. We consider the subgroups SU(N) and SO($2m$) ⊗ SO($2n$), $m + n = N$.

The lowest-dimensional spinorial representation of SO($2N$) (dimension 2^N) is reducible to the pair σ_N and σ_N^c of irreducible representations (dimension 2^{N-1} each), which are complex conjugates of each other for odd N but are real and inequivalent for even N. σ_N is reflected onto σ_N^c under conjugation effected by an outer automorphism of SO($2N$). For odd N,

$$\sigma_N(00\cdots010) \times \sigma_N(00\cdots010)$$

$$= \Sigma_{\text{odd}}\ (00\cdots020) \dot{+} \mathop{\sum\nolimits^{\cdot}}_{k=0}^{(N-3)/2} \phi_N^{\{2k+1\}}, \tag{3.1}$$

$$\sigma_N(00\cdots010) \times \sigma_N^c(00\cdots001) = \mathop{\sum\nolimits^{\cdot}}_{k'=1}^{(N-1)/2} \phi_N^{\{2k'\}}.$$

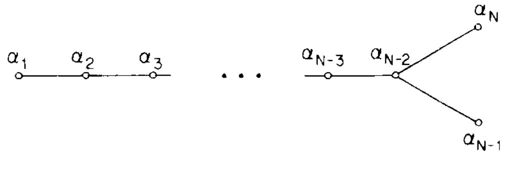
For even N,

$$\sigma_N(00\cdots010) \times \sigma_N(00\cdots010)$$

$$= \Sigma_{\text{even}}\ (00\cdots020) \dot{+} \mathop{\sum\nolimits^{\cdot}}_{k'=1}^{(N-2)/2} \phi_N{}^{\{2k'\}}, \tag{3.2}$$

$$\sigma_N(00\cdots010) \times \sigma_N^c(00\cdots001) = \mathop{\sum\nolimits^{\cdot}}_{k=0}^{(N-2)/2} \phi_N{}^{\{2k+1\}}.$$

Let us explain our notation. The N nonnegative integers within parentheses denote the n-tuples $(\lambda_1, \lambda_2, ..., \lambda_N)$ characterizing the irreducible representation denoted by Greek letters. The simple roots are labelled



Σ_{odd} and Σ_{even} are highest components, in the indicated decompositions, whose n-tuples are sums of those of the representations being multiplied. ϕ_N denotes the real $2N$-dimensional vectorial representation with n-tuples (1,0,0,...,0). $\phi^{\{p\}}$, the pth antisymmetrization of the linear representation ϕ, is defined as the linear closure of the vectors

$$\eta_1 \times \eta_2 \ldots \times \eta_p = \epsilon_{i_1 i_2 \ldots i_p} \eta_{i_1} \times \eta_{i_2} \times \ldots \times \eta_{i_p},$$

where $\eta_1 \times \eta_2 \times \ldots \times \eta_p$ denotes the vector to each system of vectors $\eta_1, \eta_2, ..., \eta_p$ from the l.v.s. on which ϕ acts and $\epsilon_{i_1 i_2 \ldots i_p} = \pm 1$ for even (odd) permutations of the indices $1,2,...,p$. If d is the dimension of ϕ, $\phi^{\{p\}}$ has dimension $\binom{d}{p} \equiv d!/p!(d-p)!$. In particular, for SO($2N$), since ϕ_N has dimension $2N$, $\phi_N^{\{p\}}$

has dimension $\binom{2N}{p}$. This is irreducible except when $p = N$, in which case it is reducible to $\Sigma_{\text{odd}} + \Sigma^c_{\text{odd}}$ $(\Sigma_{\text{even}} + \Sigma^c_{\text{even}})$ when N is odd (even). Σ^c_{odd} is the complex conjugate of Σ_{odd}, but Σ^c_{even} and Σ_{even} are real and inequivalent representations of dimension $\frac{1}{2}\binom{2N}{N}$ each. Their n-tuples are shown in Eqs. (3.1) and (3.2). The n-tuples for the remaining irreducible $\phi^{\{p\}}_N$ $(p \neq N)$ are as follows ('$\sim$' means 'is equivalent to')

$$\begin{aligned}
&\phi^{\{p\}}_N \sim \phi^{\{2N-p\}}_N &&: \quad \lambda_i = \delta_{ip}, \quad p = 1,2,\dots N-2,\\
&\phi^{\{N-1\}}_N \sim \phi^{\{N+1\}}_N &&: \quad \lambda_i = \delta_{N-1,i} + \delta_{N,i},\\
&\phi^{\{2N\}}_N &&: \quad \lambda_i = 0, \quad i = 1,2,\dots,N.
\end{aligned}$$

$\phi^{\{p\}}_N$ is the irreducible tensor representation of the (antisymmetrized) product of $p \neq N$ γ's. $\phi^{\{2N\}}_N$ corresponds simply to $\gamma_{\text{FIVE}} = \Pi^{2N}_{\mu=1}\gamma_\mu$.

Having explained the decompositions (3.1) and (3.2), we state a useful identity to get the subgroup content of $\phi^{\{p\}}_N$, $p < N$ [which may be used, in turn, to find that of $\Sigma_{\text{odd}}/\Sigma_{\text{even}}$ by substituting for the subgroup content of σ_N from Eq. (5.1) or (5.2) into Eqs. (3.1) and (3.2)]

$$\begin{aligned}
(\phi_1 \dotplus \phi_2)^{\{p\}} = &= \phi^{\{p\}}_1 \dotplus \phi^{\{p-1\}}_1 \times \phi_2 \dotplus \phi^{\{p-2\}}_1 \times \phi^{\{2\}}_2\\
&\dotplus \dots \dotplus \phi^{\{2\}}_1 \times \phi_2{}^{\{p-2\}} \dotplus \phi_1 \times \phi^{\{p-1\}}_2 \dotplus \phi^{\{p\}}_2.
\end{aligned} \tag{3.3}$$

The identity (3.3) is useful for $\mathrm{SO}(2m) \otimes \mathrm{SO}(2n) \subset \mathrm{SO}(2N)$ because $\phi_N = (\phi_m, 1) \dotplus (1, \phi_n)$ under $\mathrm{SO}(2m) \otimes \mathrm{SO}(2n)$ and therefore

$$\begin{aligned}
\phi_N{}^{\{p\}} = &(\phi^{\{p\}}_m, 1) \dotplus (\phi^{\{p-1\}}_m, \phi_n) \dotplus (\phi^{\{p-2\}}_m, \phi^{\{2\}}_n)\\
&\dotplus \dots \dotplus (\phi^{\{1\}}_m, \phi^{\{p-1\}}_n) \dotplus (1, \phi^{\{p\}}_n).
\end{aligned} \tag{3.4}$$

As stated above, $\phi^{\{p\}}_N$ is irreducible for $p \neq N$ and $(\phi^{\{p-q\}}_m, \phi^{\{q\}}_n)$ is irreducible except when either $p - q = m$ or $q = n$:

$$\begin{aligned}
&(\phi^{\{p-q\}}_m, \phi^{\{q\}}_n) = (\Sigma + \Sigma^c, \phi^{\{q\}}_n), \quad m = p - q,\\
&(\phi^{\{p-q\}}_m, \phi^{\{q\}}_n) = (\phi^{\{p-q\}}_m, \Sigma + \Sigma^c), \quad n = q.
\end{aligned} \tag{3.5}$$

Hence further decomposition of the r.h.s. of Eq. (3.5) may be carried out easily and the subgroup content of $\phi^{\{p\}}_N$ obtained.

For $\mathrm{SU}(N) \subset \mathrm{SO}(2N)$, we use $\phi_N = N + \overline{N}$ under $\mathrm{SU}(N)$. In order to be able to use Eq. (3.3) in this case, we must know how the products taken between fundamental representations of $\mathrm{SU}(N)$ decompose. The is facilitated by writing the products in the form $\psi^{[p]} \times \psi_{[q]}$ and using

$$\psi^{[p]} \times \psi_{[q]} = \psi^{[p]}_{[q]} + \psi^{[p-1]}_{[q-1]} + \cdots + \begin{cases} \psi^{[p-q]} & \text{if } p > q\\ 1 & \text{if } p = q.\\ \psi_{[q-p]} & \text{if } q > p \end{cases} \tag{3.6}$$

To explain the notation, $\psi^{[p]}_{[q]}$ is an irreducible $\mathrm{SU}(N)$ tensor with the property that it has $p(q)$ totally antisymmetric upper (lower) indices. The condition of tracelessness reduces the number of independent components in $\psi^{[p]}_{[q]}$ to

$$\binom{N+1}{p}\binom{N+1}{q}\frac{N+1-p-q}{N+1}.$$

$\psi^{[q]}_{[p]}$ is the complex conjugate of $\psi^{[p]}_{[q]}$ and $\psi^{[q]}_{[p]}$ transforms identically as $\psi^{[N-p]}_{[N-q]}$. This information suffices to work out all the decompositions of products between the fundamental representations of $\mathrm{SU}(N)$ (which correspond to antisymmetrizations of the N-dimensional defining representation). Hence the $\mathrm{SU}(N)$ content of $\phi^{\{p\}}_N$ follows from

$$\begin{aligned}
\phi^{\{p\}}_N &= N^{\{p\}} \dotplus N^{\{p-1\}} \times \overline{N} + N^{\{p-2\}} \times \overline{N}^{\{2\}} \dotplus \dots \dotplus N\\
&\quad \times \overline{N}^{\{p-1\}} \dotplus \overline{N}^{\{p\}}\\
&= \psi^{[p]} \dotplus \psi^{[p-1]} \times \psi_{[1]} \dotplus \psi^{[p-2]} \times \psi_{[2]} \dotplus \dots \dotplus \psi^{[1]}\\
&\quad \times \psi_{[p-1]} \dotplus \psi_{[p]}.
\end{aligned} \tag{3.7}$$

The decompositions and subgroup content, useful in physical applications, are independent of the ensuing discussion.

4. SHIFT ACTION OF SO(2N) AND W.V.R.

Let $E_\gamma (E_{-\gamma} = E^\dagger_\gamma)$ denote the shift operator with its shift action or ladder operation along (antiparallel to) the root γ. In this section, we use the anticommutation relations given by Eq. (1.2) and the w.v.r.

$$\begin{aligned}
&E_{\alpha_i} = b^\dagger_i b_{i+1}, \quad i = 1,2,\dots N-1,\\
&E_{\alpha_N} = b^\dagger_N b^\dagger_{N-1}
\end{aligned} \tag{4.1}$$

for the shift operators, that may well be called simple, to deduce the Cartan matrix (see also Eq. (2.4)) for $\mathrm{SO}(2N)$ given by

$$\begin{aligned}
h^{ij} &= \frac{4(\alpha_i, \alpha_j)}{(\alpha_i, \alpha_i)(\alpha_j, \alpha_j)}\\
&= \left(\begin{array}{cccccccc}
2 & -1 & 0 & & & & & \\
-1 & 2 & -1 & & & & \bigcirc & \\
0 & -1 & 2 & & & & & \\
 & & & \ddots & & & & \\
 & & & & & 2 & -1 & -1\\
 & \bigcirc & & & & -1 & 2 & 0\\
 & & & & & -1 & 0 & 2
\end{array}\right).
\end{aligned} \tag{4.2}$$

The labelling for the simple roots was given in Sec. 3.

For this section and the next, the key result is the following[3,4] shift rule for the Dynkin labels Λ^i of an arbitrary weight Λ:

$$\Lambda^i = p_i - q_i. \tag{4.3}$$

Here the end points p_i and q_i are defined by the condition that $E_{+\alpha_i}(E_{-\alpha_i})$ may act on a weight vector ξ_Λ with weight Λ without annihilating it for a maximum of $q_i(p_i)$ times in succession. R_ϕ, the l.v.s. in which ϕ acts, admits of the direct-sum weight–subspace decomposition, $R_\phi = \Sigma_{\Lambda \in \Delta_\phi} R^\Lambda_\phi$, where Δ_ϕ denotes the complete set of weights of ϕ. By definition, $\xi_\Lambda \in R^\Lambda_\phi$, the weight subspace with weight Λ, and

$$\begin{aligned}
&(E_{+\alpha_i})^{q_i}\xi_\Lambda \neq 0, \quad (E_{-\alpha_i})^{p_i}\xi_\Lambda \neq 0,\\
&(E_{+\alpha_i})^{q_i+1}\xi_\Lambda = (E_{-\alpha_i})^{p_i+1}\xi_\Lambda = 0.
\end{aligned} \tag{4.4}$$

Before being able to use Eq. (4.3), we must state how the simple shift operators and their adjoints, with their w.v.r. given by Eq. (4.1), act on a weight subspace represented suitably in terms of b_i and $b^\dagger_i$. We may consider a pair of l.v.s.: $\mathscr{V}_1$, the (real) l.v.s. of vectors (dimension 2^N).

The space $\mathscr{V}_1$ is spanned by $\Pi_{i=1}^{N} P_i|0\rangle$ where P_i is either 1 or $b_i^\dagger|0\rangle$ is defined by Eq. (1.3); $\mathscr{V}_2$ is the (real) l.v.s. of operators (dimension 2^{2N}).

The space $\mathscr{V}_2$ is the set of all linear transformations in $\mathscr{V}_1$. Such a linear transformation (represented by a $2^N \times 2^N$ matrix on choosing a suitable basis) consists of sum of products of b_i and $b_i^\dagger$. The reducibility of $\mathscr{V}_1$ and $\mathscr{V}_2$ under SO($2N$) is the subject of the discussion in the next section. It suffices here to add that under the action of $E_{\pm\alpha;}$, whose w.v.r. appears in Eq. (4.1),

$$|\xi\rangle \in \mathscr{V}_1 \rightarrow E_{\pm\alpha_i}|\xi\rangle \in \mathscr{V}_1, \qquad (4.5)$$
$$\Omega \in \mathscr{V}_2 \rightarrow [E_{\pm\alpha_i}, \Omega] \in \mathscr{V}_2.$$

For a weight vector represented by $|\xi\rangle \in \mathscr{V}_1$, the endpoints p_i and q_i [see Eq. (4.4)] are determined by

$$(E_{+\alpha_i})^{q_i}|\xi\rangle$$
$$= \underbrace{(E_{+\alpha_i}(E_{+\alpha_i}(E_{+\alpha_i}\,(E_{+\alpha_i}}_{q_i}|\xi\rangle)))) \neq 0,$$
$$\text{but } E_{+\alpha_i}^{q_i+1}|\xi\rangle = 0,$$
$$(E_{-\alpha_i})^{p_i}|\xi\rangle \neq 0, \quad \text{but } (E_{-\alpha_i})^{p_i+1}|\xi\rangle = 0. \qquad (4.6)$$

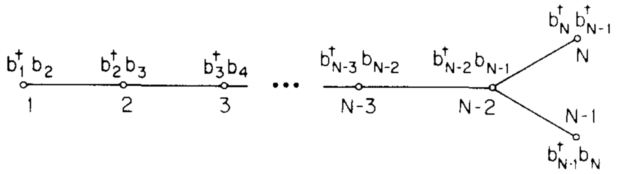
For a weight vector represented by $\Omega \in \mathscr{V}_2$, the end points are determined by

$$\underbrace{[E_{+\alpha_i}[E_{+\alpha_i}[\cdots[E_{+\alpha_i},}_{q_i}\Omega]]] \neq 0$$
$$\underbrace{[E_{-\alpha_i}[E_{-\alpha_i}[\cdots[E_{-\alpha_i},}_{p_i}\Omega]]] \neq 0$$
$$\underbrace{[E_{+\alpha_i},[E_{+\alpha_i}[\cdots[E_{+\alpha_i},}_{q_1+1}\Omega]]] = 0$$
$$= \underbrace{[E_{-\alpha_i}[E_{-\alpha_i}[E_{-\alpha_i}[\cdots[E_{-\alpha_i},}_{p_1+1}\Omega]]]] = 0. \qquad (4.7)$$

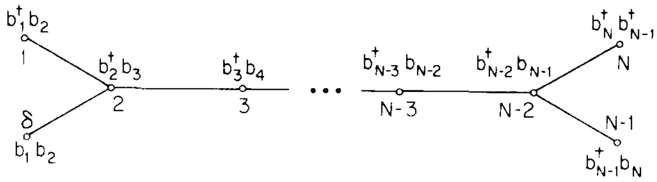
Dynkin labels for the weight, with its end points, determined by Eq. (4.6) or (4.7), as the case may be, are gotten by substituting them in Eq. (4.3).

Let us now complete the derivation of the matrix given by Eq. (4.2). The weight subspace to which $E_{+\alpha_i}$ belongs is characterized by a nonzero weight (of the adjoint) whose Dynkin labels may be read off the ith row of Cartan matrix. Knowing the w.v.r. of $E_{+\alpha_i}$, and the w.v.r. for $E_{\pm\alpha_j}$ given by Eq. (4.1), we may determine the end points p_{ij} and q_{ij} via Eq. (4.7), for the weight associated with $E_{+\alpha_i}$, and verify $h^{ij} = p_{ij} - q_{ij}$. Q.E.D.

Designating w.v.r. for $E_{+\alpha_i}$ against each simple root α_i in the ordinary Dynkin diagram for SO($2N$), we get ('α_i' is labelled by 'i' in Dynkin diagram)



Similarly, for the extended Dynkin diagram[2] we have



where $b_1 b_2$ is the w.v.r. for the E_δ, δ being the shortest root with Dynkin labels (0 — 1 0 0 ... 0). The maximal regular embedding SU(N) ⊂ SO($2N$)(SO($2m$) ⊗ SO($2N-2m$)) corresponds to our deleting the simple root labelled as N (m) in the first (second) ordinary (extended) Dynkin diagram.

TABLE I. Complete set of weights expressed in the e-basis.

Irreducible Subcomponent	Weights
σ_N	$\frac{1}{2}(\pm e_1 \pm e_2 \pm e_3 \pm \cdots \pm e_{N-1} \pm e_N)$ with an odd number of negative signs.
σ_N^c	$\frac{1}{2}(\pm e_1 \pm e_2 \pm e_3 \pm \cdots \pm e_{N-1} \pm e_N)$ with an even number (zero inclusive) of negative signs.
$\phi_N{}^{\{1\}}$ (vectorial)	$\pm e_i$, $(i = 1,2,\ldots,N)$.
$\phi_N{}^{\{2\}}$ (adjoint)	$\pm(e_i + e_j)$, $i<j$; $e_i - e_j$ $(i,j = 1,2\ldots N)$.
$\phi_N{}^{\{3\}}$	$\pm(e_i + e_j + e_k)$, $i<j<k$; $\pm(e_i + e_j - e_k)$, $i<j$; $(i,j,k = 1,2..N)$.
$\phi_N{}^{\{4\}}$	$\pm(e_i + e_j + e_k + e_l)$, $i<j<k<l$; $\pm(e_i + e_j + e_k - e_l)$, $i<j<k$; $\pm(e_i + e_j - e_k - e_l)$, $i<j$ and $k<l$; $(i,j,k,l = 1,2\ldots N)$.
$\vdots$	$\vdots$
$\phi_N{}^{\{k\}}$	$k \leqslant N-1$, quite similarly.
Σ	$e_{a_1} + e_{a_2} + \cdots + e_{a_{N-2k-1}} - e_{b_1} - e_{b_2} - \cdots - e_{b_{2k+1}}$ $a_1 < a_2 < \cdots < a_{N-2k-1}$; $b_1 < b_2 < \cdots < b_{2k+1}$; $k = 0,1,2,\cdots,\left[\frac{N-1}{2}\right]$
Σ^c	$e_{a_1} + e_{a_2} + \cdots + e_{a_{N-2k}} - e_{b_1} - e_{b_2} - \cdots - e_{b_{2k}}$ $a_1 < a_2 < \cdots < a_{N-2k}$; $b_1 < b_2 < \cdots < b_{2k}$; $k = 0,1,2,\ldots,[N/2]$.

TABLE II. Weight vector representatives of highest weights.

Subcomponent in $\mathscr{V}_1/\mathscr{V}_2$	Representative of highest weight	Dynkin labels of highest weight	Highest weight in *e*-basis
σ_N	$b_1^\dagger b_2^\dagger b_3^\dagger \cdots b_{N-1}^\dagger \|0\rangle$	(00⋯010)	$\frac{1}{2}(e_1 + e_2 + e_3 \cdots + e_{N-1} - e_N)$.
σ_N^c	$b_1^\dagger b_2^\dagger b_3^\dagger \cdots b_N^\dagger \|0\rangle$	(00⋯001)	$\frac{1}{2}(e_1 + e_2 + e_3 + \cdots + e_N)$.
$\phi_N^{\{k\}}$	$b_1^\dagger b_2^\dagger \cdots b_k^\dagger$	$(00\cdots 0\overset{k}{1}0\cdots 0)$	$e_1 + e_2 + \cdots + e_k$.
$(1 \leqslant k \leqslant N-2)$			
$\phi_N^{\{N-1\}}$	$b_1^\dagger b_2^\dagger \cdots b_{N-1}^\dagger$	(00⋯011)	$e_1 + e_2 + \cdots + e_{N-1}$.
Σ	$b_1^\dagger b_2^\dagger \cdots b_{N-1}^\dagger b_N$	(00⋯020)	$e_1 + e_2 + e_3 + \cdots e_{N-1} - e_N$.
Σ^c	$b_1^\dagger b_2^\dagger \cdots b_{N-1}^\dagger b_N^\dagger$	(00⋯002)	$e_1 + e_2 + e_3 + \cdots + e_{N-1} + e_N$.
$\phi_N^{\{N+1\}}$	$b_1^\dagger b_2^\dagger \cdots b_{N-1}^\dagger (b_N^\dagger b_N - b_N b_N^\dagger)$	(00⋯011)	$e_1 + e_2 + \cdots + e_{N-1}$.
$\phi^{\{2N-k\}}$	$b_1^\dagger b_2^\dagger \cdots b_k^\dagger \prod_{i=1}^{N-k} (b_{k+1}^\dagger b_{k+1} - b_{k+1} b_{k+1}^\dagger)$	$(00\cdots 0\overset{k}{1}0\cdots 0)$	$e_1 + e_2 + \cdots + e_k$.
$1 \leqslant k \leqslant N-2$			
$\phi_N^{\{2N\}}$	$\prod_{i=1}^{N} (b_i^\dagger b_i - b_i b_i^\dagger)$	(000⋯00)	0.

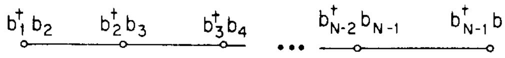
5. SO(2*N*) IRREDUCIBLE SUBCOMPONENTS IN $\mathscr{V}_1$ AND $\mathscr{V}_2$ AND W.V.R.

In the notation of Sec. 3, $\mathscr{V}_1$ and $\mathscr{V}_2$ are completely reducible to $\sigma_N \dotplus \sigma_N^c$ and

$$1 \dotplus \phi_N \dotplus \phi_N^{\{2\}} \dotplus \cdots \dotplus \phi_N^{\{N-1\}} + \Sigma + \Sigma^c$$
$$\dotplus \phi_N^{\{N+1\}} \dotplus \cdots \dotplus \phi_N^{\{2N\}},$$

respectively (we have suppressed the subscripts on Σ and Σ^c which distinguish even N from odd N, since the distinction is irrelevant here). In Table I, we enumerate the complete set of weights of the SO(2N) irreducible subcomponents in $\mathscr{V}_1$ and $\mathscr{V}_2$, facilitated by working in the self-dual basis (Sec. 2). For these irreducible subcomponents, Table II gives the w.v.r. for their highest weights, which yield the w.v.r. for the remaining weight vectors through repeated application of $E_{\pm\alpha_i}$ [Eq. (4.1)] on the w.v.r. of the highest weight. For example, for the vectorial representation, the w.v.r. obtained in this manner are none other than b_i and $b_i^\dagger$, as was to be expected from the opening remarks of Sec. 1. Dynkin labels are readily accessible from a w.v.r. by substitution into Eq. (4.3) the end points determined via Eq. (4.6) [Eq. (4.7)] when the w.v.r. lies in $\mathscr{V}_1[\mathscr{V}_2]$, respectively. Dynkin labels thus obtained for the highest weights are also shown in Table II. The following observation suggests that the self-dual basis is natural for working with w.v.r. When a w.v.r. lies in $\mathscr{V}_2$, the associated weight is given in the self-dual basis by replacing $b_i^\dagger(b_i)$ by $+e_i(-e_i)$ wherever this appears and adding up all. (For replacement, we may pick any one of the terms, if the w.v.r. is the sum of more than one term.) When a w.v.r. lies in $\mathscr{V}_1$ (see the discussion in the following paragraph), we replace P_i by $+e_i/2(-e_i/2)$ when it equals $b_i^\dagger(1)$ in the expression $\Pi_{i=1}^N P_i|0\rangle$ and add up all to get the asssociated weight. The highest weights in the self-dual basis are also shown in Table II.

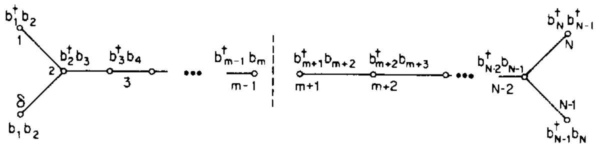
Some further remarks on $\mathscr{V}_1$ are worth adding. A w.v.r. in $\mathscr{V}_1$ is of the form $\Pi_{i=1}^N P_i|0\rangle$, P_i being either 1 or $b_i^\dagger$. Under SO(2N), a 2^{N-1}- dimensional irreducible representation is generated by the w.v.r. $|0\rangle$ and the product of an even number (p to a maximum of $[N]$) of creation operators acting on $|0\rangle$. Let us agree to call this σ_N. The remaining w.v.r. in $\mathscr{V}_1$ transform irreducibly under SO(2N) as σ_N^c. The set of w.v.r., corresponding to a definite number k of creation operators acting on $|0\rangle$, transform irreducibly as $\chi_N^{\{k\}}$ under SU(N), whose shift-action is represented by



as stated at the end of Sec. 4. Here χ_N is an N-dimensional defining representation of SU(N) with n-tuples (1, 0, 0, ..., 0). $x_N^{\{k\}}$ has dimension $\binom{N}{k}$. $\chi_N^{\{N-k\}}$ is a complex conjugate of $x_N^{\{k\}}$. Hence the SU(N) content of the spinorial is

$$\sigma_N = 1 \dotplus x_N^{\{2\}} \dotplus x_N^{\{4\}} \dotplus \ldots \dotplus x_N^{\{[N]\}}$$
$$\sigma_N^c = x_N \dotplus x_N^{\{3\}} \dotplus x_N^{\{5\}} \dotplus \ldots \dotplus x_N^{\{[N+1]-1\}}. \tag{5.1}$$

Similarly, the shift-action of SO(2m) ⊗ SO(2N − 2m) is represented by ($m \neq 1$)



as stated at the end of Sec. 4. Under SO(2m) ⊗ SO(2N − 2m),

$$\sigma_N = (\sigma_m, \sigma_{N-m}) + (\sigma_m^c, \sigma_{N-m}^c),$$
$$\sigma_N^c = (\sigma_m, \sigma_{N-m}^c) + (\sigma_m^c, \sigma_{N-m}). \tag{5.2}$$

The dimension of $\sigma_m(\sigma_{N-m})$, which equals that of $\sigma_m^c(\sigma_{N-m}^c)$, is $2^{m-1}(2^{N-m-1})$. The subgroup content of the spinorial, given by Eq. (5.1) and (5.2), is useful for finding that of Σ and Σ^c (see Sec. 3).

6. PHYSICAL APPLICATIONS I: CONJUGATION

In this section, we examine $\Sigma_{\mu\nu} \to T\Sigma_{\mu\nu}T^{-1} = \pm\Sigma_{\mu\nu}$ under the conjugation induced by a product of k γ-matrices,

(say)

$$T = T_k = \prod_{i=1}^{k} \gamma_{2i}.$$

$\Sigma_{\mu\nu}$ denotes a compact generator of SO(2N). The generators which commute with T_k span the symmetric subalgebra SO(k) ⊗ SO(2N − k) except for $k = 1$ and $k = 2$, when the symmetric subalgebra is SO(2N − 1) and SO(2N − 2) ⊗ U(1) respectively. The remaining generators which anticommute with T_k belong to the orthogonal complementary coset space. The diagonal generators

$$\Sigma_{12}, \Sigma_{34}, \Sigma_{56}, ..., \Sigma_{2k-1,2k}$$

reverse sign under the automorphism thus induced by T_k (the number k is the rank of the coset space). The maximum value (equal to n) of k corresponds to complex conjugation i.e., inversion through the origin of $D_N \simeq$SO(2N)-root-diagram. This automorphism is inner if and only if k is even. When k is even, T_k corresponds to the discrete rotation

$$\exp(\pi\Sigma_{24})\exp(\pi\Sigma_{68})\cdots\exp(\pi\Sigma_{2k-2,2k}).$$

When k is odd, T_k may be expressed as a discrete rotation multiplied by the nontrivial element of the factor group,[5] group of outer automorphisms/group of inner automorphisms (this factor group is of order 2 except for $N = 4$). A convenient choice for this element is T_N (T_{N-1}) when N is odd (even). (Here we do not have in mind the special case of $N = 4$.) For even N, T_N induces an inner automorphism which also corresponds to complex conjugation. Hence we see that all the irreducible representations of SO(4n) are real.

T_k, which is represented as $\hat{T}_k = \Pi_{i=1}^{k}(b_i + b_i^\dagger)$, acts on a w.v.r. as follows.

$$|\xi\rangle \in \mathscr{V}_1 \rightarrow \hat{T}_k |\xi\rangle \in \mathscr{V}_1$$

$$\Omega \in \mathscr{V}_2 \rightarrow \hat{T}_k \Omega \hat{T}_k^{-1} \in \mathscr{V}_2.$$

In the self-dual basis, T_k acts so as to let $e_i \rightarrow -e_i$ for $i \leqslant k$, and $e_i \rightarrow +e_i$ for $N \geqslant i \geqslant k+1$. This may also be described by a diagonal matrix, the conjugation matrix,

$$C_{\text{even}} = \text{diag}\,(\overset{\longleftarrow k \longrightarrow}{-1-1...-1}\overset{\longleftarrow N-k \longrightarrow}{+1+1...+1}),$$

whose square is obviously unity. On transforming to Dynkin basis the form for the conjugation matrix now reads

$$C_s = AC_eA^{-1},$$

where A is given by Eq. (2.9). The subscript S on C_S is for Slansky, whose choice[6] for the charge conjugation matrix in the Dynkin basis corresponds to $C_e = \text{diag}\,(-1\ -1\ -1\ +1\ -1)$ in the self-dual basis, for $N = 5$. Reflection under an automorphism is thus easy to visualize in the latter basis. Under an outer automorphism, $\sigma \rightleftarrows \sigma^c$ and $\Sigma \rightleftarrows \Sigma^c$, whereas each of the remaining irreducible subcomponents in $\sigma \times \sigma$ and $\sigma \times \sigma^c$ is reflected onto itself.

7. PHYSICAL APPLICATIONS II: YUKAWA COUPLINGS

The Clebsch–Gordan (CG) coefficients for the coupling of the operator $\sigma \times \sigma$ or $\sigma \times \sigma^c$ with any one of the irreducible subcomponents in that operator, in particular those for Yukawa couplings, may be systematically computed as discussed below.

For the purpose of illustration, let N be odd. For definiteness, we imagine a theory based on the gauge group SO(2N) in which (left-handed) fermions are assigned to σ_N. Let the set of their w.v.r. be collected in $|Lr\rangle$ wherein to each w.v.r. we associate a fermion field. Here r is a family index. Let $|Rr\rangle$ be the set of w.v.r. obtained from those in $|Lr\rangle$ by the action of $(CP)_{\text{group}} = \Pi_{i=1}^{N}(b_i + b_i^\dagger)$. Furthermore, the fermion fields associated with w.v.r. in $|Rr\rangle$ are the (right-handed) antiparticles of the corresponding fermion fields in $|Lr\rangle$. The group-invariant Yukawa couplings assume the form

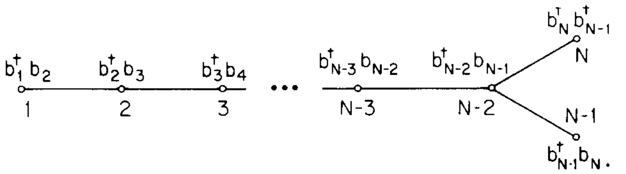
$$\mathscr{L}_{\text{Yukawa}} = \Sigma_{r,s} h_{rs} \langle Rr|\gamma_0 \Phi(\sigma \times \sigma)|Ls\rangle$$
$$+ \text{Hermitian conjugate.}$$

Here γ_0 is required to ensure Lorentz invariance. $\Phi(\sigma \times \sigma)$ is the representative of a set of scalar fields, contained in the decomposition of $\sigma \times \sigma$, which transform irreducibly under SO(2N). The w.v.r. of those fields are obtained from the w.v.r. of the highest weight by repeated action of shift operators (Secs. 3 and 4). The above expression may be regarded as the analog of the conventional expression involving γ-matrices sandwiched between a bilinear of the chiral projections of the spinorial representation. Similarly, for the coupling of the representative $\Phi(\sigma \times \sigma^c)$ contained in the decomposition of $\sigma \times \sigma^c$, we sandwich $\Phi(\sigma \times \sigma^c)$ between $|Lr\rangle$ and $\langle Ls|$, with appropriate change in Lorentz structure.

8. CONCLUSIONS

It is quite convenient to rewrite Clifford algebra in terms of N annihilation operators (and their adjoints) and to work in the self-dual basis (Sec. 2), so long as we are concerned with the group theory of the spinor space $\sigma_N + \sigma_N^c$ and linear transformations in it. Our straightforward conclusions are:

(A) Shift action for SO(2N) may be represented as (Sec. 4)



Here we have given the w.v.r. of $E_{+\alpha_i}$ ($i = 1,2,...,N$); $E_{-\alpha_i} = E_{\alpha_i}^\dagger$. It is sufficient to give w.v.r. of only those shift operators, since w.v.r. for the rest may be derived from them.

(B) The complete set of weights in σ_N, σ_N^c and in irreducible subcomponents of $\sigma_N \times \sigma_N$ and $\sigma_N \times \sigma_N^c$ is shown in Table I.

(C) For those representations, w.v.r. for their highest weights appear in Table II. For the remaining weights, w.v.r. may be developed systematically by repeated application of the shift operators given under (A).

(D) For those representations, a simple method is given in Sec. 3 for finding subgroup content under all maximal regular embeddings of SO(2N), namely SO(2m) ⊗ SO(2N − 2m) and SU(N), where $m \neq 1$.

(E) For those representations, reflection under an automorphism is easy to understand in terms of a diagonal conjugation matrix in the self-dual basis (Sec. 5).

(F) For couplings of the $\sigma_N \times \sigma_N$ and $\sigma_N \times \sigma_N^c$ operators, a systematic way to track all the CG coefficients is alluded to in Sec. 7.

An application to SO(10) based on (E) and (F) is discussed elsewhere.

[1] R. N. Mohapatra and B. Sakita, Phys. Rev. D **21**, 1062 (1980).

[2] E. B. Dynkin, Amer. Math. Soc. Trans. 2, Vol. 6 (1957), p. 111.

[3] E. B. Dynkin, Am. Math. Soc. Trans. 2, Vol. 6 (1957), p. 327.

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

[5] R. Gilmore, *Lie Groups, Lie Algebras and Some of their Applications* (Wiley, New York, 1974).

[6] R. Slansky, Phys. Rep. **79**, 1 (1982); see also R. Slanksy, Proceedings of the First Workshop on Grand Unification (Durham, N.H., 1980).

Cartesian polytensors

Jon Applequist
Department of Biochemistry and Biophysics, Iowa State University, Ames, Iowa 50011

(Received 20 July 1982; accepted for publication 12 November 1982)

A Cartesian polytensor is defined as a set of Cartesian tensors in a sequence of increasing rank. A matrix formulation of polytensors is given to express arrays of direct tensor products and series of tensor contractions in concise form. The transformation of a polytensor under rotation of coordinate axes is shown to be accomplished by means of an orthogonal matrix. The special properties of compressed polytensors, composed of totally symmetric tensors with redundant components deleted, are demonstrated. The use of polytensors is illustrated by an application to the problem of interactions among polarizable electric charge distributions.

PACS numbers: 02.10.Sp

I. INTRODUCTION

Many physical problems involving Cartesian tensors make use of series expressions whose terms are products of tensors of increasing rank. Examples are the multipole expansions of electric,[1] magnetic,[2] and gravitational[3] potentials. Similar expansions may arise whenever a Taylor series is used to express a function of tensor variables. In a study of interactions of electric multipole systems, I have found it advantageous to treat a set of Cartesian tensors of increasing rank as a single entity, which I will call here a Cartesian polytensor. Polytensors, like tensors, are subject to the operations of addition and scalar multiplication, and these operations obey the usual associative, distributive, and commutative laws. The utility of polytensors depends further on the concise manner in which they permit treatment of direct tensor products, tensor contractions, and transformations of tensors under rotation of coordinate axes. The purpose of this paper is to develop these elementary properties of polytensors. In the last section the use of polytensors is illustrated by an application to the multipole treatment of interactions among polarizable electric charge distributions.

II. CARTESIAN TENSORS—DEFINITIONS

1. An nth rank Cartesian tensor will be denoted by a boldface symbol $\mathbf{A}^{(n)}$ or the corresponding component notation $A^{(n)}_{\alpha_1\cdots\alpha_n}$, where each α_i takes the values 1, 2, or 3, corresponding to the Cartesian axes. A Cartesian tensor is defined in the usual fashion[4] by the manner in which it transforms under a rotation of coordinate axes. The components of the tensor comprise a column matrix, each element of which is identified by the set of indices $\alpha_1\cdots\alpha_n$. The *array of index sets* is the array of sets $\alpha_1\cdots\alpha_n$ in some defined order. The order will be called *canonical* if, on proceeding through the array, α_1 varies through the values 1,2,3 more rapidly than α_2, which varies more rapidly than α_3, etc. The order will be called *anticanonical* if α_n varies more rapidly than α_{n-1}, etc. The canonical order is the order in which array elements are usually stored in a computer memory. For example, the following are arrays of index sets for $n = 3$:

canonical: 111,211,311,121,221,321,...,333,
anticanonical: 111,112,113,121,122,123,...,333.

2. A Cartesian tensor of rank $m + n$ may be represented as a tensor of *subdivided rank* (m,n), in which m and n are the *rank indices*. For example $\mathbf{B}^{(m,n)}$ is of rank $m + n$, and its components are written $B^{(m,n)}_{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_n}$. The array of components comprise a rectangular matrix whose rows are indexed by the set $\alpha_1\cdots\alpha_m$ and whose columns are indexed by the set $\beta_1\cdots\beta_n$. The rank of a tensor may be subdivided into any convenient number of rank indices. Tensors $\mathbf{B}^{(0,n)}$ and $\mathbf{B}^{(n,0)}$ are nth rank tensors represented as row and column matrices, respectively.

3. The tensor $\mathbf{A}^{(n)}$ is said to be *totally symmetric* if $A^{(n)}_{\alpha_1\cdots\alpha_n}$ is unchanged on any permutation of $\alpha_1\cdots\alpha_n$. The matrix form of such a tensor is identical in canonical and anticanonical order. A tensor $\mathbf{B}^{(m,n)}$ is said to be *totally symmetric in the component indices of rank index m* if it is invariant on any permutation of the index set $\alpha_1\cdots\alpha_m$ corresponding to that rank index.

4. The *direct product* of two tensors such as $\mathbf{A}^{(l)}$ and $\mathbf{B}^{(m,n)}$ is a tensor of rank $l + m + n$, and may be represented in subdivided form by

$$\mathbf{C}^{(l,m,n)} = \mathbf{A}^{(l)}\mathbf{B}^{(m,n)}\,, \tag{1}$$

or

$$C^{(l,m,n)}_{\alpha_1\cdots\alpha_l\beta_1\cdots\beta_m\gamma_1\cdots\gamma_n} = A^{(l)}_{\alpha_1\cdots\alpha_l}B^{(m,n)}_{\beta_1\cdots\beta_m\gamma_1\cdots\gamma_n}\,. \tag{2}$$

5. An *n-fold contraction* is denoted by the symbol $\cdot n\cdot$, as in

$$\mathbf{B}^{(m,n)}\cdot n\cdot\mathbf{A}^{(n)}\,, \tag{3}$$

which is equivalent to the form

$$\sum_{[\beta]} B^{(m,n)}_{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_n}A^{(n)}_{\beta_n\cdots\beta_1}\,, \tag{4}$$

where the sum over $[\beta]$ denotes a sum over the complete array of sets $\beta_1\cdots\beta_n$. (In the present development, the common convention of implied summation over repeated indices will not be used, as it is important to indicate explicitly whether summation is over a complete array of index sets or over a "compressed" array, as will be seen below.) The contraction is equivalent to a matrix product if the column indices $\beta_1\cdots\beta_n$ of the first factor are in canonical order and the row indices $\beta_n\cdots\beta_1$ of the second factor are in anticanonical order. If both tensors are totally symmetric in the component indices of rank index n, then the same order may be

0022-2488/83/040736-06$02.50

used in both factors.

6. The *compressed form* of $\mathbf{A}^{(n)}$, denoted $\bar{\mathbf{A}}^{(n)}$, consists of the subset of components of the tensor in which no index set $\alpha_1 \cdots \alpha_n$ is a permutation of another set. If $\mathbf{A}^{(n)}$ is totally symmetric, the compressed form contains all of the information of the complete tensor in $(n+1)(n+2)/2$ components, as opposed to 3^n components in the complete tensor. Clearly, compression has significance only if $n \geqslant 2$. The canonical array of index sets of the compressed tensor is derived from that of the complete tensor by deleting any index set which does not satisfy the condition $\alpha_1 \geqslant \alpha_2 \geqslant \cdots \geqslant \alpha_n$; e.g., for $n = 3$ the compressed array in canonical order is 111, 211, 311, 221, 321, 331, 222, 322, 332, 333. The compressed form of a tensor of subdivided rank, e.g., $\bar{\mathbf{B}}^{(m,n)}$, consists of the subset of components of the tensor in which no index set $\alpha_1 \cdots \alpha_m$ is a permutation of another set corresponding to rank index m and no index set $\beta_1 \cdots \beta_n$ is a permutation of another set corresponding to rank index n. (We will not consider here tensors which are compressed in some subdivisions of their indices but not in others.)

III. CARTESIAN POLYTENSORS—DEFINITIONS

7. A *Cartesian polytensor* is a set of Cartesian tensors in a sequence of increasing rank. A polytensor of *first degree* is represented by the column matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{(0)} \\ \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \end{bmatrix} . \tag{5}$$

A_i will denote the element of the ith row of $\mathbf{A}$. One may choose to begin the sequence at a rank higher than the scalar $\mathbf{A}^{(0)}$ in special cases. The sequence of tensors is of indefinite length, though for practical purposes it will be desirable to truncate the sequence at some specified rank.

8. A polytensor of *second degree* is represented by a rectangular matrix whose blocks are tensors whose rank is subdivided into two indices; e.g.,

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}^{(0,0)} & \mathbf{B}^{(0,1)} & \mathbf{B}^{(0,2)} \cdots \\ \mathbf{B}^{(1,0)} & \mathbf{B}^{(1,1)} & \mathbf{B}^{(1,2)} \cdots \\ \vdots & & \end{bmatrix} . \tag{6}$$

B_{ij} will denote the element of the ith row and jth column of $\mathbf{B}$.

9. A polytensor of Nth degree is a sequence of tensors whose rank is subdivided into N rank indices. If $\mathbf{C}$ is a polytensor of Nth degree, its general element is $C_{ijk\cdots}$, where i spans the component indices of the first-rank index, j spans the component indices of the second-rank index, and so on.

10. The *direct product* of two polytensors $\mathbf{A}$ and $\mathbf{B}$ is denoted $\mathbf{AB}$, and is a polytensor composed of all of the possible direct products of one tensor from $\mathbf{A}$ and one tensor from $\mathbf{B}$. It is convenient, but not essential, to define the degree of $\mathbf{AB}$ as the sum of the degrees of $\mathbf{A}$ and $\mathbf{B}$. For example, if $\mathbf{A}$ is of first degree and $\mathbf{B}$ is of the second degree, then $\mathbf{AB}$ is of third degree, and the general element of the latter may be denoted by $(\mathbf{AB})_{ijk} = A_i B_{jk}$. If $\mathbf{A}$ and $\mathbf{B}$ are both of first degree, then the matrix product $\mathbf{AB}^T$ (where T denotes transpose) represents the direct product as a square matrix, a polytensor of second degree.

11. *Contractions* of polytensors are defined by analogy with ordinary tensor contractions. Let $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ be polytensors of degree 1, 2, and 3, respectively. The following represent single contractions:

$$(\mathbf{A} \cdot \mathbf{B})_i = \sum_j A_j B_{ji} , \tag{7}$$

$$(\mathbf{B} \cdot \mathbf{C})_{ikl} = \sum_j B_{ij} C_{jkl} . \tag{8}$$

We adopt the following convention: The tensor components corresponding to a repeated index [j in Eqs. (7) and (8)] are in canonical order in the first factor and anticanonical order in the second factor. Thus the contraction in Eq. (7), for example, is the polytensor composed of the tensors $\mathbf{A}^{(n)} \cdot n \cdot \mathbf{B}^{(n,m)}$. A double contraction is represented by

$$(\mathbf{B} : \mathbf{C})_k = \sum_i \sum_j B_{ij} C_{jik} . \tag{9}$$

Higher multiple contractions are defined in a corresponding manner. Only single contractions involving polytensors of first or second degree can be equated to ordinary matrix products.

12. A *compressed polytensor* is a sequence of compressed tensors, and is denoted by a bar over the symbol; e.g.,

$$\bar{\mathbf{A}} = \begin{bmatrix} \bar{\mathbf{A}}^{(0)} \\ \bar{\mathbf{A}}^{(1)} \\ \bar{\mathbf{A}}^{(2)} \\ \vdots \end{bmatrix} , \tag{10}$$

$$\bar{\mathbf{B}} = \begin{bmatrix} \bar{\mathbf{B}}^{(0,0)} & \bar{\mathbf{B}}^{(0,1)} & \bar{\mathbf{B}}^{(0,2)} \cdots \\ \bar{\mathbf{B}}^{(1,0)} & \bar{\mathbf{B}}^{(1,1)} & \bar{\mathbf{B}}^{(1,2)} \cdots \\ \vdots & & \end{bmatrix} . \tag{11}$$

If a uniform convention for ordering index arrays is followed, then $\bar{\mathbf{B}}$ is obtained by deleting certain rows and columns of $\mathbf{B}$, and is thus a minor of $\mathbf{B}$.

13. The use of polytensor expressions will be illustrated by some examples. A scalar C may occur as a series of tensor contractions,

$$C = \sum_{n=0}^{\infty} \mathbf{B}^{(n)} \cdot n \cdot \mathbf{A}^{(n)} , \tag{12}$$

or, more simply,

$$C = \mathbf{B} \cdot \mathbf{A} , \tag{13}$$

where $\mathbf{A}$ and $\mathbf{B}$ are first degree polytensors. Similarly, a tensor $\mathbf{C}^{(m)}$ may occur as a series,

$$\mathbf{C}^{(m)} = \sum_{n=0}^{\infty} \mathbf{B}^{(m,n)} \cdot n \cdot \mathbf{A}^{(n)} , \quad m = 0,1,2 \cdots, \tag{14}$$

or, more simply,

$$\mathbf{C} = \mathbf{B} \cdot \mathbf{A} , \tag{15}$$

where $\mathbf{A}$ and $\mathbf{C}$ are first degree polytensors and $\mathbf{B}$ is a second degree polytensor. $\mathbf{C}^{(m)}$ might have a more complex series form, e.g.,

$$\mathbf{C}^{(m)} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathbf{B}^{(m,i,j)} \cdot (i+j) \cdot \mathbf{A}^{(j)} \mathbf{A}^{(i)} , \quad m = 0,1,2,\cdots, \tag{16}$$

which is more easily written

$$\mathbf{C} = \mathbf{B}{:}\mathbf{A}\mathbf{A}\,, \tag{17}$$

where **B** is a third degree polytensor.

IV. TRANSFORMATION OF POLYTENSORS

14. It will be shown that a first degree polytensor transforms under a rotation of coordinate axes in a manner analogous to that of a Cartesian vector. A vector $\mathbf{v}$ in coordinate system S becomes a vector $\mathbf{v}'$ in a rotated system S', where

$$\mathbf{v}' = \boldsymbol{\lambda}\mathbf{v}\,, \tag{18}$$

and $\boldsymbol{\lambda}$ is an orthogonal matrix whose elements $\lambda_{\alpha\beta}$ are direction cosines of axes α in system S' with respect to axes β in system S.[4]

Theorem: A first degree polytensor **A** in coordinate system S is transformed into polytensor $\mathbf{A}'$ in rotated system S' according to

$$\mathbf{A}' = \mathbf{\Lambda}\mathbf{A}\,, \tag{19}$$

where $\mathbf{\Lambda}$ is an orthogonal matrix. (A matrix product is implied on the right side. Since $\mathbf{\Lambda}$ is not a polytensor, there should be no confusion with a direct product.)

Proof: The individual tensors in **A** transform according to[4]

$$A^{(n)\prime}_{\alpha_1\cdots\alpha_n} = \sum_{[\beta]} \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} A^{(n)}_{\beta_1\cdots\beta_n}\,, \tag{20}$$

where

$$\Lambda^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} = \lambda_{\alpha_1\beta_1}\lambda_{\alpha_2\beta_2}\cdots\lambda_{\alpha_n\beta_n}\,. \tag{21}$$

We regard $\mathbf{\Lambda}^{(n)}$ as square matrix whose rows are indexed by $\alpha_1\cdots\alpha_n$ in canonical order and whose columns are indexed by $\beta_1\cdots\beta_n$ in canonical order. We define the block diagonal matrix

$$\mathbf{\Lambda} = (1 \;\; \mathbf{\Lambda}^{(1)} \;\; \mathbf{\Lambda}^{(2)} \;\; \cdots)^{\mathrm{D}}\,, \tag{22}$$

where the superscript D indicates that the matrix contains the indicated blocks along the principal diagonal and zeros elsewhere. It is evident from the form of **A** in Eq. (5) that $\mathbf{\Lambda}$ so defined satisfies Eq. (19). It remains to show that $\mathbf{\Lambda}$ is orthogonal. The scalar product of any two columns of a particular block $\mathbf{\Lambda}^{(n)}$ is

$$\begin{aligned}\sum_{[\alpha]} \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\gamma_1\cdots\gamma_n} &= \sum_{[\alpha]} (\lambda_{\alpha_1\beta_1}\lambda_{\alpha_1\gamma_1})\cdots(\lambda_{\alpha_n\beta_n}\lambda_{\alpha_n\gamma_n}) \\ &= \delta_{\beta_1\gamma_1}\cdots\delta_{\beta_n\gamma_n}\,,\end{aligned} \tag{23}$$

where $\delta_{\beta\gamma}$ is the Kronecker delta and the last equality holds by virtue of the orthogonality of $\boldsymbol{\lambda}$. Hence $\mathbf{\Lambda}^{(n)}$ obeys the orthogonality relation

$$\mathbf{\Lambda}^{(n)T}\mathbf{\Lambda}^{(n)} = \mathbf{I}\,, \tag{24}$$

where $\mathbf{I}$ is the identity matrix. From Eqs. (22) and (24) one obtains the orthogonality relation

$$\mathbf{\Lambda}^T\mathbf{\Lambda} = \mathbf{I}\,. \tag{25}$$

15. The following theorem illustrates the fact that higher degree polytensors transform under rotation of coordinate axes in a manner analogous to the transformation of higher rank tensors.

Theorem: A second degree polytensor **B** in coordinate system S is transformed into polytensor $\mathbf{B}'$ in rotated system S' according to

$$\mathbf{B}' = \mathbf{\Lambda}\mathbf{B}\mathbf{\Lambda}^T\,. \tag{26}$$

Proof: A tensor $\mathbf{B}^{(m,n)}$ transforms in the same manner as a direct product $\mathbf{A}^{(m)}\mathbf{C}^{(n)}$, by definition of Cartesian tensors; i.e.,

$$\begin{aligned}B^{(m,n)\prime}_{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_n} = \sum_{[\gamma]}\sum_{[\delta]} &\Lambda^{(m)}_{\alpha_1\cdots\alpha_m\gamma_1\cdots\gamma_m} \Lambda^{(n)}_{\beta_1\cdots\beta_n\delta_1\cdots\delta_n} \\ &\times B^{(m,n)}_{\gamma_1\cdots\gamma_m\delta_1\cdots\delta_n}\,.\end{aligned} \tag{27}$$

Equation (26) follows from Eqs. (6), (22), and (27).

V. TOTALLY SYMMETRIC TENSORS

16. Let $\mathbf{A}^{(n)}$ and $\mathbf{B}^{(m,n)}$ be totally symmetric in the component indices of each of their rank indices. Let $\mathbf{C}^{(m)}$ be the totally symmetric tensor

$$\mathbf{C}^{(m)} = \mathbf{B}^{(m,n)}\cdot n\cdot\mathbf{A}^{(n)}\,. \tag{28}$$

Let

$$g(\beta_1\cdots\beta_n) = n!/n_1!n_2!n_3!\,, \tag{29}$$

where n_i is the number of times i appears in the set $\beta_1\cdots\beta_n$. Define a diagonal matrix of order $(n+1)(n+2)/2$,

$$\mathbf{g}^{(n)} = [g(11\cdots1)\,g(21\cdots1)\cdots g(33\cdots3)]^{\mathrm{D}}\,, \tag{30}$$

where indices span the compressed canonical array. For example,

$$\begin{aligned}\mathbf{g}^{(1)} &= (1\;1\;1)^{\mathrm{D}}\,, \\ \mathbf{g}^{(2)} &= (1\;2\;2\;1\;2\;1)^{\mathrm{D}}\,, \\ \mathbf{g}^{(3)} &= (1\;3\;3\;3\;6\;3\;1\;3\;3\;1)^{\mathrm{D}}\,.\end{aligned} \tag{31}$$

Then we have the following:

Theorem: The compressed form of the contraction defined by Eq. (28) is given by the matrix product

$$\overline{\mathbf{C}}^{(m)} = \overline{\mathbf{B}}^{(m,n)}\mathbf{g}^{(n)}\overline{\mathbf{A}}^{(n)}\,. \tag{32}$$

Proof: Each component of $\mathbf{C}^{(m)}$ is represented in the contraction on the right side of Eq. (28) by a sum of 3^n terms corresponding to the complete array of index sets $\beta_1\cdots\beta_n$. Because of the given symmetry, each term is repeated $g(\beta_1\cdots\beta_n)$ times in the sum. The matrix product in Eq. (32) accomplishes the same result by expressing each component of $\overline{\mathbf{C}}^{(m)}$ as a sum of $(n+1)(n+2)/2$ terms, each multiplied by the appropriate $g(\beta_1\cdots\beta_n)$.

17. Let **A**, **B**, and **C** be polytensors composed of the corresponding totally symmetric tensors in item 16; i.e.,

$$\mathbf{C} = \mathbf{B}\cdot\mathbf{A}\,. \tag{33}$$

Let g be the diagonal matrix

$$\mathbf{g} = (1\;\mathbf{g}^{(1)}\;\mathbf{g}^{(2)}\cdots)^{\mathrm{D}}\,. \tag{34}$$

Theorem: The compressed form of the polytensor contraction defined by Eq. (33) is given by the matrix product

$$\overline{\mathbf{C}} = \overline{\mathbf{B}}\mathbf{g}\overline{\mathbf{A}}\,. \tag{35}$$

Proof: The theorem follows directly from Eq. (32).

The importance of this theorem will be appreciated from the fact that a first degree polytensor truncated at rank

t contains $(3^{t+1}-1)/2$ components in complete form and $(t+1)(t+2)(t+3)/6$ components in compressed form. Thus Eq. (35) represents an operation with substantially smaller matrices than Eq. (33), and the matrix size increases much less rapidly with t.

18. It is worth summarizing some important properties of the second degree polytensor $\mathbf{B}$ defined in item 17. If $\mathbf{B}$ is truncated at rank t, it is a square matrix of order $(3^{t+1}-1)/2$. The matrix is symmetric if and only if $\mathbf{B}^{(n,m)} = \mathbf{B}^{(m,n)T}$, a condition which is not implied by the given permutation symmetry. The matrix rank of $\mathbf{B}$ is limited by the following theorem.

Theorem: If $\mathbf{B}$ is truncated at tensor rank t, its matrix rank is less than or equal to $(t+1)(t+2)(t+3)/6$.

Proof: the theorem is trivial for $t=0$ or 1. For $t>1$, $\mathbf{B}$ is the largest minor of $\mathbf{B}$ which does not, in general, contain repeated rows or columns and which may therefore be nonsingular. Thus the upper limit on the rank of $\mathbf{B}$ is $(t+1)(t+2)(t+3)/6$, the order of $\overline{\mathbf{B}}$.

The main consequence of the theorem is that $\mathbf{B}^{-1}$ does not exist for $t>1$, while $\overline{\mathbf{B}}$ has an inverse except in special cases where its rank is less than its order.

VI. TRANSFORMATION OF COMPRESSED TENSORS

19. We will require the following:

Lemma: If $X_{\alpha_1\cdots\alpha_n}$ is any function of the indices $\alpha_1\cdots\alpha_n$, then

$$\sum_{[\alpha]} X_{\alpha_1\cdots\alpha_n} = \sum_{(\alpha)} \sum_{N\{\alpha\}} X_{\alpha_1\cdots\alpha_n}\,, \tag{36}$$

where the sum over (α) denotes the sum over the compressed array of index sets, and the sum over $N\{\alpha\}$ denotes the sum over all distinguishable permutations of $\alpha_1\cdots\alpha_n$ when numerical values are assigned.

Proof: The sum over $N\{\alpha\}$ generates from each member $\alpha_1\cdots\alpha_n$ of the compressed array the deleted members of the complete array. Thus both sides of Eq. (36) are sums over the complete array of index sets.

20. We define a transformation matrix $\boldsymbol{\Gamma}^{(n)}$:

$$\Gamma^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} = \sum_{N\{\beta\}} \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n}\,, \tag{37}$$

where $\alpha_1\cdots\alpha_n$ and $\beta_1\cdots\beta_n$ span only the compressed arrays.

Theorem: A compressed totally symmetric tensor $\overline{\mathbf{A}}^{(n)}$ in coordinate system S is transformed into compressed tensor $\overline{\mathbf{A}}^{(n)\prime}$ in system S' according to

$$\overline{\mathbf{A}}^{(n)\prime} = \boldsymbol{\Gamma}^{(n)}\overline{\mathbf{A}}^{(n)}\,. \tag{38}$$

Proof: The transformation of the complete tensor is, in consequence of Eqs. (20) and (36),

$$A^{(n)\prime}_{\alpha_1\cdots\alpha_n} = \sum_{(\beta)} \sum_{N\{\beta\}} \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} A^{(n)}_{\beta_1\cdots\beta_n}\,. \tag{39}$$

Equation (38) follows by insertion of Eq. (37) into Eq. (39).

21. An example will suffice to show that a compressed tensor of subdivided rank transforms in the same manner as a direct product of compressed tensors of appropriate rank. Let $\mathbf{B}^{(m,n)}$ be totally symmetric in the component indices of both rank indices. From Eqs. (27), (36), and (37) the transform of the compressed form is

$$\begin{aligned}\overline{B}^{(m,n)\prime}_{\alpha_1\cdots\alpha_m\beta_1\cdots\beta_n} = \sum_{(\gamma)} \sum_{(\delta)} &\Gamma^{(m)}_{\alpha_1\cdots\alpha_m\gamma_1\cdots\gamma_m} \Gamma^{(n)}_{\beta_1\cdots\beta_n\delta_1\cdots\delta_n}\\ &\times \overline{B}^{(m,n)}_{\gamma_1\cdots\gamma_m\delta_1\cdots\delta_n}\,,\end{aligned} \tag{40}$$

which proves the assertion for this case.

22. The inverse transformation of a compressed tensor, accomplished by means of the matrix $\boldsymbol{\Gamma}^{(n)-1}$, is less easily obtained than that of the complete tensor, since $\boldsymbol{\Gamma}^{(n)}$ is not an orthogonal matrix. However, a simple form for the inverse matrix can be obtained, as will be demonstrated with the help of certain lemmas to be proven in this section. In what follows, a sum over $S\{\alpha\}$ will denote a sum over all permutations of the symbols $\alpha_1\cdots\alpha_n$, regardless of their numerical values.

Lemma: If $X_{\alpha_1\cdots\alpha_n}$ is any function of the indices $\alpha_1\cdots\alpha_n$, then

$$g(\alpha_1\cdots\alpha_n) \sum_{S\{\alpha\}} X_{\alpha_1\cdots\alpha_n} = n! \sum_{N\{\alpha\}} X_{\alpha_1\cdots\alpha_n}\,. \tag{41}$$

Proof: Let the set $\alpha_1\cdots\alpha_n$ contain n_i i's ($i=1,2,3$). Each term in the sum on the right side is repeated $n_1!n_2!n_3!$ times in the sum on the left side. Thus Eq. (41) follows from Eq. (29).

For the following lemmas we define the quantity $Y^{(n)}_{(\alpha,\beta,\gamma)}$ in terms of the transformation matrix $\boldsymbol{\Lambda}^{(n)}$:

$$Y^{(n)}_{(\alpha,\beta,\gamma)} = \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} \Lambda^{(n)}_{\alpha_1\cdots\alpha_n\gamma_1\cdots\gamma_n}\,. \tag{42}$$

Lemma: The transformation matrix defined by Eq. (21) obeys the relation

$$\sum_{S\{\alpha\}} \sum_{S\{\beta\}} Y^{(n)}_{(\alpha,\beta,\gamma)} = \sum_{S\{\beta\}} \sum_{S\{\gamma\}} Y^{(n)}_{(\alpha,\beta,\gamma)}\,. \tag{43}$$

Proof: From Eqs. (21) and (42),

$$Y^{(n)}_{(\alpha,\beta,\gamma)} = \lambda_{\alpha_1\beta_1}\cdots\lambda_{\alpha_n\beta_n}\lambda_{\alpha_1\gamma_1}\cdots\lambda_{\alpha_n\gamma_n}\,. \tag{44}$$

Let $p_i(\alpha)$ be the ith permutation of the set $\alpha_1\cdots\alpha_n$, with $i=1,\ldots,n!$; and let $\{p_i(\alpha), p_j(\beta)\}$ denote the set of pairs $(\alpha_1,\beta_1),\ldots,(\alpha_n,\beta_n)$ from the two permuted sets. From Eq. (44) it is seen that each term on the left side of Eq. (43) depends only on the sets $\{p_i(\alpha), p_j(\beta)\}$ and $\{p_i(\alpha), p_1(\gamma)\}$, where p_1 denotes the unpermuted set. Likewise, each term on the right side of Eq. (43) depends only on the sets $\{p_1(\alpha), p_k(\beta)\}$ and $\{p_1(\alpha), p_m(\gamma)\}$. For the two sides of Eq. (43) to be equal, it is sufficient that, there be a one-to-one correspondence between pairs (i,j) and (k,m) such that $\{p_i(\alpha), p_j(\beta)\} = \{p_1(\alpha), p_k(\beta)\}$ and $\{p_i(\alpha), p_1(\gamma)\} = \{p_1(\alpha), p_m(\gamma)\}$. That this condition is satisfied follows from the definition of the sets.

Lemma: The transformation matrix defined by Eq. (21) obeys the relation

$$g(\gamma_1\cdots\gamma_n) \sum_{N\{\alpha\}} \sum_{N\{\beta\}} Y^{(n)}_{(\alpha,\beta,\gamma)} = g(\alpha_1\cdots\alpha_n) \sum_{N\{\beta\}} \sum_{N\{\gamma\}} Y^{(n)}_{(\alpha,\beta,\gamma)}\,. \tag{45}$$

Proof: Eq. (45) follows from Eq. (43) by application of Eq. (41) to each of the sums over $S\{\alpha\}$, $S\{\beta\}$, and $S\{\gamma\}$.

We now arrive at the major theorem of this section.

Theorem: The inverse of the transformation matrix defined by Eq. (37) is given by

$$\boldsymbol{\Gamma}^{(n)-1} = \mathbf{g}^{(n)-1}\boldsymbol{\Gamma}^{(n)T}\mathbf{g}^{(n)}\,. \tag{46}$$

Proof: Define the matrix $\mathbf{Z}^{(n)}$ by

$$\mathbf{Z}^{(n)} = \mathbf{g}^{(n)-1}\mathbf{\Gamma}^{(n)T}\mathbf{g}^{(n)}\mathbf{\Gamma}^{(n)} . \tag{47}$$

The elements of $\mathbf{g}^{(n)}$ are of the form

$$g^{(n)}_{\alpha_1\cdots\alpha_n\beta_1\cdots\beta_n} = \delta_{\alpha_1\beta_1}\cdots\delta_{\alpha_n\beta_n} g(\alpha_1\cdots\alpha_n) , \tag{48}$$

where $\alpha_1\cdots\alpha_n$ and $\beta_1\cdots\beta_n$ each span the compressed arrays. Inserting Eqs. (37) and (48) into (47), one obtains

$$Z^{(n)}_{\alpha_1\cdots\alpha_n\epsilon_1\cdots\epsilon_n} = \sum_{(\delta)} g(\alpha_1\cdots\alpha_n)^{-1} g(\delta_1\cdots\delta_n) \sum_{N\{\alpha\}} \sum_{N\{\epsilon\}} Y^{(n)}_{(\delta,\alpha,\epsilon)} . \tag{49}$$

Equation (49) is transformed as follows by applying Eqs. (45), (36), and (23) in turn:

$$Z^{(n)}_{\alpha_1\cdots\alpha_n\epsilon_1\cdots\epsilon_n} = \sum_{(\delta)} \sum_{N\{\delta\}} \sum_{N\{\epsilon\}} Y^{(n)}_{(\delta,\alpha,\epsilon)} \tag{50}$$

$$= \sum_{N\{\epsilon\}} \sum_{[\delta]} Y^{(n)}_{(\delta,\alpha,\epsilon)} \tag{51}$$

$$= \sum_{N\{\epsilon\}} \delta_{\alpha_1\epsilon_1}\cdots\delta_{\alpha_n\epsilon_n} . \tag{52}$$

Since the index sets span only the compressed arrays, only the unpermuted set $\epsilon_1\cdots\epsilon_n$ can be identical to $\alpha_1\cdots\alpha_n$. Hence the only surviving term in Eq. (52) is

$$Z^{(n)}_{\alpha_1\cdots\alpha_n\epsilon_1\cdots\epsilon_n} = \delta_{\alpha_1\epsilon_1}\cdots\delta_{\alpha_n\epsilon_n} . \tag{53}$$

That is, $\mathbf{Z}^{(n)}$ is the identity matrix. Thus Eq. (46) follows from Eq. (47).

VII. TRANSFORMATION OF COMPRESSED POLYTENSORS

23. The preceding results on compressed tensors lead directly to a number of theorems on the transformation of the corresponding compressed polytensors. For this purpose we define a transformation matrix $\mathbf{\Gamma}$ in the block diagonal form

$$\mathbf{\Gamma} = (1\ \mathbf{\Gamma}^{(1)}\ \mathbf{\Gamma}^{(2)} \cdots)^{\mathrm{D}} . \tag{54}$$

24. **Theorem**: A compressed first degree polytensor $\overline{\mathbf{A}}$ in coordinate system S is transformed into polytensor $\overline{\mathbf{A}}'$ in rotated system S' according to

$$\overline{\mathbf{A}}' = \mathbf{\Gamma}\overline{\mathbf{A}} . \tag{55}$$

Proof: The theorem follows from Eq. (38).

25. **Theorem:** A compressed second degree polytensor $\overline{\mathbf{B}}$ in coordinate system S is transformed into polytensor $\overline{\mathbf{B}}'$ in rotated system S' according to

$$\overline{\mathbf{B}}' = \mathbf{\Gamma}\overline{\mathbf{B}}\mathbf{\Gamma}^T . \tag{56}$$

Proof: The theorem follows from Eq. (40).

26. **Theorem:** The inverse of the transformation matrix defined by Eqs. (37) and (54) is given by

$$\mathbf{\Gamma}^{-1} = \mathbf{g}^{-1}\mathbf{\Gamma}^T\mathbf{g} , \tag{57}$$

where $\mathbf{g}$ is defined by Eq. (34).

Proof:The theorem follows from Eq. (46).

27. Let $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ be compressed polytensors of first and second degree, respectively. Let $\overline{\mathbf{C}}$ be their contraction as in Eq. (35).

Theorem: The transform $\overline{\mathbf{C}}'$ of the contraction $\overline{\mathbf{C}}$ is equal to the contraction of the transforms $\overline{\mathbf{A}}'$ and $\overline{\mathbf{B}}'$; i.e.,

$$\overline{\mathbf{C}}' = \overline{\mathbf{B}}'\mathbf{g}\overline{\mathbf{A}}' . \tag{58}$$

Proof: The right side of Eq. (58) becomes, on substitution of Eqs. (55), (56), and (57), in turn,

$$\overline{\mathbf{B}}'\mathbf{g}\overline{\mathbf{A}}' = \mathbf{\Gamma}\overline{\mathbf{B}}\mathbf{\Gamma}^T\mathbf{g}\mathbf{\Gamma}\overline{\mathbf{A}} = \mathbf{\Gamma}\overline{\mathbf{B}}\mathbf{g}\overline{\mathbf{A}} = \overline{\mathbf{C}}' . \tag{59}$$

VIII. PHYSICAL DIMENSIONS OF POLYTENSORS

28. The physical dimensions of tensor components generally depend on the tensor rank. Thus the components of a polytensor do not all have the same dimensions, and the units associated with numerical values of polytensors cannot be treated by means of scalar factors containing the appropriate units, as is usually the case in matrix problems. Instead, the dimensions of a first degree polytensor **A** may be expressed by

$$\mathbf{A} = \mathbf{D}_A\hat{\mathbf{A}} , \tag{60}$$

where $\hat{\mathbf{A}}$ is the dimensionless form of the polytensor and $\mathbf{D}_A$ is a diagonal matrix whose diagonal elements are the units of the corresponding polytensor components. In consequence, we have the following theorems.

29. **Theorem:** If **B** is the direct product of first degree polytensors **A** and **C**, then **B** is related to its dimensionless form $\hat{\mathbf{B}}$ by

$$\mathbf{B} = \mathbf{D}_A\hat{\mathbf{B}}\mathbf{D}_C . \tag{61}$$

Proof: The direct product may be expressed as the matrix product

$$\mathbf{B} = \mathbf{A}\mathbf{C}^T = \mathbf{D}_A\hat{\mathbf{A}}\hat{\mathbf{C}}^T\mathbf{D}_C = \mathbf{D}_A\hat{\mathbf{B}}\mathbf{D}_C . \tag{62}$$

30. **Theorem:** If a first degree polytensor **C** is the contraction **B·A**, where **A** and **B** are first and second degree polytensors, respectively, then **B** is related to its dimensionless form by

$$\mathbf{B} = \mathbf{D}_C\hat{\mathbf{B}}\mathbf{D}_A^{-1} . \tag{63}$$

Proof: By definition,

$$\mathbf{D}_C\hat{\mathbf{C}} = \mathbf{B}\mathbf{D}_A\hat{\mathbf{A}} , \tag{64}$$

or

$$\hat{\mathbf{C}} = \mathbf{D}_C^{-1}\mathbf{B}\mathbf{D}_A\hat{\mathbf{A}} . \tag{65}$$

The prefactor of $\hat{\mathbf{A}}$ in Eq. (65) must be a dimensionless square matrix whose elements are numerically equal to those of **B**; the prefactor thus defines $\hat{\mathbf{B}}$, and the theorem follows from this definition.

31. A special case of the problem in item 30 is that in which **A** is an eigenvector of **B**, i.e.,

$$\mathbf{C} = \mathbf{B}\cdot\mathbf{A} = \mathbf{k}\mathbf{A} , \tag{66}$$

where **k** is what is usually regarded as the scalar eigenvalue associated with the eigenvector. The following theorem shows that **k** is actually the product of a scalar and a matrix of units.

Theorem: If **A** is an eigenvector defined by Eq. (66) and **k** is the associated eigenvalue, then

$$\mathbf{k} = \hat{k}\,\mathbf{D}_C\mathbf{D}_A^{-1} , \tag{67}$$

where $\hat{k}$ is the dimensionless eigenvalue.

Proof: It follows from Eq. (66) that **B** and **k** have the same dimensions. If we take $\hat{k}\,\mathbf{I}$ as the dimensionless form of $\mathbf{k}$, then the theorem follows from Eq. (63).

IX. ELECTRIC MULTIPOLE INTERACTIONS

32. The application of polytensor formalism to a physical problem will be illustrated by the case of an electric charge distribution and its interaction with an external electrostatic field or with other charge distributions. Let $\rho(\mathbf{r})$ be the charge density at position $\mathbf{r}$ in the region of space containing the system of interest. The nth-order multiple moment of the charge distribution about a point $\mathbf{R}$ is the nth-rank tensor

$$\boldsymbol{\mu}^{(n)} = (n!)^{-1} \int_v (\mathbf{r} - \mathbf{R})^n \rho(\mathbf{r})\, dv\,, \tag{68}$$

where the integration is over the volume containing the charge distribution. Let $\mathbf{E}^{(k)}$ be the k th-order gradient of the potential ϕ due to external charges:

$$\mathbf{E}^{(k)} = -\nabla^k \phi\,, \tag{69}$$

where the gradient is evaluated at $\mathbf{r} = \mathbf{R}$. The expansion of multipole moments in powers of the field gradients can be written in the form[5,6]

$$\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}_0^{(n)} + \sum_{k=0}^{\infty} \mathbf{p}^{(n,k)} \cdot k \cdot \mathbf{E}^{(k)} + \tfrac{1}{2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \mathbf{p}^{(n,k,l)} \cdot (k + l) \cdot \mathbf{E}^{(l)} \mathbf{E}^{(k)} + \cdots\,, \tag{70}$$

where $\boldsymbol{\mu}_0^{(n)}$ is the multipole moment in the absence of an external field and the $\mathbf{p}$'s are generalized polarizabilities. Let $\mathbf{M}$ be a first degree polytensor composed of the tensors $\boldsymbol{\mu}^{(n)}$ and $\mathbf{E}$ be a first degree polytensor composed of the tensors $\mathbf{E}^{(k)}$. These polytensors are thus single entities which characterize the charge distribution and the external field, respectively. Equation (70) can be recast in the more concise form

$$\mathbf{M} = \mathbf{M}_0 + \mathbf{P} \cdot \mathbf{E} + \tfrac{1}{2}\mathbf{Q} : \mathbf{E}\mathbf{E} + \cdots\,, \tag{71}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are second and third degree polarizability polytensors, respectively.

Now consider a system composed of N nonoverlapping charge distributions which interact by way of the fields of their multipole moments. For simplicity, we will assume that the permanent multipole moments $\mathbf{M}_0$ vanish and that the field is small enough to neglect terms in Eq. (70) higher than the linear term in $\mathbf{E}$. Let the multipole moments of charge distribution I be defined with respect to a local origin $\mathbf{R}_I$. The multipole polytensor $\mathbf{M}_I$ of the Ith charge distribution is then

$$\mathbf{M}_I = \mathbf{P}_I \cdot \left(\mathbf{E}_I - \sum_{\substack{J=1 \\ J \neq I}}^{N} \mathbf{T}_{IJ} \cdot \mathbf{M}_J \right), \quad I = 1,\ldots,N\,, \tag{72}$$

where $\mathbf{E}_I$ is the external field gradient polytensor evaluated at $\mathbf{R}_I$ and $\mathbf{T}_{IJ}$ is the second degree polytensor composed of the multipole field tensors,[1]

$$\mathbf{T}_{IJ}^{(m,n)} = (-1)^n \nabla_I^{m+n} R_{JI}^{-1}\,, \tag{73}$$

where $\nabla_I = \partial / \partial \mathbf{R}_I$ and $\mathbf{R}_{JI} = \mathbf{R}_I - \mathbf{R}_J$. Thus Eq. (72) expresses the response of charge distribution I to the external field and the superimposed fields of the induced multipole moments of all other charge distributions in the system. The system of equations represented by Eq. (72) can be rewritten in the convenient matrix form

$$\mathscr{A}\mathscr{M} = \mathscr{E}\,, \tag{74}$$

where the matrices are composed of blocks of polytensors in the manner

$$\mathscr{M} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_N \end{pmatrix}, \tag{75}$$

$$\mathscr{E} = \begin{pmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_N \end{pmatrix}, \tag{76}$$

$$\mathscr{A} = \begin{pmatrix} \mathbf{P}_1^{-1} & \mathbf{T}_{12} \cdots & \mathbf{T}_{1N} \\ \mathbf{T}_{21} & \mathbf{P}_2^{-1} \cdots & \mathbf{T}_{2N} \\ \vdots & & \\ \mathbf{T}_{N1} \cdots & & \mathbf{P}_N^{-1} \end{pmatrix}. \tag{77}$$

Equation (74) may be solved for the multipole moments in the form

$$\mathscr{M} = \mathscr{A}^{-1}\mathscr{E}\,. \tag{78}$$

Equation (78) expresses the response of the system of interacting charge distributions to an arbitrary external field. This has important applications in the calculation of properties of molecular systems. The elaboration of this result will be given in future publications, but for the present, a few points are worth noting: (i) Polytensors $\mathbf{M}_I$, $\mathbf{E}_I$, $\mathbf{P}_I$, and $\mathbf{T}_{IJ}$ are composed of tensors which are totally symmetric in the component indices of each rank index; hence, the methods of the compressed polytensors can be used to reduce the matrix size. (ii) The formalism of Eqs. (72)–(78) reduces the multipole interaction problem to the same form as the dipole interaction problem which has been applied to the calculation of various molecular properties.[7] (iii) Even if one is interested, say, only in the dipole response of the system, which is expressed by appropriate elements of the $\mathscr{A}^{-1}$ matrix, that response is influenced by all orders of multipole response by virtue of the dependence of each element of $\mathscr{A}^{-1}$ on all elements of $\mathscr{A}$. (iv) The theorems proven here concerning transformation of polytensors are important in this application because the polytensors in Eqs. (74)–(78) must be specified in a single coordinate system, while certain of these, especially $\mathbf{P}_I$, are likely to be given in a local coordinate system for the particular charge distribution.

ACKNOWLEDGMENT

This investigation was supported by a research grant from the National Institute of General Medical Sciences (GM-13684).

[1] C. J. F. Böttcher, O. C. van Belle, P. Bordewijk, and A. Rip, *Theory of Electric Polarization* (Elsevier, Amsterdam, 1973), 2nd ed., Vol. 1, Chap. 1.
[2] R. E. Raab, Mol. Phys. **29**, 1323 (1975).
[3] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973), p. 991.
[4] H. Jeffreys, *Cartesian Tensors* (Cambridge U. P., Cambridge, 1931), Chap. 1.
[5] A. D. McLean and M. Yoshimine, J. Chem. Phys. **47**, 1927 (1967).
[6] J. Applequist, J. Chem. Phys. (to be published).
[7] J. Applequist, Acc. Chem. Res. **10**, 79 (1977).

Some character theory for groups of linear and antilinear operators

J. D. Newmarch [a]
Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia

(Received 5 October 1981; accepted for publication 20 November 1981)

Elementary group concepts are recast into a form applicable to finite magnetic groups of linear and antilinear operators. Analogs of useful definitions for linear groups such as the Frobenius–Schur invariant, commutator subgroups, and ambivalent classes are considered. These are applied to the 180 magnetic single and double point groups and it is shown that only seven require independent treatment of characters.

PACS numbers: 02.20. + b

1. INTRODUCTION

The use of group theory in certain areas of physics and chemistry is now well established. This generally proceeds through some form of representation theory (vector representations, ray representations, vector corepresentations, or ray corepresentations) of a group of operators on a Hilbert space.[1,2] The form used depends critically on the nature of the operators, as to whether they are linear or antilinear (Wigner[3] has shown that only these two types of operator need be considered in quantum mechanics). For vector representations of groups of linear operators an extensive literature exists, with contributions from mathematicians, physicists and chemists. Qualitative applications of vector representations (such as selection rules) are based on character theory[4,5] whereas semiquantitative calculations through the Wigner–Eckart theorem use both basis dependent information in the *n-jm* symbols[6,7] and character theory in the *n-j* symbols and isoscalars.[8,9] Characters, of course, need not be considered as any information obtainable from them can also be obtained from any realization of the vector representation, but their use enormously simplifies many calculations and justifies their detailed considerations.

Surprisingly, the character theory for the other types of representation is extremely ill-developed. Backhouse[10] has shown that a character table exists for ray representations of finite groups and Newmarch and Golding[11] (henceforth denoted as N-G) for the vector corepresentations of finite magnetic groups of linear and antilinear operators, while even this is missing for ray corepresentations. Standard vector representation concepts such as the Frobenius–Schur invariant do not appear to have been considered. In part the purpose of this paper is to fill in some of these gaps for the vector corepresentations of finite magnetic groups by considering one-dimensional irreducible corepresentations (ICRs), faithful ICRs, and complex conjugates of ICRs (Secs. 6 and 8).

During the course of this investigation an even more important gap became apparent. Magnetic groups are rather special groups in that they possess a certain subgroup of index two. This subgroup corresponds to the linear operators and its coset to the antilinear operators. The subgroup is obviously fixed by physical considerations and linear operators cannot be changed into antilinear ones without changing their physical applicability. This is reflected in the mathematics of corepresentations, and shows that a magnetic group must be considered as a *pair* of groups. The following three sections are devoted to the elementary group theory of this situation where isomorphism, homomorphism, etc., are discussed. Our own opinion is that much of the material of these sections should be self-evident. However, inappropriate statements—particularly in regard to isomorphism—have appeared sufficiently often to prompt us to spell them out. This vein is followed in Sec. 5, where it is shown that direct products of magnetic groups can be usefully defined.

The third aim of this paper is to reduce the number of magnetic single and double point groups (180 in all) requiring separate treatment of characters. Assuming known character theory of linear groups, by isomorphism (Sec. 2), factor groups (Sec. 3), direct products (Sec. 5), and an examination of the intertwining numbers (Sec. 7), it is shown that only seven groups need be considered.

Examples are drawn from the finite magnetic point groups. The theory is applicable by finite approximations to the magnetic space groups,[1] the spin groups,[12,13] and the line groups of stereo-regular polymers.[14] Much is readily transferable to compact groups, where it should find applications due to the PCT theorem for elementary particles.[15]

In general the notation is that of N-G. Magnetic point groups are labelled as in Bradley and Cracknell,[1] with an asterisk to denote the double groups. The ICRs of these groups are also labelled as in Bradley and Cracknell,[1] save for the typographical omission of the prefix D when there is no increase in degeneracy in inducing ICRs from the linear subgroup. E or e denotes the identity of the group, or the unit matrix. A prefix M denotes the magnetic group analog of a linear group concept. Proofs are usually omitted whenever they are simple modifications of those for linear groups.

2. ISOMORPHISMS AND HOMOMORPHISMS

It is only the physical importance of the time reversal operator which leads to the study of magnetic groups. Such a group contains a subgroup of linear operators and a coset of antilinear operators and clearly, to maintain their applicability, we cannot arbitrarily change linear operators into antilinear ones or vice versa. The subgroup of linear operators is just as important as the group itself. An abstract definition which indicates this is

Definition 2.1: A magnetic group M is an ordered pair of groups $M = (G,H)$ where H has index two in G.

---

[a] Present address: School of Electrical Engineering and Computer Science, University of New South Wales, P. O. Box 1, Kensington, NSW 2033, Australia.

0022-2488/83/040742-15$02.50

Classifications of groups into families are made according to various criteria. For example, there is the equivalence family of D_2, consisting of all point groups mapped onto one another by automorphisms of $O(3)$. Such an equivalence concept gives 90 families of grey and nongrey magnetic single point groups and a further 90 families of double groups.[1] In addition the fundamental group concept of isomorphism may be applied to linear groups to reduce, say, the 32 families of crystallographically distinct point groups down to 11 nonisomorphic families. However, all statements sighted on "isomorphic magnetic groups" have been rather misleading [$M_1 = (G_1, H_1)$ is isomorphic to $M_2 = (G_2, H_2)$ if $G_1 \cong G_2$] as the position of the linear subgroup need not be preserved. For example, $6'22'$ and $62'2'$ have ICRs of different dimension so degeneracies cannot be transferred despite $G_1 \cong G_2$. An appropriate definition is

Definition 2.2: Two magnetic groups $M_1 = (G_1,H_1)$ and $M_2 = (G_2,H_2)$ are M-isomorphic iff there is a group isomorphism $\phi:G_1 \rightarrow G_2$ for which $\phi(H_1) = H_2$.

This is a very stringent condition and generally requires explicit construction of the isomorphism. It cannot, for example, be weakened to *an* isomorphism $\phi:G_1 \rightarrow G_2$ and another from H_1 to H_2. To see this, consider the group $16\Gamma_2c$ of order 16 with presentation $\langle x,y | x^4 = y^4 = e, xy = yx^3 \rangle$ from the tables of Hall and Senior.[16] This group contains $(21) = Z_4 \otimes Z_2$ once characteristically (i.e., invariant under all automorphisms of $16\Gamma_2c_2$) and twice noncharacteristically. Setting $G_1 = G_2 = 16\Gamma_2c_2$, H_1 the characteristic subgroup and H_2 one of the noncharacteristic ones, then there is no M-isomorphism of M_1 onto M_2 (which here would be an automorphism) despite G_1 and G_2,H_1 and H_2 being pairwise isomorphic. The two magnetic groups are essentially different. (In fact, the first has seven ICRs and the second has eight.) A calculation for the 180 single and double magnetic point groups yields 64 nonisomorphic families which are collected in Table I.

We have dwelt on the concept of isomorphism at length primarily to show that a magnetic group must be considered as a pair of groups. These should now be obvious:

Definition 2.3: An M-homomorphism ϕ of $M_1 = (G_1,H_1)$ into $M_2 = (G_2,H_2)$ is a homomorphism ϕ of G_1 into G_2 such that $\phi(H_1) \subseteq H_2$ and $\phi(G_1 - H_1) \subseteq G_2 - H_2$.

This ensures that linear elements are mapped onto linear elements and antilinear onto antilinear. This definition has been used by Janssen in discussing projective corepresentations.[17]

Definition 2.4: An M-normal subgroup of $M = (G,H)$ is a subgroup of H (and hence of G) which is normal in G (and hence normal in H).

The subgroups of G for the magnetic single point groups have been listed by Ascher and Janner,[18] and of course only a few are M-normal. Later it is shown that they may be obtained from the character table. For the moment,

Theorem 2.5 (First Isomorphism Theorem): Let $M = (G,H)$ be a magnetic group and ϕ an M-homomorphism of M. Then the kernel of ϕ is an M-normal subgroup L and the image of M is naturally M-isomorphic to $(G/L,H/L)$. Conversely, each M-normal subgroup L defines an M-homomorphism of M onto $(G/L,H/L)$.

The other isomorphism theorems can be similarly adapted. However, this is all we need for now.

3. COREPRESENTATIONS

Definition 3.1: A corepresentation D is an M-homomorphism of a magnetic group into a magnetic group of operators (G,H) over a complex vector space, where the operators of H are linear and of G-H are antilinear.

Herbut *et al.*[19] have given a similar definition for their unitary/antiunitary representations of magnetic groups and introduced the term "antimatrix" for the matrix of an antilinear operator. Whilst we support their viewpoint in which sense we have interpreted corepresentations, we consider that the tensor notation from spinor calculus used by Newmarch and Golding[20] handles antilinear operators in the most effective manner. We regard both the common notation used here and that of Herbut *et al.*[19] to be "approximations" to the tensor notation, and use the common notation on the grounds of familiarity and a mild preference for seeing complex conjugates explicitly.

Matrices of linear and antilinear operators of a corepresentation and irreducible corepresentations (ICRs) are defined in the normal way. From these we have

Lemma 3.2: Let D be a corepresentation of $M = (G,H)$ with character χ and let $u \epsilon H$. If n is the order of u and f the degree of D [$f = \chi(e)$] then

(a) $D(u)$ is similar to diag. $(\epsilon_1,\epsilon_2,\ldots,\epsilon_f)$,

(b) $\epsilon_i^n = 1$ for all i,

(c) $\chi(u) = \sum_{I=1}^{f} \epsilon_i$,

(d) $|\chi(u)| \leqslant \chi(e) = f$.

Lemma 3.3: If D is a corepresentation of M, then the kernel of D (ker D) is an M-normal subgroup of M, and $u \epsilon$ ker D iff $\chi(u) = \chi(e)$.

Lemma 3.4: Let $D = \Sigma n_i D_i$ be a corepresentation of M and D_i be ICRs. Then ker $D = \cap\{\ker D_i : n_i > 0\}$ and $\cap\{\ker D_i : \text{all ICRs}\} = \{e\}$.

These are all proved in exactly the same manner as for representations (e.g., Isaacs[21]). The regular corepresentation and its properties are given in N-G.

Every M-normal subgroup of a magnetic group may be found from the character table by taking irreducible characters and sums of characters and finding those elements u for which $\chi(u) = \chi(e)$. For example, the group $4'/mmm$ has M-normal subgroups $\{E,I\}$ from E_g, $\{E,C_{2z},\sigma_z\}$ from B_{1g},$\{E,C_{2x},C_{2y},C_{2z}\}$ from A_u, $\{E,\sigma_z\}$ from E_u,$\{E,C_{2z},\sigma_x,\sigma_y\}$ from B_{1u}, and $\{E,C_{2z}\}$ from $B_{1g} \oplus A_u$. (The character table is given in N-G).

One of the major features which distinguishes corepresentation theory from representation theory is the different form of Schur's lemmas for the two theories. For linear groups any matrix commuting with an IR is a constant diagonal matrix (quantitatively, the set of all such matrices form an algebra of dimension one over ℂ). In N-G it was shown

TABLE I. The M-isomorphic families of magnetic point groups. The families are listed by ascending orders of the groups. The notation for groups and group elements is that of Bradley and Cracknell[1] with an asterisk to distinguish double groups. Elements isomorphic to each other in each family are listed in the same order in rows of the "Isomorphism" column. A point group label is given for G under "Popular name for G" although for double groups, $\theta^2 = \bar{E}$ and G is not in fact the point group. A "dash" here indicates a grey group. The comments are illustrative, not exhaustive.

Family	Order	Magnetic group	Popular name for G	H	Isomorphism	Comments
1	2	$11'$	C_1'	C_1	θ	Character table as H
		$\bar{1}'$	C_i	C_1	θI	
		$2'$	C_2	C_1	θC_{2z}	
		m'	C_{1h}	C_1	$\theta\sigma_h$	
2	4	$22'2'$	D_2	C_2	$C_{2z},\theta C_{2y}$	Character table as H
		$2/m'$	C_{2h}	C_2	$C_{2z},\theta I$	
		$21'$	C_2'	C_2	C_{2z},θ	
		$2'/m'$	C_{2h}	C_i	$I,\theta C_{2z}$	
		$2'/m$	C_{2h}	C_{1h}	$\sigma_z,\theta I$	
		$\bar{1}1'$	C_i'	C_i	I,θ	
		$*m'$	C_{1h}^*	C_1^*	$\bar{E},\theta\sigma_z$	
		$*2'$	C_2^*	C_1^*	$\bar{E},\theta C_{2z}$	
		$2m'm'$	C_{2v}	C_2	$C_{2z},\theta\sigma_y$	
		$2'm'm$	C_{2v}	C_{1h}	$\sigma_y,\theta C_{2x}$	
		$m1'$	C_{1h}'	C_{1h}	σ_z,θ	
3	4	$4'$	C_4	C_2	θC_{4z}^+	Homomorphic image of family 8
		$\bar{4}'$	S_4	C_2	θS_{4z}^-	
		$*11'$	$C_1^{*\prime}$	C_1^*	θ	
		$*\bar{1}'$	C_i^*	C_1^*	θI	
4	6	$32'$	D_3	C_3	$C_3^+,\theta C_{21}'$	Character table as H
		$\bar{3}m'$	D_{3d}	S_6	$C_3^+,\theta\sigma_{d1}$	
5	6	$6'$	C_6	C_3	θC_6^+	Homomorphic image of, e.g., family 13
		$\bar{6}'$	C_{3h}	C_3	θS_3^-	
		$\bar{3}'$	S_6	C_3	θS_6^+	
		$31'$	C_3'	C_3	θC_3^+	
6	8	$4/m'$	C_{4h}	C_4	$C_{4z}^+,\theta I$	Homomorphic image of, e.g., family 23
		$4'/m'$	C_{4h}	S_4	$S_{4z}^-,\theta I$	
		$*m1'$	$C_{1h}^{*\prime}$	C_{1h}^*	σ_z,θ	
		$*2'/m$	C_{2h}^*	C_{1h}^*	$\sigma_z,\theta I$	
		$*21'$	$C_2^{*\prime}$	C_2^*	C_{2z},θ	
		$41'$	C_4'	C_4	C_{4z}^+,θ	
		$\bar{4}1'$	S_4'	S_4	S_{4z}^-,θ	
		$*2/m'$	C_{2h}^*	C_2^*	$C_{2z},\theta I$	
7	8	$4'22'$	D_4	D_2	$C_{2x},\theta C_{4z}^+$	Homomorphic image of family 19
		$\bar{4}'2m'$	D_{2d}	D_2	$C_{2x},\theta S_{4z}^-$	
		$4'mm'$	C_{4v}	C_{2v}	$\sigma_x,\theta C_{4z}^+$	
		$\bar{4}'2'm$	D_{2d}	C_{2v}	$\sigma_x,\theta S_{4z}^-$	
		$*\bar{1}1'$	$C_i^{*\prime}$	C_i^*	I,θ	
8	8	$*4'$	C_4^*	C_2^*	θC_{4z}^+	
		$*\bar{4}'$	S_4^*	C_2^*	θS_{4z}^{*-}	
9	8	$42'2'$	D_4	C_4	$C_{4z}^+,\theta C_{2x}$	Character table as H
		$4m'm'$	C_{4v}	C_4	$C_{4z}^+,\theta\sigma_x$	
		$\bar{4}2'm'$	D_{2d}	S_4	$S_{4z}^-,\theta C_{2x}$	
		$*2'2'2$	D_2^*	C_2^*	$C_{2z},\theta C_{2y}$	
		$*m'm'2$	C_{2v}^*	C_2^*	$C_{2z},\theta\sigma_y$	
		$*m'm2'$	C_{2v}^*	C_{1h}^*	$\sigma_y,\theta\sigma_x$	
10	8	$2221'$	D_2'	D_2	C_{2x},C_{2y},θ	Character table as H
		$m'm'm'$	D_{2h}	D_2	$C_{2x},C_{2y},\theta I$	
		$mm21'$	C_{2v}'	C_{2v}	σ_x,σ_y,θ	
		mmm'	D_{2h}	C_{2v}	$\sigma_x,\sigma_y,\theta I$	
		$m'm'm$	D_{2h}	C_{2h}	$C_{2z},I,\theta C_{2y}$	

TABLE I. (*Continued*)

Family	Order	Magnetic group	Popular name for G	H	Isomorphism	Comments
		$*2'/m'$	C_{2h}^{*}	C_{i}^{*}	$I,\bar{E},\theta C_{2z}$	
		$2/m1'$	C'_{2h}	C_{2h}	C_{2z},I,θ	
11	8	$4'/m$	C_{4h}	C_{2h}		Direct product of $\bar{1}1'$ with $4'$ (family 3)
12	12	$*31'$	$C_{3}^{*\prime}$	C_{3}^{*}	C_{3}^{+},θ	
		$*\bar{3}'$	S_{6}^{*}	C_{3}^{*}	$C_{3}^{+},\theta I$	
13	12	$*6'$	C_{6}^{*}	C_{3}^{*}	$C_{3}^{+},\theta C_{2}$	Direct product of $\bar{1}1'$ with $6'$ (family 5)
		$*\bar{6}'$	C_{3h}^{*}	C_{3}^{*}	$C_{3}^{+},\theta\sigma_{h}$	
		$6/m'$	C_{6h}	C_{6}	$C_{6}^{+}\theta\sigma_{h}$	
		$6'/m$	C_{6h}	C_{3h}	$S_{3}^{-},\theta\sigma_{h}$	
		$6'/m'$	C_{6h}	S_{6}	$S_{6}^{-},\theta\sigma_{h}$	
14	12	$*3m'$	C_{3v}^{*}	C_{3}^{*}	$C_{3}^{+},\theta\sigma_{d1}$	Character table as H
		$*32'$	D_{3}	C_{3}^{*}	$C_{3}^{+},\theta C'_{21}$	
15	12	$\bar{3}m'$	D_{3d}	S_{6}		Character table as H
16	12	$6'22'$	D_{6}	D_{3}	$C_{3}^{+},C'_{21},\theta C_{2}$	Character table as H
		$6'mm'$	C_{6v}	C_{3v}	$C_{3}^{+},\sigma_{d1},\theta C_{2}$	
		$\bar{6}'m'2$	D_{3h}	D_{3}	$C_{3}^{+},C'_{21},\theta\sigma_{h}$	
		$\bar{6}'m2'$	D_{3h}	C_{3v}	$C_{3}^{+},\sigma_{v1},\theta\sigma_{h}$	
		$\bar{3}'m$	D_{3d}	D_{3}	$C_{3}^{+},C'_{21},\theta I$	
		$\bar{3}'m$	D_{3d}	C_{3v}	$C_{3}^{+},\sigma_{d1},\theta I$	
		$321'$	D'_{3}	D_{3}	C_{3}^{+},C'_{21},θ	
		$3m1'$	C'_{3v}	C_{3v}	$C_{3}^{+},\sigma_{d1},\theta$	
17	12	$62'2'$	D_{6}	C_{6}	$C_{6}^{+},\theta C'_{21}$	Character table as H
		$6m'm'$	C_{6v}	C_{6}	$C_{6}^{+},\theta\sigma_{d1}$	
		$\bar{6}m'2'$	D_{3h}	C_{3h}	$S_{3}^{-},\theta C'_{21}$	
18	12	$\bar{3}1'$	S'_{6}	S_{6}	S_{6}^{-},θ	Homomorphic image of, e.g., family 29
		$61'$	C'_{6}	C_{6}	C_{6}^{+},θ	
		$\bar{6}1'$	C'_{3h}	C_{3h}	S_{3}^{-},θ	
19	16	$*4'22'$	D_{4}^{*}	D_{2}^{*}	$C_{2x},C_{2y},\theta C_{4z}^{+}$	
		$*\bar{4}'2m'$	D_{2d}^{*}	D_{2}^{*}	$C_{2x},C_{2y},\theta S_{4z}^{-}$	
		$*4'mm'$	C_{4v}^{*}	C_{2v}^{*}	$\sigma_{x},\sigma_{y},\theta C_{4z}^{+}$	
		$*\bar{4}'m2'$	D_{2d}^{*}	C_{2v}^{*}	$\sigma_{x},\sigma_{y},\theta S_{4z}^{-}$	
20	16	$*4m'm'$	C_{4v}^{*}	C_{4}^{*}	$C_{4z}^{-},\theta\sigma_{x}$	Character table as H
		$*42'2'$	D_{4}^{*}	C_{4}^{*}	$C_{4z}^{-},\theta C_{2x}$	
		$*\bar{4}2'm'$	D_{2d}^{*}	S_{4}^{*}	$S_{4z}^{+},\theta C_{2x}$	
21	16	$*2221'$	$D_{2}^{*\prime}$	D_{2}^{*}	C_{2x},C_{2y},θ	Character table as H
		$*mm21'$	$C_{2v}^{*\prime}$	C_{2v}^{*}	$\sigma_{x},\sigma_{y},\theta$	
		$*m'm'm'$	D_{2h}^{*}	D_{2}^{*}	$C_{2x},C_{2y},\theta I$	
		$*mmm'$	D_{2h}^{*}	C_{2v}^{*}	$\sigma_{x},\sigma_{y},\theta I$	
22	16	$4/m'm'm'$	D_{4h}	D_{4}	$C_{4z}^{+},C_{2x},\theta I$	Character table as H
		$4/m'mm$	D_{4h}	C_{4v}	$C_{4z}^{+},\sigma_{x},\theta I$	
		$4'/m'm'm$	D_{4h}	D_{2d}	$S_{4z}^{-},C_{2x},\theta I$	
		$4221'$	D'_{4}	D_{4}	C_{4z}^{+},C_{2x},θ	
		$\bar{4}2m1'$	D'_{2d}	D_{2d}	S_{4z}^{-},C_{2x},θ	
		$4mm1'$	C'_{4v}	C_{4v}	$C_{4z}^{+},\sigma_{x},\theta$	
23	16	$*4/m'$	C_{4h}^{*}	C_{4}^{*}	$C_{4z}^{+},\theta I$	
		$*41'$	$C_{4}^{*\prime}$	C_{4}^{*}	C_{4z}^{+},θ	
		$*4'/m'$	C_{4h}^{*}	S_{4}^{*}	$S_{4z}^{-},\theta I$	

TABLE I. (*Continued*)

Family	Order	Magnetic group	Popular name for G	H	Isomorphism	Comments
		$*\bar{4}1'$	$S_4^{*\prime}$	S_4^*	S_{4z}^-,θ	
24	16	$*4'/m$	C_{4h}^*	C_{2h}^*		Direct product of $\bar{1}1'$ with $*4'$ (family 8)
25	16	$*m'm'm$ $4/mm'm'$	D_{2h}^* D_{4h}	C_{2h}^* C_{4h}	$C_{2z},I,\theta C_{2x}$ $C_{4z}^+,I,\theta C_{2x}$	Character table as H
26	16	$*2/m1'$ $4/m1'$	$C_{2h}^{*\prime}$ C_{4h}'	C_{2h}^* C_{4h}	C_{2z},I,θ $C_{4z}^+,I,\theta C_{4z}^+$	Direct product of $\bar{1}1'$ with $41'$ (family 6)
27	16	$4'/mmm$	D_{4h}	D_{2h}		Homomorphic image of family 42
28	16	$mmm1'$	D_{2h}'	D_{2h}		Character table as H
29	24	$*61'$ $*6/m'$ $*6'/m$ $*\bar{6}1'$	$C_6^{*\prime}$ C_{6h}^* C_{6h}^* $C_{3h}^{*\prime}$	C_6^* C_6^* C_{3h} C_{3h}^*	C_6^+,θ $C_6^+,\theta I$ $S_3^-,\theta I$ S_3^-,θ	
30	24	$*3m1'$ $*\bar{3}'m$ $*321'$ $*\bar{3}'m'$	$C_{3v}^{*\prime}$ D_{3d}^* $D_3^{*\prime}$ D_{3d}^*	C_{3v}^* C_{3v}^* D_3^* D_3^*	C_3^+,θ_{d1},θ $C_3^+,\sigma_{d1},\theta I$ C_3^+,C_{21}',θ $C_3^+,C_{21}',\theta I$	
31	24	$*62'2'$ $*6m'm'$ $*\bar{6}m'2'$	D_6^* C_{6v}^* D_{3h}^*	C_6^* C_6^* C_{3h}^*	$C_6^+,\theta C_{21}'$ $C_6^+,\theta\sigma_{d1}$ $S_3,\theta C_{21}'$	Character table as H
32	24	$*\bar{3}m'$	D_{3d}^*	S_6^*		Character table as H
33	24	$*6'2'2$ $*6'm'm$ $*\bar{6}'m'2$ $*\bar{6}'m2'$	D_6^* C_{6v}^* D_{3h}^* D_{3h}^*	D_3^* C_{3v}^* D_3^* C_{3v}^*	$C_3^+,C_{21}',\theta C_2$ $C_3^+,\sigma_{d1},\theta C_2$ $C_3^+,C_{21}',\theta\sigma_h$ $C_3^+,\sigma_{v1},\theta\sigma_h$	Character table as H
34	24	$*6'/m'$	C_{6h}^*	S_6^*		Direct product of $\bar{1}1'$ with $*6'$ (family 13)
35	24	$*\bar{3}1'$	$S_6^{*\prime}$	S_6^*		Direct product of $\bar{1}1'$ with $*31'$ (family 12)
36	24	$6/m1'$	C_{6h}'	C_{6h}		Direct product of $\bar{1}1'$ with $61'$ (family 12)
37	24	$6'/m'm'm$ $6'/mm'm$ $6/m'mm$ $6221'$ $6mm1'$ $\bar{6}2m1'$ $\bar{3}m1'$	D_{6h} D_{6h} D_{6h} D_6' C_{6v}' D_{3h}' D_{3d}'	D_{3d} D_{3h} C_{6v} D_6 C_{6v} D_{3h} D_{3d}	$S_6^-,C_{21}',\theta C_2$ $S_3^-,C_{21}',\theta I$ $C_6^+,\sigma_{d1},\theta I$ C_6^+,C_{21}',θ C_6^+,σ_{d1},θ C_3^-,C_{21}',θ S_6^-,C_{21}',θ	Character table as H
38	24	$6/mm'm'$	D_{6h}	C_{6h}		Character table as H

TABLE I. (*Continued*)

39	24	$231'$ $m'3$	T' T_h	T T	C_{31}^{-},C_{2x},θ $C_{31}^{-},C_{2x},\theta I$	Homomorphic image of family 53
40	24	$\bar{4}'3m'$ $4'32'$	T_d 0	T T	$C_{31}^{-},C_{2x},\theta\sigma_{da}$ $C_{31}^{-},C_{2x},\theta C_{2a}$	Character table as *H*
41	32	$4/mmm1'$	D_{4h}'	D_{4h}		Character table as *H*
42	32	$*4'/mmm'$	D_{4h}^{*}	D_{2h}^{*}		Direct product of $\bar{1}1'$ with $*4'22'$ (family 19)
43	32	$*4/mm'm'$	D_{4h}^{*}	C_{4h}^{*}		Character table as *H*
44	32	$*4/m'm'm'$ $*4/m'mm$ $*4'/m'm'm$ $*4221'$ $*4mm1'$ $*\bar{4}2m1'$	D_{4h}^{*} D_{4h}^{*} D_{4h}^{*} $D_{4}^{*\prime}$ $C_{4v}^{*\prime}$ $D_{2d}^{*\prime}$	D_{4}^{*} C_{4v}^{*} D_{2d}^{*} D_{4}^{*} C_{4v}^{*} D_{2d}^{*}	$C_{4z}^{+},C_{2x},\theta I$ $C_{4z}^{+},\sigma_x,\theta I$ $S_{4z}^{-},C_{2x},\theta I$ C_{4z}^{+},C_{2x},θ $C_{4z}^{+},\sigma_x,\theta$ S_{4z}^{-},C_{2x},θ	Character table as *H*
45	32	$*mmm1'$	$D_{2h}^{*\prime}$	D_{2h}^{*}		Character table as *H*
46	32	$*4/m1'$	$C_{4h}^{*\prime}$	C_{4h}^{*}		Direct product of $\bar{1}1'$ with $*41'$ (family 23)
47	48	$*6'/m'm'm$	D_{6h}^{*}	D_{3d}^{*}		Character table as *H*
48	48	$*6/mm'm'$	D_{6h}^{*}	C_{6h}^{*}		Character table as *H*
49	48	$*6'/mmm'$ $*6/m'm'm'$ $*6/m'mm$ $*6221'$ $*6mm1'$ $*\bar{6}2m1'$	D_{6h}^{*} D_{6h}^{*} D_{6h}^{*} $D_{6}^{*\prime}$ $C_{6v}^{*\prime}$ $D_{3h}^{*\prime}$	D_{3h}^{*} D_{6}^{*} C_{6v}^{*} D_{6}^{*} C_{6v}^{*} D_{3h}^{*}	$S_{3}^{-},C_{21}',\theta I$ $C_{6}^{+},C_{21}',\theta I$ $C_{6}^{+},\sigma_{d1},\theta I$ C_{6}^{+},C_{21}',θ $C_{6}^{+},\sigma_{d1},\theta$ S_{3}^{-},C_{21}',θ	Character table as *H*
50	48	$6/mmm1'$	D_{6h}'	D_{6h}		Character table as *H*
51	48	$*m'3$ $*231'$	T_{h}^{*} $T^{*\prime}$	T^{*} T^{*}	$C_{31}^{-},C_{2x},\bar{C}_{2y},\theta I$ $C_{31}^{-},C_{2x},\bar{C}_{2y},\theta$	
52	48	$*\bar{4}'3m'$ $*4'32'$	T_{d}^{*} 0^{*}	T^{*} T^{*}	$C_{31}^{-},C_{2x},\bar{C}_{2y},\theta\sigma_{da}$ $C_{31}^{-},C_{2x},\bar{C}_{2y},\theta C_{2a}$	Character table as *H*
53	48	$*\bar{3}m1'$	$D_{3d}^{*\prime}$	D_{3d}^{*}		Direct product of $\bar{1}1'$ with $*321'$ (family 30)
54	48	$*6/m1'$	$C_{6h}^{*\prime}$	C_{6h}^{*}		Direct product of $\bar{1}1'$ with $*61'$ (family 29)
55	48	$m'3m'$ $4321'$ $m'3m$ $\bar{4}3m1'$	0_h $0'$ 0_h T_d'	0 0 T_d T_d	$C_{4x}^{+},C_{31}^{-},C_{2b},\theta I$ $C_{4x}^{+},C_{31}^{-},C_{2b},\theta$ $S_{4x}^{-},C_{31}^{-},\sigma_{db},\theta I$ $S_{4x}^{-},C_{31}^{-},\sigma_{db},\theta$	Character table as *H*
56	48	$m31'$	T_h'	T_h		Homomorphic image of family 62

TABLE I. (*Continued*)

57	48	$m3m'$	0_h	T_h		Character table as H
58	64	$*4/mmm1'$	$D_{4h}^{*\prime}$	D_{4h}^{*}		Character table as H
59	96	$*m'3m'$ $*m'3m$ $*4321'$ $*\bar{4}3m1'$	0_h^* 0_h^* $0^{*\prime}$ $T_d^{*\prime}$	0^* T'_d 0^* T_d^*	$C_{4x}^+,\bar{C}_{31}^-,C_{2b},\theta I$ $S_{4x}^-,\bar{C}_{31}^-,\sigma_{db},\theta I$ $C_{4x}^+,\bar{C}_{31}^-,C_{2b},\theta$ $S_{4x}^-,\bar{C}_{31}^-,\sigma_{db},\theta$	Character table as H
60	96	$*m31'$	$T_h^{*\prime}$	T_h^*		Direct product of $\bar{1}1'$ with $231'$ (family 53)
61	96	$*m3m'$	0_h^*	T_h^*		Character table as H
62	96	$m3m1'$	$0'_h$	0_h		Character table as H
63	96	$*6/mmm1'$	$D_{6h}^{*\prime}$	D_{6h}^{*}		Direct product of $\bar{1}1'$ with $*6221'$ (family 51)
64	192	$*m3m1'$	$0_h^{*\prime}$	0_h^*		Character table as H

that the algebra of such matrices is of dimension one, four or two over $\mathbb{R}$ (i.e., is isomorphic to $\mathbb{R}$,$\mathbb{Q}$, or $\mathbb{C}$). Labelling the ICRs as types (a), (b), and (c), respectively, in concordance with standard usage, the intertwining number I was introduced: for an ICR of type (a), $I = 1$, for an ICR of type (b), $I = 4$, and for an ICR of type (c), $I = 2$. The row orthogonality relation for ICRs was then shown to be

$$\sum_u \chi_i(u)\chi_j(u)^* = \delta_{ij} I_i |H|.$$

These next results all follow as for representation theory (e.g., Isaacs[21]).

Theorem 3.5: Let D be a corepresentation of $M = (G,H)$ and $L \subseteq \ker D$ an M-normal subgroup of M. Define $\hat{D}$ on M/L by $\hat{D}(gL) = D(g)$ for all $g \in G$. Then

(a) $\hat{D}$ is a corepresentation of M/L,

(b) $\hat{D}$ is irreducible iff D is irreducible,

(c) if D is irreducible with intertwining number I and $\hat{D}$ has intertwining number $\hat{I}$, $I = \hat{I}$.

Conversely,

Theorem 3.6: Let L be an M-normal subgroup of $M = (G,H)$ and $\hat{D}$ a corepresentation of M/L. Define D on M by $D(g) = \hat{D}(gL)$ for all $g \in G$. Then

(a) D is a corepresentation of M with $L \subseteq \ker D$,

(b) D is irreducible iff $\hat{D}$ is irreducible,

(c) the intertwining numbers of D and $\hat{D}$ are equal.

In terms of characters:

Corollary 3.7: Let χ be a function on M, $\hat{\chi}$ a function on M/L, and $\chi(gL) = \hat{\chi}(g)$. Then

(a) χ is a character iff $\hat{\chi}$ is a character,

(b) χ is irreducible iff $\hat{\chi}$ is irreducible, and then they have the same intertwining number.

These three results can be used in exactly the same manner as they are in representation theory. In particular, every magnetic *single* point group is an M-homomorphic image of a magnetic *double* point group and hence the single group does not require separate treatment. While this result has been implicitly assumed by many authors we feel a proof is important as many other equally "obvious" transfers from representation theory are known to be false. In this case we may eliminate the 31 isomorphism families containing single groups from any separate calculations, to leave 33 nonisomorphic families of magnetic double point groups.

4. MAGNETIC CLASSES

An M-class C of $M = (G,H)$ was defined in N-G to be an equivalence class of elements of H: $u_1,u_2 \in C$ if there exists *either* $u \in H$ with $u\, u_1\, u^{-1} = u_2$ *or* $a \in G$-H with $a\, u_1\, a^{-1} = u_2^{-1}$ (or both). (The term C class was used in N-G for what we here call an M-class. The prefix M is more appropriate as it is a *group* concept rather than a *corepresentation* one.) The character χ of a corepresentation is an M-class function and from this follows the column orthogonality relation for ICRs

$$\sum_i \frac{\chi_i(u_1)\chi_i(u_2)^*}{I_i} = \delta(C_{u_1},C_{u_2})\frac{|H|}{n_{u_1}},$$

where $n_u = |C_u|$ (N-G, Theorem 16).

It is well known that for ordinary groups the order of a class equals the order of the group divided by the order of the centralizer of any element of the class. A similar result holds for magnetic groups—once the centralizer is defined.

Definition 4.1: The M-centralizer $\mathbb{C}(L)$ of a set of linear operators L in M is

$$\mathbb{C}(L) = \{u,a \in M:\ ul = lu, al = l^{-1}a\ \forall l \in L\}.$$

Lemma 4.2: $\mathbb{C}(u)$ is a subgroup of M.
$\mathbb{C}(u)$ may consist of linear elements only or of both linear and antilinear elements.

Theorem 4.3: If u is an element of an M-class C of M then

$$|C|\times|\mathbb{C}(u)| = |M|.$$

This may be shown by adapting the ordinary group proof of, say, Jansen and Boon.[22] This means considering linear and antilinear elements separately and consequently is a little tedious. Similar adaptations are required in dealing with the class multiplication constants:

Definition 4.4: Let C_i and C_j be two M-classes. The class multiplication constant h_{ij}^{k} is the number of pairs $u_i \in C_i$ and $u_j \in C_j$ whose product is any fixed element $u_k \in C_k$.

Lemma 4.5: h_{ij}^{k} is independent of the element $u_k \in C_k$.

Proof: We prove this simple result only to demonstrate the alterations necessary for magnetic groups. Let $u_k, u_k' \in C_k$

(a) If $u_k = u u_k' u^{-1}$, then to each pair $u_i \in C_i, u_j \in C_j$ with

$u_i u_j = u_k$ corresponds another pair $u_i' = u^{-1} u_i u \in C_i$

and $u_j' = u^{-1} u_j u \in C_j$ with $u_i' u_j' = u_k'$.

(b) If $u_k = a u_k'^{-1} a^{-1}$, then to each pair $u_i \in C_i, u_j \in C_j$ with

$u_i u_j = u_k$ corresponds another pair

$u_i' = a^{-1} u_i^{-1} a \in C_i$, $u_j' = (a^{-1} u_i) u_j^{-1} (a^{-1} u_i)^{-1} \in C_j$

with
$u_i' u_j' = u_k'$.

Hence the number of pairs $u_i' u_j' = u_k'$ equals the number of pairs $u_i u_j = u_k$.

Similarly,

Theorem 4.6: With $C_1 = \{e\}$ and $C_{-i} = (C_i)^{-1}$

(a) $h_{ij}^{1} = |C_i| \delta_{-i,j}$,

(b) $h_{ij}^{k} = h_{ji}^{k} = h_{i-j}^{-k}$,

(c) if D is irreducible and $S_u = \sum_{u' e C_u} D(u')$ then

$$S_u = \frac{|C_u| \chi(u) I}{f},$$

where I is the intertwining number of D and f the degree of D (This follows from Theorem 12 of N-G.),

(d) with S as in (c)

$$S_{u_i} S_{u_j} = \sum_{\text{all classes}} h_{ij}^{k} S_{u_k},$$

(e) $|C_{u_i}| \cdot |C_{u_j}| \chi(u_i) \chi(u_j) = f \sum_{\text{all classes}} h_{ij}^{k} |C_{u_k}| \chi(u_k)$,

(f) $h_{ij}^{k} = \dfrac{1}{|H|} \sum_{\substack{\text{all ICRs} \\ p}} \dfrac{|C_{u_i}| \cdot |C_j^{u}}{f_p I_p} \chi_p(u_i) \chi_p(u_j) \chi_p(u_k)^*$,

(g) (van Zanten and de Vries[23])

$$\sum_{i,j,k} \frac{C_{u_k}}{|C_{u_i}| \cdot |C_{u_j}|} (h_{ij}^{k})^2 = |H| \sum_{\substack{\text{all ICRs} \\ p}} \frac{I_p}{f_p^2}.$$

5. DIRECT PRODUCT GROUPS

On the face of it, direct product groups are not particularly useful. For example, Cracknell[24] considers $m'3$ as the direct product $32 \times \bar{1}'$ and concludes that this is not profitable as the three ICRs of $m'3$ are not direct products of the four IRs of 32 and the one ICR of $\bar{1}'$. However, if we return to the group idea of direct products being formed of ordered pairs, then $32 \times \bar{1}'$ is a very odd group indeed as it contains the element $(E, \theta I)$, for example, which acts linearly in one space and antilinearly in another. Whilst it is possible that such mixed magnetic/linear groups may yet find applications, we investigate in this section a direct product which is conceptually simpler.

Definition 5.1: Let M_1 and M_2 be magnetic groups and set the magnetic group

$$M = M_1 \overset{M}{\times} M_2$$

to be their M-direct (outer) product if

(a) the linear subgroup of M is the direct product of the linear subgroups of M_1 and M_2.

(b) the antilinear coset of M is the direct product of antilinear cosets of M_1 and M_2.

Symbolically, if $M_1 = (G_1, H_1)$ and $M_2 = (G_2, H_2)$ then

$$M_1 \overset{M}{\times} M_2 = ((G_1 - H_1) \times (G_2 - H_2) \cup (H_1 \times H_2), H_1 \times H_2).$$

Some standard result for ordinary direct products do not transfer to magnetic direct products:

(a) $|M_1 \overset{M}{\times} M_2| = |M_1| \cdot |M_2| / 2$. This follows from the orders of the linear subgroups.

(b) Neither M_1 nor M_2 need be M-isomorphic to a subgroup of

$$M_1 \overset{M}{\times} M_2$$

as neither $\{M_1, e\}$ nor $\{e, M_2\}$ are subgroups. For example

$$4' \overset{M}{\times} 21' = \{(E, E), (E, C_2), (C_2, E), (C_2, C_2),$$

$$(\theta C_4^{+}, \theta), (\theta C_4^{+}, \theta C_2), (\theta C_4^{-}, \theta), (\theta C_4^{-}\ \theta C_2)\}$$

and $21'$ is not a subgroup.

(c) If C_i and C_j are M-classes of M_1 and M_2, respectively,
then $C_i \times C_j$ need not be an M-class of

$$M_1 \overset{M}{\times} M_2.$$

Again this is because of the absence of elements (a_1, e) and (e, a_2) from

$$M_1 \overset{M}{\times} M_2.$$

However, each direct product of M-classes splits into at most two M-classes of

$$M_1 \overset{M}{\times} M_2.$$

For example, in

$$31' \overset{M}{\times} 31'$$

the product $\{C_3^+, C_3^-\} \times \{C_3^+, C_3^-\}$ gives the two M-classes $\{(C_3^+, C_3^+), (C_3^-, C_3^-)\}$ and $\{(C_3^+, C_3^-), (C_3^-, C_3^+)\}$. On the other hand, many results are transferable:

(d) The M-direct product is commutative,

$$M_1 \overset{M}{\times} M_2 \cong M_2 \overset{M}{\times} M_1,$$

and associative,

$$(M_1 \overset{M}{\times} M_2) \overset{M}{\times} M_3 = M_1 \overset{M}{\times} (M_2 \overset{M}{\times} M_3).$$

(e) M_1 is naturally M-isomorphic to

$$(M_1 \overset{M}{\times} M_2)/H_2.$$

(f) M is naturally M-isomorphic to the diagonal subgroup of

$$M \overset{M}{\times} M \overset{M}{\times} \cdots \overset{M}{\times} M.$$

So as for ordinary groups the inner direct product may, if desired, be treated by descent in symmetry from the outer direct product.

(g) If d_1 and d_2 are corepresentations of M_1 and M_2, respectively, then $d = d_1 \times d_2$ is a corepresentation of

$$M = M_1 \overset{M}{\times} M_2.$$

From (c), irreducibility of d_1 and d_2 does not necessarily imply irreducibility of $d = d_1 \times d_2$ as the number of M-classes may increase. The ICRs of

$$M_1 \overset{M}{\times} M_2$$

are, however, easily obtained:

Theorem 5.2: Let d_1 and d_2 be ICRs of M_1 and M_2, respectively, and $d = d_1 \times d_2$ be a corepresentation of

$$M = M_1 \overset{M}{\times} M_2.$$

(a) If d_1 is of type (a), then d is irreducible and of the same type as d_2.

(b) If d_1 and d_2 are both of type (b) then d is reducible to four equivalent ICRs of type (a).

(c) If d_1 is of type (b) and d_2 of type (c) then d is reducible to two equivalent ICRs of type (c).

(d) If d_1 and d_2 are both of type (c) then d is reducible to two inequivalent ICRs D_1 and D_2 which have the same degree and are both of type (c). Further, let C_{ij} be the product of M-classes $C_i \times C_j$ of M_1 and M_2, respectively. If C_{ij} is an M-class of M, then D_1 and D_2 have equal characters on C_{ij}. If, however, C_{ij} reduces to two classes C and C', the character of $D_1(D_2)$ on C equals the character of $D_2(D_1)$ on C'.

Proof: A character based proof is possible but not particularly useful for finding the ICRs for

$$M_1 \overset{M}{\times} M_2,$$

as in general a transformation is required to reduce d. Consequently we give a constructive proof based on the definite matrix forms given in the Appendix of N–G. Most of the ICRs so far given in the literature are of this form or differing by a simple transformation.

(a) This is irreducible so no transformation is required.

(b) Let $d_i(u_i) = \begin{pmatrix} \Delta_i(u_i) & 0 \\ 0 & \Delta_i(u_i) \end{pmatrix}$ and

$$d_i(a_0^i) = \begin{pmatrix} 0 & P_i \\ -P_i & 0 \end{pmatrix}$$

with Δ_i an IR of H_i. Then d is equivalent to $d' = 4D$, where D is the ICR of type (a),

$$D((u_1, u_2)) = \Delta_1(u_1) \times \Delta_2(u_2)$$

and

$$D((a_0^1, a_0^2)) = P_1 \times P_2.$$

(c) Let $d_1(u_1) = \begin{pmatrix} \Delta_1(u_1) & 0 \\ 0 & \Delta_1(u_1) \end{pmatrix}, \quad d_1(a_0^1) = \begin{pmatrix} 0 & P_1 \\ -P_1 & 0 \end{pmatrix}$

and

$$d_2(u_2) = \begin{pmatrix} \Delta_2(u_2) & 0 \\ 0 & \Delta_3(u_2) \end{pmatrix}, \quad d_2(a_0^2) = \begin{pmatrix} 0 & P_2 \\ P_3 & 0 \end{pmatrix}.$$

Then d is equivalent to $d' = 2D$, where D is the ICR of type (c),

$$D((u_1, u_2)) = \begin{pmatrix} \Delta_1(u_1) \times \Delta_2(u_2) & 0 \\ 0 & \Delta_1(u_1) \times \Delta_3(u_2) \end{pmatrix},$$

$$D((a_0^1, a_0^2)) = \begin{pmatrix} 0 & P_1 \times P_3 \\ -P_1 \times P_2 & 0 \end{pmatrix}.$$

(d) Let $d_i(u_i) = \begin{pmatrix} \Delta_i(u_i) & 0 \\ 0 & \Delta_i'(u_i) \end{pmatrix}$ and $d_i(a_0^i) = \begin{pmatrix} 0 & P_i \\ P_i' & 0 \end{pmatrix}$.

Then d is equivalent to $d' = D_1 \oplus D_2$, where D_1 and D_2 are the two ICRs of type (c),

$$D_1((u_1, u_2)) = \begin{pmatrix} \Delta_1(u_1) \times \Delta_2'(u_2) & 0 \\ 0 & \Delta_1'(u_1) \times \Delta_2(u_2) \end{pmatrix},$$

$$D_1((a_0^1, a_0^2)) = \begin{pmatrix} 0 & P_1 \times P_2' \\ P_1' \times P_2 & 0 \end{pmatrix}$$

and

$$D_2((u_1, u_2)) = \begin{pmatrix} \Delta_1(u_1) \times \Delta_2(u_2) & 0 \\ 0 & \Delta_1'(u_1) \times \Delta_2'(u_2) \end{pmatrix},$$

$$D_2((a_0^1, a_0^1)) = \begin{pmatrix} 0 & P_1 \times P_2 \\ P_1' \times P_2' & 0 \end{pmatrix}.$$

The second part of (d) follows by the equality of the traces of $D_1((u_1, u_2))$ and $D_2((u_1, au_2^{-1}a^{-1}))$.

We still have to show that this gives *all* ICRs of

$$M_1 \overset{M}{\times} M_2.$$

Firstly, if an ICR D is contained in both $d_1 \times d_2$ and $d_3 \times d_4$

then $d_1 = d_3$ and $d_2 = d_4$ by nonequivalence of characters in M_1 and M_2. Secondly, Theorem 14 of N-G related the degrees and intertwining numbers of all ICRs of a magnetic group to its order, and by calculating degrees and intertwining numbers of the ICRs of

$$M_1 \overset{M}{\times} M_2$$

obtained from those of M_1 and M_2,

Theorem 5.3: Each ICR of

$$M = M_1 \overset{M}{\times} M_2$$

is a component of $d_1 \times d_2$ for some ICRs d_1 and d_2 of M_1 and M_2, respectively.

Restating all this in terms of characters

Corollary 5.4: Let ψ_1 and ψ_2 be irreducible characters of M_1 and M_2 with intertwining numbers I_1 and I_2, respectively, and let $\chi = \psi_1 \psi_2$ be a character of

$$M = M_1 \overset{M}{\times} M_2 .$$

(a) If $I_1 = 1$ then χ is irreducible with intertwining number $I = I_2$. (Of course, the subscripts "one" and "two" may be interchanged throughout).

(b) If $I_1 = I_2 = 4$ then $\chi' = \chi/4$ is irreducible with intertwining number one.

(c) If $I_1 = 4$ and $I_2 = 2$ then $\chi' = \chi/2$ is irreducible with intertwining number two.

(d) If $I_1 = I_2 = 2$ then $\chi = \chi' + \chi''$, where χ' and χ'' are both irreducible of the same degree with intertwining number two. Further, $\chi'((u_1, u_2)) = \chi''((u_1, a u_2^{-1} a^{-1}))$.

Example: *4′ has three ICRs A, DE, and DB with intertwining numbers one, two, and four, respectively (the character table appears in Table II). The M-classes of

$$*4' \overset{M}{\times} *4'$$

are $C_1 = \{(E, E)\}$, $C_2 = \{(C_{2z}, E), (\bar{C}_{2z}, E)\}$, $C_3 = \{(E, C_{2z}), (E, \bar{C}_{2z})\}$, $C_4 = \{(\bar{E}, E)\}$, $C_5 = \{(E, \bar{E})\}$, $C_6 = \{(C_{2z}, \bar{E}), (C_{2z}, \bar{E})\}$, $C_7 = \{(\bar{E}, C_{2z}), (\bar{E}, \bar{C}_{2z})\}$, $C_8 = \{(C_{2z}, C_{2z})\}$, $C_9 = \{(C_{2z}, \bar{C}_{2z}), (\bar{C}_{2z}, C_{2z})\}$, and $C_{10} = \{(\bar{E}, \bar{E})\}$. Only one direct product of M- classes splits, to $C_8 \oplus C_9$. $A \times A, A \times D\bar{E}, A \times DB, D\bar{E} \times A$, and $DB \times A$ are all irreducible. $(DB \times DB)/4$ is irreducible with $I = 1$, and $(DE \times DB)/2$ and $(DB \times DE)/2$ are both irreducible with $I = 2$. $D\bar{E} \times D\bar{E} = D_1 \oplus D_2$, both irreducible with $I = 2$. They have characters χ_1 and χ_2 (respectively) equal on C_1 through C_7, and C_{10}. $\chi_1(C_8) = \chi_2(C_9) = a$ and $\chi_1(C_9) = \chi_2(C_8) = b$. Since the character of $D\bar{E}$ on C_{2z} is zero, $a = -b$. Row or column orthogonality fixes $|a| = 2$. To determine the argument of a, additional information appears to be necessary. For example, C_8 and C_9 are ambivalent and so a is real (see next section). Alternatively, $a^2 = 4$ follows by the class multiplication rule [Theorem 4.6, part (e)].

The major problem with direct product groups is that of *identifying* when a group is a direct product. One case is always easy to spot, though: a group containing the inversion group $\bar{1} = \{E, I\}$.

Corollary 5.5: Let M contain the linear subgroup $\bar{1}$. Then

$$M \cong M' \overset{M}{\times} \bar{1}1' ,$$

where $M' = (G', H') = M/\bar{1}$. To each ICR D of M' corresponds exactly two ICRs D_g and D_u of M with equal matrices on (H', E) and opposite matrices (in sign) on (H', I).

As in representation theory, such "inversion" magnetic groups may now be dealt with trivially from the "noninversion" groups. In Sec. 3 the number of families of magnetic point groups requiring separate calculations was reduced to 33. Eliminating now the inversion groups leaves only 16.

6. SPECIAL GROUPS AND ICRs

Groups with only one-dimensional ICRs have, of course, a particularly simple character theory (inner direct products, for example, are trivial). If the degree of an ICR is only one, the intertwining algebra can only be $\mathbb{R}$ and so the intertwining number must be one. If all ICRs have degree one, from Theorem 14 of N-G the number of ICRs, which is also the number of M-classes, equals the order of the linear subgroup. Every M-class consequently has only one element and the group must satisfy the relations

$$u_1 u_2 = u_2 u_1, \quad \forall u_1, u_2 \in H$$

and

$$a u = u^{-1} a, \quad \forall u \in H, a \in G - H .$$

Conversely, this guarantees that ICRs have degree one. While H is abelian, G is in general nonabelian [for example, $M = (D_n, C_n)$ for all $n \geqslant 1$ satisfies the relations]. Abelian G may have two-dimensional ICRs (for example, 4′).

To calculate the number of one-dimensional ICRs for general M we need the commutator subgroup.

Definition 6.1: The M-commutator subgroup M' is the subgroup of M generated by

$$\{u_1^{-1} u_2^{-1} u_1 u_2, a^{-1} u_1 a u_1 : u_1, u_2 \in H, a \in G - H\}.$$

Lemma 6.2: M' is an M-normal subgroup of M.

Proof:

(a) $u(u_1^{-1} u_2^{-1} u_1 u_2) u^{-1} \in M'$,

(b) $$u(a^{-1} u_1 a u_1) u^{-1} = [(au^{-1})^{-1} u_1 (au^{-1}) u_1] \cdot [u_1^{-1} u u_1 u^{-1}] \in M',$$

(c) $$a^{-1}(u_1^{-1} u_2^{-1} u_1 u_2) a = [(u_1 a)^{-1} u_2^{-1} (u_1 a) u_2^{-1}] \cdot [(a u_2^{-1})^{-1} u_2 (a u_2^{-1}) u_2] \in M' ,$$

(d) $$a^{-1}(a_1^{-1} u a_1 u) a = [(a_1 a)^{-1} u (a_1 a) u^{-1}] \cdot [(au^{-1})^{-1} u (au^{-1}) u] \in M'.$$

The proof then follows that for ordinary groups.

Theorem 6.3: M' is the minimal normal subgroup L such that M/L possesses only one-dimensional ICRs.

Corollary 6.4: The number of one-dimensional ICRs is

$$\tfrac{1}{2}[M:M'] = |H|/|M'|.$$

Recently Butler *et al.*[8,9,25–28] have developed and used a recursive method for generating $6j$ and $3jm$ tensors for

TABLE II. Selected character tables of the magnetic point groups. The group appears on the upper left with the ICR labels beneath. In the middle, the classes are listed along the top with the characters beneath. To the right are successively the intertwining number I, the Frobenius–Schur invarant c, and n, the minimal power for the occurence of each ICR in an (arbitrarily chosen) faithful ICR.

(a)

$*4'$	E	$C_{2z},\bar{C}_{2z}$	$\bar{E}$	I	c	n
A	1	1	1	1	1	0
DB	2	-2	2	4	2	2
$D\bar{E}$	2	0	-2	2	0	1

(b)

$*4'22'$	E	$\bar{E}$	$C_{2x,y},\bar{C}_{2x,y}$	$C_{2z},\bar{C}_{2z}$	I	c	n
A	1	1	1	1	1	1	0
DE	2	2	0	-2	2	2	2
B_1	1	1	-1	1	1	1	2
$\bar{E}$	2	-2	0	0	1	-1	1

(c)

$*41'$	E	$\bar{E}$	$C_{2z},\bar{C}_{2z}$	$C_{4z}^{\pm}$	$\bar{C}_{4z}^{\pm}$	I	c	n
A	1	1	1	1	1	1	1	0
B	1	1	1	-1	-1	1	1	4
DE	2	2	-2	0	0	2	0	2
$D\bar{E}_1$	2	-2	0	$\surd 2$	$-\surd 2$	2	0	1
$D\bar{E}_2$	2	-2	0	$-\surd 2$	$\surd 2$	2	0	3

(d)

$*31'$	E	$\bar{E}$	$C_3^{\pm}$	$\bar{C}_3^{\pm}$	I	c	n
A	1	1	1	1	1	1	0
DE	2	2	-1	-1	2	0	2
$D\bar{A}$	2	-2	-2	2	4	2	3
$D\bar{E}$	2	-2	1	-1	2	0	1

(e)

$*61'$	E	$\bar{E}$	$C_6^{\pm}$	$\bar{C}_6^{\pm}$	$C_3^{\pm}$	$\bar{C}_3^{\pm}$	$C_2,\bar{C}_2$	I	c	n
A	1	1	1	1	1	1	1	1	1	0
B	1	1	-1	-1	1	1	-1	1	1	6
DE_1	2	2	-1	-1	-1	-1	2	2	0	4
DE_2	2	2	1	1	-1	-1	-2	2	0	2
$D\bar{E}_1$	2	-2	0	0	-2	2	0	2	0	3
$D\bar{E}_2$	2	-2	$-\surd 3$	$\surd 3$	1	-1	0	2	0	1
$D\bar{E}_3$	2	-2	$\surd 3$	$-\surd 3$	1	-1	0	2	0	5

(f)

$*321'$	E	$\bar{E}$	$C_3^{\pm}$	$\bar{C}_3^{\pm}$	$C'_{21,2,3},\bar{C}'_{21,2,3}$	I	c	n
A_1	1	1	1	1	1	1	1	0
A_2	1	1	1	1	-1	1	1	2
E	2	2	-1	-1	0	1	1	2
$\bar{E}_1$	2	-2	1	-1	0	1	-1	1
$D\bar{E}$	2	-2	-2	2	0	2	0	3

(g)

$*231'$	E	$\bar{E}$	$C_{2x,y,z},\bar{C}_{2x,y,z}$	$C_{31,2,3,4}^{\pm}$	$\bar{C}_{31,2,3,4}^{\pm}$	I	c	n
A	1	1	1	1	1	1	1	0
DE	2	2	2	-1	-1	2	0	4
T	3	3	-1	0	0	1	1	2
$\bar{E}$	2	-2	0	1	-1	1	-1	1
$D\bar{F}$	4	-4	0	-1	1	2	0	3

groups of linear operators. When certain problems regarding the $6j$ tensor for grey groups have been resolved[29] it is likely that the method can be adapted to magnetic groups. It is one of the few which involve properties of faithful representations and in anticipation of future use.

Theorem 6.5 [Burnside–Brauer (Ref. 21)]: Let χ be a faithful character of a magnetic group $M = (G,H)$ and suppose $\chi(u)$ takes on exactly m different values for $u \in H$. Then every irreducible character of M occurs in the nth inner Kronecker power χ^n for $0 \leqslant n < m$.

In the accompanying character tables a faithful ICR is given wherever possible and the minimum value of n. A faithful ICR has kernel $\{e\}$ and hence $\chi(u) \neq \chi(e)$ for $u \neq e$.

Finally, in this section, we consider real-valued characters.

Definition 6.6: An M-class C is ambivalent if for each $u \in C$ its inverse is also in C. Alternatively, if u is any arbitrary element of C, then there exists $u_1 \in H$ with $u_1 u u_1^{-1} = u^{-1}$ or

$a \in G - H$ with $au = ua$.

Every grey group (i.e., containing the commuting operator θ) has every M-class ambivalent.

Theorem 6.7: The number of ICRs of a magnetic group with real character equals the number of ambivalent M-classes.

Thus the grey groups have only real characters.

7. THE INTERTWINING NUMBERS

The intertwining numbers have been seen to play a crucial role in the character theory of magnetic groups. It we already know the character table then of course the intertwining numbers are already known. However, as in the last section it is of interest to see if information can be obtained purely from group or class properties, particularly if the character table has not been determined. Here is a simple result.

Theorem 7.1: The number of ICRs with intertwining numbers one or four equals the number of M-classes C_i with the following property: for any $u \in C_i$ there exists $a \in G - H$ such that $au = u^{-1}a$.

Proof: Let u_1, u_2 be arbitrary elements of C_i and $a_1 \in G - H$ with $a_1 u_1 = u^{-1} a_1$. Then u_1 and u_2 are equivalent by a linear element. To see this, suppose they are equivalent by a nonlinear element a_2:

$$u_1 = a_2 u_2^{-1} a_2^{-1}.$$

By substitution,

$$(a_1 a_2) u_2^{-1} a_2^{-1} = u_1^{-1} a_1$$

or

$$u_2 (a_1 a_2)^{-1} = (a_1 a_2)^{-1} u_1$$

so they are equivalent by a linear element. It is readily checked that $au = u^{-1}a$ is a class property independent of the choice of u. Hence any such M-class remains irreducible on restriction to ordinary classes of H. Conversely, if an M-class does not possess this property then it branches into two ordinary classes of H. But from the relations between IRs of H and ICRs of M (N-G, Appendix) the number of ICRs with intertwining number two equals the number of M-classes which split on H, and hence the number with intertwining number one or four is the number of irreducible M-classes on H as required.

This theorem completely determines the number of ICRs with $I = 2$ by M-class properties. The problem of deciding between the number with $I = 1$ and the number with $I = 4$ is much more complex. For example, for a grey group it becomes the calculation of the number of IRs of the first and second kinds, respectively. van Zanten and de Vries[30] and Gow[31] have given various lower bounds for these, but only in certain cases are there presently exact solutions.

For the remainder of this section we aim at a special case, namely when all ICRs have $I = 1$. In this case the character theory of the magnetic group reduces to that of the linear subgroup and, especially for magnetic point groups, this may be very well known. Extensions along the lines of van Zanten and de Vries[30] will be obvious.

Definition 7.2: Let $\zeta^{(2)}(u)$ be the number of square roots of u in G-H.

Lemma 7.3: $\zeta^{(2)}$ is an M-class function.

Lemma 7.4: $\zeta^{(2)} = \Sigma_i c_i \chi_i$, where $c_i = 1$ if $I_i = 1$, $c_i = -\frac{1}{2}$ if $I_i = 4$, and $c_i = 0$ if $I_i = 2$.

Proof: From N-G, row orthogonality gives

$$c_i = \frac{1}{I_i |H|} \sum_u \zeta^{(2)}(u) \chi_i(u)^*.$$

But

$$\zeta^{(2)}(u) \chi_i(u)^* = \sum_{a \in G - \mathrm{H}: a^2 = u} \chi_i(a^2),$$

so

$$c_i = \frac{1}{I_i |H|} \sum_a \chi_i(a^2).$$

Substituting by Eqs. (20), (24), and (28) of N-G the result follows.

Theorem 7.5: All ICRs of a magnetic group have intertwining number one iff

$$\zeta^{(2)}(e) = \sum_i \chi_i(e).$$

Proof: Immediate from the possible values of c_i.

Corollary 7.6: Let the set of irreducible characters of M be ICR(M) and the set of linear irreducible characters of H be Irr(H). Then ICR(M) = Irr(H) iff

$$\zeta^{(2)}(e) = \sum_i \varphi_i(e) \quad \text{for } \varphi_i \in \mathrm{Irr}(H).$$

Proof: If ICR(M) = Irr(H) then all ICRs of M must have intertwining numbers of one to avoid branching, and hence the previous theorem applies with $\varphi \in$ Irr(H) replacing $\psi \in$ ICR(M).

Conversely, suppose

$$\zeta^{(2)}(e) = \sum_i \varphi_i(e).$$

Break this up into a sum over j [IRs inducing ICRs of type (a)], k [IRs inducing ICRs of type (b)], and 1 [IRs inducing ICRs of type (c)]:

$$\zeta^{(2)}(e) = \sum_j \varphi_j(e) + \sum_k \varphi_k(e) + \sum_l \varphi_l(e).$$

By the relations between IRs and ICRs this is

$$\zeta^{(2)}(e) = \sum_j \chi_j(e) + \frac{1}{2} \sum_k \chi_k(e) + \frac{1}{2} \sum_l \chi_l(e).$$

But we know from Lemma 7.4 that

$$\zeta^{(2)}(e) = \sum_j \chi_j(e) - \frac{1}{2} \sum_k \chi_k(e).$$

and as the sums over k and l are nonnegative they must vanish. Hence all intertwining numbers are one and ICR(M) = Irr(H).

Example: The group *6′2′2 has antilinear elements $\{\theta C_6^{\pm}, \theta \bar{C}_6^{\pm}, \theta C_2, \theta \bar{C}_2, \theta C''_{21,2,3}, \theta \bar{C}''_{21,2,3}\}$ and eight of these square to the identity. The linear subgroup *32 has six IRs and the sum of their degrees is also eight. Hence *6′2′2 has the same character table as *32.

The character theory of this type of group follows from the linear group and may be found in many places.[1,5,9] Eliminating these from the remaining 16 families of magnetic groups leaves only seven families—a very manageable number! Their character tables are given in Table II.

8. COMPLEX CONJUGATES OF ICRs WITH REAL CHARACTER

As with representations, a corepresentation is equivalent to its complex conjugate iff it has real character. The row orthogonality relations of N–G give an immediate character test

$$\sum_u (\chi(u))^2 \neq 0 \quad \text{iff } D \cong D^*.$$

For linear groups the well-known Frobenius–Schur invariant[21] divides IRs with real characters into orthogonal IRs ($c = 1$) and symplectic IRs ($c = -1$). This division is of great importance for Racah algebra methods of linear groups as it completely determines the $1-j$ phase which is required for, amongst other things, permutation properties of the $6j$ tensor.[6,7] (For complex IRs the phase is undetermined. However, the concept of quasiambivalence[32] has proved useful for a partial determination of the phase.[7,33]) Newmarch and Golding[29] have similarly found the $1-j$ phase important for the Racah algebra of grey groups but have noted that for ICRs of types (b) or (c) of these groups the phase is not uniquely determined.[34] Thus the relations between complex conjugates and to the $1-j$ phase deserves further investigation.

For the remainder of this section, D will be a unitary ICR with real character, D^* the ICR with matrices complex conjugate to D (that D^* is an ICR is easily shown) and $\mathbf{p}$ the set of matrices giving equivalence of D to D^*:

$$\mathbf{p} = \{ P : PD(u) = D(u)^*P, PD(a) = D(a)^*P^* \forall u,a \in M \}.$$

$\mathbf{m}$ is the commutator algebra of D, i.e., the set of all matrices commuting with D.

These are all simple generalizations of results on representations:

Lemma 8.1:

(a) If P is any element of $\mathbf{p}$ and M any element of $\mathbf{m}$ then both PM and M^*P are elements of $\mathbf{p}$.

(b) If P, Q are two elements of $\mathbf{p}$, there exist $M,M' \in \mathbf{m}$ such that $P = QM' = M^*Q$.

(c) If $P \in \mathbf{p}$, $P^*P \in \mathbf{m}$.

While not affecting his conclusions, Rudra[35] makes an error in stating $P^*P = \lambda E$, as can be shown by example (a similar error is made by Kotzev and Aroyo[36] in connection with isoscalars). Consider the two-dimensional ICR of 4′ generated by

$$D(\theta C_{4z}^+) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

From the reality of D, $\mathbf{p} = \mathbf{m}$ and the most general form of $P \in \mathbf{p}$ is

$$P = \begin{pmatrix} z_1 & z_2 \\ -z_2^* & z_1^* \end{pmatrix}.$$

Trivially, $P^*P \neq E$ except for special cases. The meat of this section is that such special cases *must* occur.

Theorem 8.2: Let D be a unitary ICR with intertwining number I. Then there exist $P_0,P_1,\ldots,P_{I-1} \in \mathbf{p}$ such that

(a) $P_i = P_0 M_i$, where the M_i form a group of

$$\mathbf{m}\{ \pm E, \pm M_1, \pm M_2, \ldots \},$$

(b) $P_i^*P_i = c_i E$ with $c_i^2 = 1$, $i = 0,1,\ldots,I-1$,

(c) if $I = 4$, $c_0c_1c_2c_3 = -1$.

Proof: It is sufficient to take D of type (b) with intertwining number four as the other two types follow as special cases. Choose any unitary $P \in \mathbf{p}$ and set $M = P^*P$. M is also unitary and may be written

$$M = x_1E + x_2M_1$$

with x_1,x_2 real, $x_1^2 + x_2^2 = 1$, and $M_1 \in \mathbf{m}' = \mathbf{m} - \{\lambda E: \lambda \in \mathbb{R}\}$ with $M_1^2 = -E$. Suppose $x_2 \neq 0$. Then M possesses an inverse square root

$$M^{-1/2} = \frac{\sqrt{1+x_1}}{\sqrt{2}} E - \frac{\sqrt{1-x_1}}{\sqrt{2}} M_1.$$

From $P^*P = M$ and $PP^* = M^*$, $PM = M^*P$ and so $PM_1 = M_1^*P$. Hence

$$PM^{-1/2} = M^{-(1/2)*}P.$$

If now we set

$$P_0 = PM^{-1/2} \text{ and } P_1 = PM^{-1/2}M_1$$

it follows that

$$P_0^* P_0 = E \text{ and } P_1^* P_1 = -E.$$

Continuing with this case of $x_2 \neq 0$, by a suitable $4-D$ rotation in $\mathbf{m}$ M_1 can be taken as an element of the group of $\mathbf{m} : \{ \pm E, \pm M_1, \pm M_2, \pm M_1M_2\}$ with M_2 arbitrarily lying in a plane orthogonal to E and M_1. Set $P_2 = P_0 M_2$ and define $M' \in \mathbf{m}$ by

$$M' = P_2^* P_2.$$

A simple equation relates M' and M_1. Consider

$$(P_0 M_1M_2)^* (P_0 M_1M_2) = (P_0^*M_1^*P_0)(P_0^*M_2^*P_0)M_1M_2$$

$$\text{as } P_0 P_0^* = E$$

$$= -M_1^{-1}M'M_2^{-1}M_1M_2$$

$$\text{as } P_1^*P_1 = P_0^*M_1^*P_0 M_1 = -E$$

$$\text{and } P_0^*M_2^*P_0 = P_2^*P_2M_2^{-1} = M'M_2^{-1}$$

$$= -M_1M'M_1 \text{ as } M_1M_2 = -M_2M_1 \text{ and } M_2^2 = -E.$$

But this also equals

$$(P_0 M_2M_1)^*(P_0 M_2M_1) = M'$$

in a similar manner. Writing M' as a linear combination of M_1,M_2, and M_1M_2 and equating these gives

$$M' = y_1E + y_2M_1.$$

By unitarity of all matrices, $y_1^2 + y_2^2 = 1$. If in this equation $y_2 = 0$ with $y_1 = \pm 1$, set

$$P_3 = P_0 M_1M_2.$$

Then $P_2^*P_2 = P_3^*P_3 = y_1E$ and indeed, for all real linear combinations $P' = z_1P_2 + z_2P_3$ with $z_1^2 + z_2^2 = 1$,

$$P'^*P' = y_1E.$$

If, however, $y_2 \neq 0$, a contradiction rapidly follows. For then taking the inverse square root of M' as with M, a real linear combination P_2' of P_2 and P_3 exists with $P_2'^* P_2' = yE$ and this P_2' may be used in place of P_2. From this it follows that y_2 must have been zero after all.

The case excluded so far was of $x_2 = 0$. However, either for all $P' \in \mathbf{p}$ the corresponding x_2' is zero, in which case there is nothing to show, or there is at least one for which $x_2' \neq 0$ and this P' may be used in place of P in the above analysis.

For an ICR of type (a), $\mathbf{m} \cong \mathbb{R}$ and there is nothing to show. For an ICR of type (b) $m \cong \mathbb{C}$ and either all $P \in \mathbf{p}$ satisfy $P^*P = cE$ or P_0 and P_1 can be constructed as above.

Part (c) follows from the equation

$$c_0 P_3^* P_3 = -P_0^* M_1^* P_0 P_0^* M_2^* P_0 M_2 M_1$$

to complete the proof.

Character tests may be established in a fairly straightforward manner from the orthogonality relations for ICRs given in N-G Sec. 3.

Theorem 8.3: Let D be a unitary ICR equivalent to D^*, and let P_i, c_i be as in the preceding theorem. Then

$$\frac{1}{|H|} \sum_u \chi(u^2) = \sum_i c_i.$$

In conjunction with $c_i^2 = 1$ and $c_0 c_1 c_2 c_3 = -1$ for ICRs of type (b), this shows that the c_i are essentially determined by character theory alone, independent of any specific choice of P_i or of the basis for D. As usual, $c_i = 1$ means P_i is symmetric, $c_i = -1$ means P_i is antisymmetric. Setting the Frobenius–Schur invariant to be

$$c = \sum c_i$$

gives

Corollary 8.4: Let D be a unitary ICR equivalent to D^*.

(a) If D is of type (a), $c = c_0 = \pm 1$.

(b) If D is of type (b), then $c = \pm 2$. If $c = 2$, three of the c_i are positive and one negative, whereas if $c = -2$,

three of the c_i are negative, one positive.

(c) If D is of type (c), $c = 2, 0, -2$. If $c = 2$, $c_0 = c_1 = 1$, if $c = -2$, $c_0 = c_1 = -1$, and if $c = 0$, c_0 and c_1 are of opposite sign.

It can be seen that for all type (b) and some type (c) ICRs there is a freedom in the choice of $1-j$ phase which does not exist for linear groups. For quasiambivalent linear groups a useful simplification for Racah methods is that the product of three $1-j$ phases is one whenever the triple product of IRs contains the identity IR.[7,8,32,33] By considering, for example, *31' and *4' it can be verified that the product of phases is not unity for all choices in magnetic groups even when the character is real. Whilst we do not wish to pursue this here in any depth, we do note a special case of particular relevance to grey groups: if θ is some antilinear element of a magnetic group M which commutes with all elements of the group and for which $D(\theta^2) = \pm E$ for all ICRs of M, then as in Newmarch and Golding,[20] $D(\theta) \in \mathbf{p}$ for each ICR. As

$$D(\theta^2) = D(\theta) D(\theta)^*,$$

$$(D_1(\theta) \otimes D_2(\theta))^* (D_1(\theta) \otimes D_2(\theta))$$
$$= D_1(\theta^2)^* \otimes D_2(\theta^2)^* = \pm E.$$

Hence for any D_3 in this direct product, $c_1 c_2 c_3 = 1$.

Another well-known property of the Frobenius–Schur invariant is its relation to the multiplicity of the identity IR in the symmetrized and antisymmetrized Kronecker squares $D^{[2]}$ and $D^{[1^2]}$ of an IR. From Eq. (20), (23), and (27) and Sec. 5 of N-G the Frobenius–Schur invariant for magnetic groups similarly characterizes these multiplicities for ICRs with real character. The results are summarized in Table III, from which it may be observed that the occurrence of the identity ICR in the symmetrized (antisymmetrized) Kronecker square equals the number of c_i with value one (minus one).

Finally, a word about matrix forms. If $P \in \mathbf{p}$ with $P^*P = E$ is symmetric then exactly as for linear groups, D is equivalent to a real ICR.[3] Similarly, if $P \in \mathbf{p}$ with $P^*P = -E$ is antisymmetric, then D is equivalent to a symplectic ICR.[4,34,37] Any type (b) and some type (c) ICRs (with $c = 0$) with real character are equivalent to both real and symplectic ICRs. For example, consider the ICR of type (c) with $c = 0$, DE of 41'. Constructing the ICR in the usual way from the linear subgroup gives

$$D(C^{-}_{4z}) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad \text{and} \quad D(\theta) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A symmetric $P \in \mathbf{p}$ is

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$
$$\times \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} = r^{-1} \omega r.$$

Transforming D by $\omega_1 r$, where $\omega_1^2 = \omega$, gives

$$D'(C^{-}_{4z}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad D'(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

which is real. On the other hand, transforming D by λE with $\lambda^2 = i$ gives

$$D''(C^{-}_{4z}) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad \text{and} \quad D''(\theta) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

which is in symplectic form. These provide alternative

TABLE III. The multiplicity of the identity ICR 1 in symmetrized and antisymmetrized Kronecker squares of ICRs with real character.

Type of ICR	Frobenius–Schur invariant	Multiplicity of 1 in $D^{[2]}$	Multiplicity of 1 in $D^{[1^2]}$
(a)	1	1	0
	-1	0	1
(b)	2	3	1
	-2	1	3
(c)	2	2	0
	0	1	1
	-2	0	2

"standard" forms to the one obtained by induction from the linear subgroup.

9. CONCLUSION

In a single paper it is, of course, impossible to consider all aspects of character theory used for linear groups and we have singled out a few of general interest. They should be sufficient, however, to show that character theory is a viable tool for the examination of magnetic groups.

[1] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids* (Clarendon, Oxford, 1972).

[2] *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1968).

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

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

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

[6] J.-R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).

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

[8] P. H. Butler and B. G. Wybourne, Int. J. Quantum Chem. **10**, 581 (1976).

[9] P. H. Butler, *Point Group Symmetry Applications: Methods and Tables* (Plenum, New York, 1981).

[10] N. B. Backhouse, Physica **70**, 503 (1973).

[11] J. D. Newmarch and R. M. Golding, "The Character Table for the Corepresentations of Magnetic Groups," J. Math. Phys. **23**, 695 (1982).

[12] D. B. Litvin and W. Opechowski, Physica **76**, 538 (1974).

[13] D. B. Litvin, Acta Crystallogr. **A33**, 279 (1977).

[14] M. Vujičić, I. B. Božović, and F. Herbut, J. Phys. A. **10**, 1271 (1977).

[15] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics and All That* (Benjamin, New York, 1964).

[16] M. Hall, Jr. and J. K. Senior, *The Groups of Order* $2^n(n\leqslant6)$' (Macmillan, New York, 1964).

[17] T. Janssen, J. Math. Phys. **13**, 342 (1972).

[18] E. Ascher and A. Janner, Acta Crystallogr. **18**, 325 (1965).

[19] F. Herbut, M. Vujičić, and Z. Papadopolos, J. Phys. A **13**, 2577 (1980).

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

[21] I. Martin Isaacs, *Character Theory of Finite Groups* (Academic, New York, 1976).

[22] L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967).

[23] A. J. van Zanten and E. de Vries, Physica **49**, 532 (1970).

[24] A. P. Cracknell, Prog. Theor. Phys. **35**, 196 (1966).

[25] P. H. Butler, Int. J. Quantum Chem. **10**, 599 (1976).

[26] P. H. Butler and B. G. Wybourne, Int. J. Quantum Chem. **10**, 615 (1976).

[27] P. H. Butler, R. W. Haase, and B. G. Wybourne, Aust. J. Phys. **31**, 131 (1978).

[28] P. H. Butler and M. F. Reid, J. Phys. A **12**, 1655 (1979).

[29] J. D. Newmarch and R. M. Golding, J. Math. Phys. **24**, 441 (1983).

[30] A. J. van Zanten and E. de Vries, Can. J. Math. **27**, 528 (1975).

[31] R. Gow, J. Algebra **40**, 258 (1976).

[32] G. Mackay, Pac. J. Math. **8**, 503 (1958).

[33] P. H. Butler and R. C. King, Can. J. Math. **26**, 328 (1974).

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

[35] P. Rudra, J. Math. Phys. **15**, 2031 (1974).

[36] J. N. Kotzev and M. I. Aroyo, J. Phys. A **13**, 2275 (1980).

[37] T. Damhus, Linear Algebra Appl. **32**, 125 (1980).

On the $3j$ symmetries

J. D. Newmarch
Department of Physics, Universiti Pertanian Malaysia, Serdang, Selangor, Malaysia [a)]

(Received 4 November 1981; accepted for publication 8 January 1982)

The symmetry properties of the $3jm$ tensor for any finite or compact linear group are discussed using a wreath product construction. This is shown to provide a complete group theoretic explanation for all symmetry properties whether "essential" or "arbitrary." The link with the similar—but distinct—method of inner plethysms is considered.

PACS numbers: 02.20.Hj

1. INTRODUCTION

Many aspects of the so-called Wigner–Racah algebra of angular momentum have been generalized to arbitrary groups, with the most complete generalizations being to compact and finite groups. While the Clebsch–Gordan series, coupling coefficients, and the Wigner–Eckart theorem now form standard components of applied group theory, not as much work has appeared on the $3jm$ tensor based on the various $3j, V, \bar{V}$ symbols of Wigner, Racah, Fano, and Racah, etc. This is largely due to the complexities in dealing both with the permutational properties of the tensor and with its relation to the coupling coefficient. The permutation properties are the reason for dealing with the $3jm$ tensor as they allow condensed tabulation and easier manipulation in equations, whereas the relation to the coupling coefficient allows the tensor to be actually used, mainly through the Wigner–Eckart theorem.

These problems were essentially solved by Derome and Sharp[1,2] in 1965 and 1966, with the first paper detailing (amongst other things) the relation between the $3jm$ and coupling coefficient tensors and the second, the symmetry properties. Their results have formed the basis for further work of both a theoretical and a computational nature.[3–16] However, judging by the number of papers appearing on coupling coefficients either without any mention of permutation properties or with some complex convention, their results are not sufficiently widely known. To improve this and also to make exact the connection with Littlewood's algebra of plethysms used by some authors,[17,18] we give in this paper an alternative derivation of the symmetries of the $3jm$ tensor. The material details a conceptual approach to the problem rather than a more efficient calculation method. Thus while a method is stated for producing symmetrized $3jm$ tensors it is unlikely that it will be used except for special classes of groups (most promising candidate: the symmetric group S_n?).

First, some background. The *coupling coefficient* is defined to be the tensor which reduces an inner direct product of two irreducible representations (irreps) $j_1 \otimes j_2$ of a group G to a third irrep of G. The $3jm$ tensor may be defined in a number of ways, but the cleanest is probably the one used by Fano and Racah[19] for SU(2) which is to reduce the *triple* product $j_1 \otimes j_2 \otimes j_3$ to the trivial irrep 1_G. The invariance of the modulus of the $3jm$ tensor under permutations follows very easily from this for G finite or compact (in addition, it forms the only really workable definition for groups containing antilinear operators[10]). The problem of relating the two tensors may be tackled in two ways: juggle the double product reduction until it becomes a triple product which introduces the $1jm$ or Wigner tensor relating j_3 to j_3^*, or expand the triple product into two double products which introduces the $2jm$ tensor reducing $j_3^* \otimes j_3$ to 1_G. Whatever, we take this problem as solved[3] and merely note three points: (a) The double product leads to the coupling coefficient and is not particularly appropriate for discussing $3jm$ permutations. (b) Relating the two tensors is not trivial as it involves the use of a third tensor. (c) The $3jm$ tensor possesses a (weak) orthogonality property through reducing the triple product, whereas the coupling coefficient possesses a stronger one by reducing only the double product. This strong orthogonality may be transferred to the $3jm$ tensor through (b).

In the Derome and Sharp approach to the $3jm$ symmetries, all possible $3jm$ tensors for the various permutations of irreps are taken and then relations sought between them. This gives a set of permutation matrices called $3j$ tensors [not to be confused with Wigner's $3j$ symbol which is a $3jm$ tensor for SU(2). In SU(2) the $3j$ tensors are just phase factors. A complete list of and explanations for this nomenclature is given in the Appendix]. By counting up the number of independent matrices and exploring their properties, Derome and Sharp were able to detail those symmetry properties which are essential and those which are arbitrary. To some extent this matrix work can be given a group interpretation by noting that the $3j$ matrices generate representations of S_3, S_2, or S_1, but for the last two cases when not all irreps are equivalent this is not sufficient to explain group theoretically all the permutuation properties.

In this paper a complete derivation for all cases is given by transferring the permutations of the $3jm$ tensor to where they act equally naturally but without regard to equivalence or inequivalence of irreps of G, namely, to the direct product group $G \times G \times G$. The permutation action of S_3 on elements of the triple product group defines a semidirect or wreath product group $\Gamma = (G \times G \times G) \ltimes S_3 = G \wr S_3$. This group is discussed in the next section and its irreps are dealt with there and in the following section. These irreps are labelled

---

[a)] Present address: School of Electrical Engineering and Computer Science, University of New South Wales, P.O. Box 1, Kensington, NSW 2033, Australia.

0022-2488/83/040757-08$02.50

quite naturally by irreps of G and of S_1, S_2, or S_3, with the correspondence being the same as for the $3j$ permutation matrices. However, this correspondence arises at an earlier stage.

The $3jm$ tensor is obtained by reducing irreps $j_1 \times j_2 \times j_3$ of $G \times G \times G$ to $1_{G'}$ in the diagonal subgroup $G' = \text{diag } G \times G \times G$; in a similar manner we obtain the $[\lambda]\, r - 3jm$ tensor by reducing irreps of Γ to $1_{\Gamma'}$ in the subgroup $\Gamma' = G' \otimes S_3$ (Sec. 4). Section 5 deals with the permutation properties of the tensor by further descent to S_3 and the relation to the symmetrized $3jm$ tensor by subduction to G' instead.

The method discussed here is not that of the plethysm algebra, although for certain cases there is a strong link. This is discussed in Sec. 6. The paper closes with a possible (though probably trivial) generalization of the symmetrized $3jm$ tensor.

Only linear groups are dealt with; as for linear/antilinear groups the direct product $j_1 \times j_2$ of two irreducible corepresentations is not generally irreducible in the direct product group.[20] The consequences of this will be discussed elsewhere.

A tensor notation is used with implied summation over repeated indices. An inner index (m) in a tensor $T^m{}_n$ corresponds to the row index for the matrix T, while an outer index(n) corresponds to the column index of the matrix. While the "niceties" of this notation are not generally used, the Hermitian adjoint of a unitary tensor $U^m{}_{m'}$ is written $U^{m'}{}_m$ when there is no risk of confusion. The notation is used because quite generally in this area some tricky points get obscured in the simple $3j$ or $\bar{V}$ notations.

2. THE GROUP $G \wr S_3$

Consider a finite or compact group G of elements u with irreducible representations (irreps) j. In order to form inner direct products $j_1 \otimes j_2 \otimes \cdots \otimes j_n$ of G, one way is to first find the irreps $j_1 \times j_2 \times \cdots \times j_n$ of the outer direct product group $G \times G \times \cdots \times G$ and then restrict this group to the diagonal subgroup diag $G \times G \times \cdots \times G$. In discussing the $3jm$ tensor it is possible to use the double product $G \times G$, but this leads directly to the coupling coefficient rather than the $3jm$, and cleaner results are obtained by considering the triple product $G \times G \times G$. The irreps of $G \times G \times G$ are just $j_1 \times j_2 \times j_3$, and on restriction to $G' = \text{diag } G \times G \times G$ are generally reducible. The $3jm$ tensor is nothing more than that part of the unitary matrix which reduces these representations to $1_{G'}$, the trivial representation of G'.

The permutation properties of the $3jm$ tensor were derived by Derome and Sharp by considering S_3 permutations directly on the tensor, but we observe that S_3 acts quite naturally on $G \times G \times G$ by permuting elements $(u_1 u_2 u_3)$. [Strictly, elements of $G \times G \times G$ should be written (u_1, u_2, u_3) but no confusion should arise by omitting the commas.] This action is sufficient to produce the semidirect product group $\Gamma = (G \times G \times G) \otimes S_3$. Definitions of the semidirect product vary in the literature, so we state explicitly the conventions used here. A permutation π is given in the cycle notation, with usual product [e.g., $(12)(123) = (23)$]. Each $\pi \in S_3$ acts automorphically on $G \times G \times G$ by permuting positions (not indices). Thus $(123)\,(u_a u_b u_c) = (u_b u_c u_a)$. The combination law for the permutations is then $\pi_1(\pi_2(u_a u_b u_c)) = (\pi_2 \pi_1) \times (u_a u_b u_c)$. This gives

$$\Gamma = \{(\pi, u_1 u_2 u_3):\ \pi \in S_3,\ u_i \in G\}$$

as a semidirect or wreath product group $\Gamma = (G \times G \times G) \otimes S_3 = G \wr S_3$ with multiplication

$$(\pi_1, u_1 u_2 u_3)(\pi_2, v_1 v_2 v_3) = (\pi_1 \pi_2, \pi_2(u_1 u_2 u_3) \cdot (v_1 v_2 v_3)).$$

To use this group for deriving the $3jm$ symmetries, we next need its irreps. One of the quickest ways of finding them is to start with irreps $j_1 \times j_2 \times j_3$ of $G \times G \times G$ and lift them to Γ by a two-stage process. For the first stage, consider the little group $\mathrm{L}(j_1 \times j_2 \times j_3)$ of $j_1 \times j_2 \times j_3$ in Γ. This is also a semidirect product $(G \times G \times G) \otimes S_n$, where S_n is a subgroup of S_3, and Jansen and Boon (Ref. 21, pp. 157–160) have described the process whereby $j_1 \times j_2 \times j_3$ and the irreps of S_n yield directly the irreps of the little group. These may then be induced to Γ by the usual process and these are irreducible in Γ. Further, all irreps of Γ may be obtained by this process. The details are not particularly exciting and we give the results only. Some notational aspects of the following theorem require explanation: in component form the matrix of $(u_1 u_2 u_3)$ is $j_1(u_1)^{m_1}{}_{n_1} j_2(u_2)^{m_2}{}_{n_2}\, j_3(u_3)^{m_3}{}_{n_3}$. This may be abbreviated to

$$j_1 j_2 j_3 (u_1 u_2 u_3)^{m_1 m_2 m_3}{}_{n_1 n_2 n_3}$$

or to

$$jjj_{(123)}(u_1 u_2 u_3)^{m_1 m_2 m_3}{}_{n_1 n_2 n_3}$$

without confusion. Secondly, given an irrep $[\lambda]$ of a (symmetric) group with matrices $\lambda(\pi)^{r'}{}_{s'}$ induction to a larger group gives a representation with matrices $D_{[\lambda]}(\pi)^{r'r}{}_{s's}$, where the indices r and s label coset representatives of the group.

Theorem: Let $j_1 \times j_2 \times j_3$ be an irrep of $G \times G \times G$ with little group $\mathrm{L}(j_1 \times j_2 \times j_3)$ in $\Gamma = (G \times G \times G) \otimes S_3$, and $S_n = \mathrm{L}(j_1 \times j_2 \times j_3)/(G \times G \times G)$. Let $\{m_r \in S_3\}$ be coset representatives of S_n in S_3, $[\lambda]$ an irrep of S_n, and $D_{[\lambda]} = [\lambda] \uparrow S_3$. Then $D_{[\lambda]\, j_1 j_2 j_3}$ is an irrep of Γ, where

$$D_{[\lambda]\, j_1 j_2 j_3}(\pi, u_1 u_2 u_3)^{r'r m_1 m_2 m_3}{}_{s's n_1 n_2 n_3}$$
$$= D_{[\lambda]}(\pi)^{r'r}{}_{s's}\ jjj_{\pi_s^{-1}(123)}(u_1 u_2 u_3)^{\pi_r^{-1}\pi(m_1 m_2 m_3)}{}_{\pi_s^{-1}(n_1 n_2 n_3)}.$$

Furthermore, all irreps of Γ are of this form.

An alternative method of constructing these irreps is given by Kerber[22] in connection with the plethysm algebra. However, the emphasis in this is different and will be examined in Sec. 6.

3. EQUIVALENT IRREPS IN Γ

It must be admitted that while the theorem is given in a form eminently suitable for generalization, it does not display its salient features at a glance. We take the opportunity in this section to take a more graphic look, and also to give the equivalence transformations we shall allow.

Being heavily dependent on the little group, it is not surprising that the theorem breaks down into a number of cases when examined closer. We deal with each in turn.

A. j_1, j_2, and j_3 all equivalent

The little group here is equal to Γ itself, so the coset representatives are trivial and the indices r,s may be suppressed. $[\lambda]$ is an irrep of S_3 and can be [3], [21], or $[1^3]$. Generators of this irrep are

$$D_{[\lambda]j_1j_1j_1}((12),u_1u_2u_3)^{r'm_1m_2m_3}{}_{s'n_1n_2n_3} = \lambda((12))^{r'}{}_{s'}j_1j_1j_1(u_1u_2u_3)^{m_2m_1m_3}{}_{n_1n_2n_3} \tag{3.1}$$

and

$$D_{[\lambda]j_1j_1j_1}((123),u_1u_2u_3)^{r'm_1m_2m_3}{}_{s'n_1n_2n_3} = \lambda((123))^{r'}{}_{s'}j_1j_1j_1(u_1u_2u_3)^{m_2m_3m_1}{}_{n_1n_2n_3}. \tag{3.2}$$

Any equivalence transformation applied to this has the consequence of transforming $[\lambda]$, j_1, or both. However, for present purposes we are not interested in such transformations and conclude that no nontrivial transformations are allowed.

B. Exactly two of j_1, j_2, and j_3 equivalent

Without loss of generality we may take $j_1 = j_2 \neq j_3$. The little group is the $\mathrm{L}(j_1 \times j_1 \times j_3) = (G \times G \times G) \circledS (S_2 \times S_1)$ so $S_n = S_2$ with $[\lambda] = [2]$ or $[1^2]$ (as these irreps are one-dimensional, the indices r',s' are suppressed) and coset representatives e, (123),(132). Generators are then

$$D_{[\lambda]jjj_3}((12),u_1u_2u_3)^{rm_1m_2m_3}{}_{sn_1n_2n_3} = \lambda((12))\begin{pmatrix} j_1j_1j_3(u_1u_2u_3)^{m_2m_1m_3}{}_{n_1n_2n_3} & 0 & 0 \\ 0 & 0 & j_1j_3j_1(u_1u_2u_3)^{m_1m_3m_2}{}_{n_2n_3n_1} \\ 0 & j_3j_1j_1(u_1u_2u_3)^{m_3m_2m_1}{}_{n_3n_1n_2} & 0 \end{pmatrix}^{r}_{s} \tag{3.3}$$

and

$$D_{[\lambda]j_1j_1j_3}((123),u_1u_2u_3)^{rm_1m_2m_3}{}_{sn_1n_2n_3} = \lambda(e)\begin{pmatrix} 0 & 0 & j_1j_3j_1(u_1u_2u_3)^{m_2m_3m_1}{}_{n_2n_3n_1} \\ j_1j_1j_3(u_1u_2u_3)^{m_1m_2m_3}{}_{n_1n_2n_3} & 0 & 0 \\ 0 & j_3j_1j_1(u_1u_2u_3)^{m_3m_1m_2}{}_{n_3n_1n_2} & 0 \end{pmatrix}^{r}_{s}. \tag{3.4}$$

Again, transformations are restricted by the requirement that they do not alter bases in S_2 and G. In addition we impose the requirement that they do not alter $D_{[\lambda]j_1j_1j_2}(e,u_1u_2u_3)$. This restriction is imposed in order that the particularly simple structure of this irrep is not lost. [It is the diagonal matrix diag. $(j_1j_1j_3, j_3j_1j_1, j_1j_3j_1)(u_1u_2u_3)^{rm_1m_2m_3}{}_{sn_1n_2n_3}$.]

Schur's lemma in $G \times G \times G$ then shows that only diagonal transformations are allowed:

$$\begin{pmatrix} \exp(i\phi_1)I & 0 & 0 \\ 0 & \exp(i\phi_2)I & 0 \\ 0 & 0 & \exp(i\phi_3)I \end{pmatrix}. \tag{3.5}$$

C. None of j_1, j_2, or j_3 equivalent

For this last case the little group is trivial with $S_n = S_1$, and r', s' may again be suppressed. Using the coset representatives e, (12), (13), (23), (123), and (132), the generators are

$$D_{[1]j_1j_2j_3}((12),u_1u_2u_3)^{rm_1m_2m_3}{}_{sn_1n_2n_3} = \begin{pmatrix} 0 & j_2j_1j_3(u_1u_2u_3)^{m_2m_1m_3}{}_{n_2n_1n_3} & 0 & 0 & 0 & 0 \\ j_1j_2j_3(u_1u_2u_3)^{m_1m_2m_3}{}_{n_1n_2n_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & j_2j_3j_1(u_1u_2u_3)^{m_2m_3m_1}{}_{n_2n_3n_1} \\ 0 & 0 & 0 & 0 & j_3j_1j_2(u_1u_2u_3)^{m_3m_1m_2}{}_{n_3n_1n_2} & 0 \\ 0 & 0 & 0 & j_1j_3j_2(u_1u_2u_3)^{m_1m_3m_2}{}_{n_1n_3n_2} & 0 & 0 \\ 0 & 0 & j_3j_2j_1(u_1u_2u_3)^{m_3m_2m_1}{}_{n_3n_2n_1} & 0 & 0 & 0 \end{pmatrix} \tag{3.6}$$

and

$$D_{[1]j_1j_2j_3}((123),\ u_1u_2u_3)^{rm_1m_2m_3}{}_{sn_1n_2n_3} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & j_2j_3j_1(u_1u_2u_3)^{m_2m_3m_1}{}_{n_2n_3n_1} \\ 0 & 0 & 0 & j_1j_3j_2(u_1u_2u_3)^{m_1m_3m_2}{}_{n_1n_3n_2} & 0 & 0 \\ 0 & j_2j_1j_3(u_1u_2u_3)^{m_2m_1m_3}{}_{n_2n_1n_3} & 0 & 0 & 0 & 0 \\ 0 & 0 & j_3j_2j_1(u_1u_2u_3)^{m_3m_2m_1}{}_{n_3n_2n_1} & 0 & 0 & 0 \\ j_1j_2j_3(u_1u_2u_3)^{m_1m_2m_3}{}_{n_1n_2n_3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & j_3j_1j_2(u_1u_2u_3)^{m_3m_1m_2}{}_{n_3n_1n_2} & 0 \end{pmatrix} \tag{3.7}$$

Applying the same restrictions as in the last case, the only transformations allowed are those by diagonal matrices

$$\begin{pmatrix} \exp(i\phi_1)I & & & & & 0 \\ & \exp(i\phi_2)I & & & & \\ & & \exp(i\phi_3)I & & & \\ & & & \exp(i\phi_4)I & & \\ & & & & \exp(i\phi_5)I & \\ 0 & & & & & \exp(i\phi_6)I \end{pmatrix} \quad (3.8)$$

Before leaving this section we note that for each case we have been able to ignore a pair of indices as their range was trivial, although it did depend on which case was considered. Nevertheless, we seize the chance to discard them as in this type of work indices tend to proliferate. In the sequel such indices will be omitted wherever possible.

4. DESCENT TO $G'\otimes S_3$

We have constructed the irreps of $\Gamma = G^3\otimes S_3$ and shown that they may be labelled as $[\lambda\,]j_1j_2j_3$, where $[\lambda\,]$ is an irrep of S_1, S_2, or S_3, and $j_1,j_2\,j_3$ are irreps of G. We now restrict ourselves to $\Gamma' = G'\otimes S_3$ and consider the reduction of each irrep of Γ to the trivial irrep $1_{\Gamma'}$ of Γ'. First we deal with the multiplicity which we label by

$$m^{1_{\Gamma'}}_{[\lambda\,]j_1j_2j_3}.$$

To calculate this multiplicity the character alone is needed, and the specific matrix forms of the last section may be used. Each case is considered in turn.

When none of j_1, j_2, j_3 are equivalent, the trace of $D_{[1]j_1j_2j_3}(\pi,u_1u_2u_3)$ is only nonzero for $\pi = e$ when it is

$$\chi_{[1]j_1j_2j_3}(e,uuu) = 6\chi_1(u)\chi_2(u)\chi_3(u), \tag{4.1}$$

where χ_i is the character of j_i. The multiplicity of $1_{\Gamma'}$ is

$$\begin{aligned} m^{1_{\Gamma'}}_{[\lambda\,]j_1j_2j_3} &= \frac{1}{|\Gamma'|}\int_G \sum_{\pi e S_3}\chi_{[1]j_1j_2j_3}(\pi,uuu)\,du \\ &= \frac{6}{|\Gamma'|}\int_G \chi_1(u)\chi_2(u)\chi_3(u)\,du. \end{aligned} \tag{4.2}$$

But $|\Gamma'| = 6|G\,|$, and so the right-hand side is the multiplicity of $1_{G'}$ in $j_1\otimes j_2\otimes j_3$, i.e.,

$$m^{1_{\Gamma'}}_{[1]j_1j_2j_3} = m^{1_{G'}}_{[1]j_1j_2j_3}. \tag{4.3}$$

When exactly two are equivalent (say $j_1 = j_2 \neq j_3$), the character is nonvanishing for more elements of Γ':

$$\chi_{[\lambda\,]j_1j_1j_3}(e,uuu) = 3\chi_{[\lambda\,]}(e)(\chi_1(u))^2\chi_3(u), \tag{4.4}$$

$$\chi_{[\lambda\,]j_1j_1j_3}((12),uuu) = \chi_{[\lambda\,]}((12))\,\chi_1(u^2)\,\chi_3(u), \tag{4.5}$$

with the same trace for $\pi = (13)$ and (23). This gives for the multiplicity of $1_{\Gamma'}$,

$$\begin{aligned} m^{1_{\Gamma'}}_{[\lambda\,]j_1j_1j_2} = \frac{1}{2|G\,|}\int_G \{&\chi_{[\lambda\,]}(e)(\chi_1(u))^2\chi_3(u) \\ &+ \chi_{[\lambda\,]}((12))\chi_1(u^2)\chi_3(u)\}\,du. \end{aligned} \tag{4.6}$$

Now $[\lambda\,]$ can only be [2] or $[1^2]$, and the right-hand side may be recognized as the multiplicity of $1_{G'}$ in $(j_1\odot[\lambda\,])\otimes j_3$ (or of j_3^* in the symmetric or antisymmetric square of j_1). Thus

$$m^{1_{\Gamma'}}_{[\lambda\,]j_1j_1j_2} = m^{1_{G'}}_{[\lambda\,]j_1j_1j_2}. \tag{4.7}$$

The third case, when all j-values are equivalent, may be evaluated in the same way. Here $[\lambda\,]$ may be [3], $[1^3]$, or [21] and the multiplicity of $1_{\Gamma'}$ is the same as the multiplicity of $1_{G'}$ in the symmetric, antisymmetric, and mixed symmetry cubes of j, i.e.,

$$m^{1_{\Gamma'}}_{[\lambda\,]jjj} = m^{1_{G'}}_{[\lambda\,]jjj}. \tag{4.8}$$

We now *define* the $[\lambda\,]$ r-$3jm$ tensor to be the part of the unitary tensor which reduces $D_{[\lambda\,]j_1j_2j_3}$ to $1_{\Gamma'}$. In addition to the G irrep labels j and component labels m, it also contains an S_n irrep label $[\lambda\,]$ and component label r. It may be defined by the eigenvector equation

$$\begin{aligned} D_{[\lambda\,]j_1j_2j_3}(\pi,uuu)^{r'rm_1m_2m_3}{}_{s'sn_1n_2n_3}([\lambda\,]\,j_1j_2j_3)^{s'sn_1n_2n_3}{}_t \\ = ([\lambda\,]\,j_1j_2j_3)^{r'rm_1m_2m_3}{}_t \end{aligned} \tag{4.9}$$

for all $\pi\in S_3, u\in G$, where t is a "multiplicity label" with values from one to $m^{1_{\Gamma'}}_{[\lambda\,]j_1j_2j_3}$.

The unitary condition is

$$([\lambda\,]\,j_1j_2j_3)^{r'rm_1m_2m_3}{}_{t_1}([\lambda\,]\,j_1j_2j_3)^{t_2}{}_{r'rm_1m_2m_3} = \delta^{t_2}{}_{t_1} \tag{4.10}$$

which is a "weak orthogonality" property. In the next section we shall show that this can be tightened in each case.

5. THE SYMMETRIZED $3jm$ TENSOR

For the last stage in the derivation of the symmetrized $3jm$ tensor, consider the descents $\Gamma'\downarrow G'$ and $\Gamma'\downarrow S_3$. This section divides into three cases according to the previous section.

When all three j-values are equal, the coset representatives are trivial and the label r may be omitted from the $[\lambda\,]r$-$3jm$ tensor. Restriction to S_3 with $u = e$ makes Eq. (4.9) read

$$\lambda\,(\pi)^{r'}{}_{s'}([\lambda\,]j_1j_1j_1)^{s'\pi(m_1m_2m_3)}{}_t = ([\lambda\,]\,j_1j_1j_1)^{r'm_1m_2m_3}{}_t, \tag{5.1}$$

or, in more detail,

$$\lambda\,((12))^{r'}{}_{s'}([\lambda\,]\,j_1j_1j_1)^{s'm_2m_1m_3}{}_t = ([\lambda\,]\,j_1j_1j_1)^{r'm_1m_2m_3}{}_t \tag{5.2}$$

and

$$\lambda\,((123))^{r'}{}_{s'}([\lambda\,]\,j_1j_1j_1)^{s'm_2m_3m_1}{}_t = ([\lambda\,]\,j_1j_1j_1)^{r'm_1m_2m_3}{}_t. \tag{5.3}$$

Thus if $[\lambda\,] = [3]$, the value of the tensor is unchanged under permutation of m-values; if $[\lambda\,] = [1^3]$ it changes sign under tranpositions but not under cyclic permutations; if $[\lambda\,] = [21]$ there are *two* tensors for each t corresponding to r' equals 1 or 2. These two tensors transform as a basis for [21] and hence their transformation properties are much more complex. Many examples of such "nonsimple phase" cases are now known to occur in both the finite[23–25] and compact Lie[18] groups, although few numerical cases have appeared. Schindler and Mirman[26] have given some for the symmetric groups but apparently without complete recognition.

On descent to G', Eq. (4.9) becomes

$$j_1(u)^{m_1}{}_{n_1} j_1(u)^{m_2}{}_{n_2}\, j_1(u)^{m_3}{}_{n_3}([\lambda\,]\, j_1j_1j_1)^{r'n_1n_2n_3}{}_t$$
$$=([\lambda\,]\, j_1j_1j_1)^{r'm_1m_2m_3}{}_t. \tag{5.4}$$

However, this is just one of the ways of defining the ordinary $3jm$ tensor: its columns are a complete set of independent solutions satisfying a certain orthogonality condition. The strongest orthogonality condition is that derived from the coupling coefficient,

$$(j_1j_1j_1)^{m_1m_2m_3}{}_{t'_1}(j_1j_1j_1)^{t'_2}{}_{m_1m_2m'_3}=|j_1|^{-1}\delta^{t'_2}{}_{t'_1}\delta^{m_3}{}_{m'_3}, \tag{5.5}$$

where $|j_1|$ is the dimension of j_1. (This equation holds for a sum over any pair of m-values, not just m_1 and m_2.) The columns of the $3jm$ tensor form a basis for a vector space called the multiplicity space, and any column of the $[\lambda\,]r$-$3jm$ tensor must therefore be expressible in terms of the $3jm$ tensor

$$([\lambda\,]\, j_1j_1j_1)^{r'm_1m_2m_3}{}_t=A_{[\lambda\,]}{}^{r't'}{}_t(j_1j_1j_1)^{m_1m_2m_3}{}_{t'}. \tag{5.6}$$

From the unitary condition (4.10) we already have one orthogonality property. The last two equations allow this to be tightened. Form the inner product (Hermitian adjoint)

$$([\lambda_1]\, j_1j_1j_1)^{r'_1m_1m_2m_3}{}_{t_1}([\lambda_2]\, j_1j_1j_1)^{t_2}{}_{r'_2m_1m_2m'_3}$$
$$=A_{[\lambda_1]}{}^{r'_1t'_1}{}_{t_1}A_{[\lambda_2]}{}^{t_2}{}_{r'_2t'_2}|j_1|^{-1}\delta^{m_3}{}_{m'_3}\delta^{t'_2}{}_{t'_1}$$
$$=B^{r'_1t_2}{}_{r'_2t_1}|j_1|^{-1}\delta^{m_3}{}_{m'_3}, \tag{5.7}$$

where $B^{r'_1t_2}{}_{r'_2t_1}=A_{[\lambda_1]}{}^{r'_1t'_1}{}_{t_1}A_{[\lambda_2]}{}^{t_2}{}_{r'_2t'_1}$. This may be improved further by equating m_3 and m'_3, summing, and then utilizing Eq. (5.1):

$$B^{r'_1t_2}{}_{r'_2t_1}=([\lambda_1]\, j_1j_1j_1)^{r'_1m_1m_2m_3}{}_{t_1}([\lambda_2]\, j_1j_1j_1)^{t_1}{}_{r'_2m_1m_2m_3}$$
$$=\lambda_1(\pi)^{r'_1}{}_{s'_1}\lambda_2\,(\pi^{-1})^{s'_2}{}_{r'_2}([\lambda_1]\, j_1j_1j_1)^{s'_1\pi(m_1m_2m_3)}{}_{t_1}$$
$$\times([\lambda_2]\, j_1j_1j_1)^{t_2}{}_{s'_2\pi(m_1m_2m_3)}$$
$$=\lambda_1(\pi)^{r'_1}{}_{s'_1}\lambda_2(\pi^{-1})^{s'_2}{}_{r'_2}\,B^{s'_1t_2}{}_{s'_2t_1}\,, \tag{5.8}$$

i.e.,

$$\lambda_1(\pi)^{r'_1}{}_{s'_1}B^{s'_1t_2}{}_{s'_2t_1}=B^{r'_1t_2}{}_{r'_2t_1}\lambda_2(\pi)^{r'_2}{}_{s'_2}. \tag{5.9}$$

For each pair t_1,t_2, the matrix B intertwines λ_1 and λ_2 and hence by Schur's lemma is zero for $\lambda_1\neq\lambda_2$ and diagonal for $\lambda_1=\lambda_2$. Hence

$$B^{r'_1t_2}{}_{r'_2t_1}=|\lambda_1|^{-1}\delta^{r'_1}{}_{r'_2}C^{t_2}{}_{t_1}\delta(\lambda_1,\lambda_2). \tag{5.10}$$

Substituting this into Eq. (5.7), equating r'_1 and r'_2, m_3 and m'_3, and summing invokes the orthogonality of the $[\lambda\,]r$-$3jm$ tensor so that C is also a delta tensor. Hence

$$([\lambda_1]\, j_1j_1j_1)^{r'_1m_1m_2m_3}{}_{t_1}([\lambda_2]\, j_1j_1j_1)^{t_2}{}_{r'_2m_1m_2m'_3}$$
$$=|\lambda_1|^{-1}|j_1|^{-1}\delta(\lambda_1,\lambda_2)\delta^{r'_1}{}_{r'_2}\delta^{t_2}{}_{t_1}\delta^{m_3}{}_{m'_3}. \tag{5.11}$$

Counting dimensions, this shows that the columns of the $[\lambda\,]r$-$3jm$ tensors for $[\lambda\,]=[3]$, $[21]$, and $[1^3]$ form an orthogonal basis for the $3jm$ multiplicity space. Further, if the $[\lambda\,]r$-$3jm$ tensor is divided by $|\lambda\,|^{1/2}$, this basis satisfies the same orthogonality property as any $3jm$ tensor and hence *defines* a particular $3jm$ tensor—the "symmetrized" $3jm$ tensor. That is, if the columns of the $[\lambda\,]r$-$3jm$ tensor are divided by $|\lambda\,|^{1/2}$, they form columns (or pairs of columns when $[\lambda\,]=[21]$) of the $3jm$ tensor, and these columns transform as basis vectors for $[\lambda\,]$ when the m-values are permuted. Any other $3jm$ tensor derived by any method is related to this one by a unitary transformation in the multiplicity space. This completes the discussion for all irreps equal.

When exactly two are equal, say $j_1=j_2\neq j_3$, the primed labels may be omitted as $[\lambda\,]=[2]$ or $[1^2]$ is one-dimensional. On descent to G', Eq. (4.9) gives

$$\begin{pmatrix} j_1j_1j_3(uuu)^{m_1m_2m_3}{}_{n_1n_2n_3} & 0 & 0\\ 0 & j_3j_1j_1(uuu)^{m_3m_1m_2}{}_{n_3n_1n_2} & 0\\ 0 & 0 & j_1j_3j_1(uuu)^{m_2m_3m_1}{}_{n_2n_3n_1}\end{pmatrix}^r{}_s\times([\lambda\,]\, j_1j_1j_3)^{sm_1m_2m_3}{}_t=([\lambda\,]\, j_1j_1j_3)^{rn_1n_2n_3}{}_t. \tag{5.12}$$

This breaks into three equations, and each equation defines subspaces of $3jm$ multiplicity spaces. For r equals 1, 2, and 3, these are subspaces of the $3jm$ tensors $(j_1j_1j_3)$, $(j_3j_1j_1)$, and $(j_1j_3j_1)$, respectively. As for the last case, orthogonality of the $3jm$ tensors produces

$$([\lambda_1]\, j_1j_1j_3)^{1m_1m_2m_3}{}_{t_1}([\lambda_2]\, j_1j_1j_3)^{t_2}{}_{1m_1m_2m'_3} \tag{5.13}$$
$$=B^{1t_2}{}_{1t_1}|j_3|^{-1}\delta^{m_3}{}_{m'_3}$$

with similar conditions for $r=2,3$. Under the e and (12) permutations applied to this, Schur's lemma in S_2 gives $\delta(\lambda_1,\lambda_2)$ on the right, whereas applying (13) and (23) permutations shows

$$B^{1t_2}{}_{1t_1}=B^{2t_2}{}_{2t_2}=B^{3t_2}{}_{3t_1}.$$

Orthogonality of the $[\lambda\,]r$-$3jm$ tensor in addition gives

$$B^{rt_2}{}_{rt_1}=\delta^{t_2}{}_{t_1}$$

so that $1/\sqrt{3}\,([\lambda\,]\, j_1j_1j_3)$ for $[\lambda\,]=[2],[1^2]$ defines three particular $3jm$ tensors $(j_1j_1j_3)$,$(j_3j_1j_1)$, and $(j_1j_3j_1)$.

The permutation properties of these symmetrized $3jm$ tensors follow by descent $\Gamma'\downarrow S_3$. A (12) permutation is completely defined for the $(j_1j_1j_3)$ tensor:

$$\lambda\,(12)([\lambda\,]\, j_1j_1j_3)^{1m_2m_1m_3}{}_t=([\lambda\,]\, j_1j_1j_3)^{1m_1m_2m_3}{}_t$$

so that this tensor changes sign if $[\lambda\,]=[1^2]$ but is left invariant if $[\lambda\,]=[2]$. Identical properties hold for $(j_3j_1j_1)$ under (23) transpositions, and for $(j_1j_3j_1)$ under (13) transpositions.

The effect of permutations can also be calculated for the off-diagonal elements, but it must be noted that the allowed equivalence transformations in Γ produce an arbitrariness in this. For the off-diagonal (12) elements,

$$\lambda((12))([\lambda]\,j_1j_1j_3)^{3m_1m_2m_3}{}_t$$
$$= \exp i(\phi_3 - \phi_2)([\lambda]\,j_1j_1j_3)^{2m_1m_2m_3}{}_t$$

and for (123),

$$([\lambda]\,j_1j_1j_3)^{3m_1m_2m_3}{}_t = \exp i(\phi_3 - \phi_1)([\lambda]\,j_1j_1j_3)^{1m_1m_2m_3}{}_t,$$
$$([\lambda]\,j_1j_1j_3)^{1m_1m_2m_3}{}_t = \exp i(\phi_1 - \phi_2)([\lambda]\,j_1j_1j_3)^{2m_1m_2m_3}{}_t,$$

etc. The choices of ϕ_1, ϕ_2, and ϕ_3 are entirely upto the user, and are determined by choosing an irrep in Γ from its equivalence class.

The final case of all irreps inequivalent follows this last case very closely. When divided by $\sqrt{6}$, the $[\lambda]r$-$3jm$ defines six symmetrized $3jm$ tensors corresponding to the six orderings of $(j_1j_2j_3)$. Permutations in S_3 map these tensors onto each other with at most a change of phase. These phase factors are completely arbitrary (to within consistency imposed by multiplications) and are determined by the choice of irrep in Γ. For the irreps given explicitly in Sec. 3 all phase factors are 1 so that these symmetrized $3jm$ tensors are invariant under any permutation of j and m-values.

Particular $3jm$ tensors have thus been shown to arise naturally on reducing irreps of $(G \times G \times G)\circledS S_3$ to the trivial irrep of the subgroup $(\text{diag } G \times G \times G)\circledS S_3$. These $3jm$ tensors are "symmetrized" in that their transformation properties under permutations are as simple as possible. Any other $3jm$ tensors are related to these by unitary transformations in the multiplicity space. While there are arbitrary phase factors in some of these permutation properties, they are explained as arising from equivalence transformations in $G \wr S_3$.

6. THE METHOD OF PLETHYSMS

It was mentioned in the introduction that some authors have used Littlewood's algebra of plethysms to obtain results about $3j$ symmetries, mainly as to the existence or nonexistence of nonsimple phase irreps. As we have *not* used plethysms but something quite closely related, it is worth detailing this link. There are in fact two plethysm algebras, the "inner" and the "outer." However, as the outer plethysm algebra is really just the inner plethysm algebra for the general linear group transferred to the symmetric group via the duality between the two groups, we need only talk about inner plethysms to cover all cases.

The inner plethysm construction, as detailed for example by Kerber,[22] is quite heavily dependent on the irreps and carrier spaces of a group. Given an irrep j of G with carrier space V, the action of S_n on $V \times V \times \cdots \times V$ (n times) is defined by permutations of the basis vectors. This gives a representation of S_n over this direct product space which may be combined with the irrep j to give an irrep of $G \sim S_n$ of dimension $|j|^n$. For the case n equals 3, this is in fact the irrep of Γ we have called $D_{[3]jjj}$. The other irreps with $[\lambda] = [21]$ and $[1^3]$ may be obtained quite readily, but this is a "second-stage" calculation and they do not appear quite so naturally. For our other irreps when not all j-values are equivalent, the plethysm method works best in a subgroup of S_n [for example $G \wr (S_2 \times S_1)$ for exactly two j-values equal] and this is not sufficient to give all $3j$ symmetries for these cases. The irreps in this subgroup may, however, be induced to $G \sim S_n$ as we have done.

The subgroup Γ' is isomorphic to $G \wr S_3$ and the plethysm method reduces representations of Γ' to $j \times [\lambda]$ of $G \times S_3$. In the context of $3j$ symmetries the reduction is to $1_G \times [\lambda]$. In our notation, this is reduction to $D_{[\lambda]1_G1_G1_G}$ with $[\lambda] = [3]$, $[21]$, and $[1^3]$. A complete list of the multiplicities of these irreps in all irreps of Γ' is given in Table I. From there it may be seen that the multiplicity of $D_{[\lambda]1_G1_G1_G}$ in $D_{[3]jjj}$ is the same as the multiplicity of $1_{\Gamma'}$ in $D_{[\lambda]jjj}$ but that many other entries on this table (in particular when not all j-values are equivalent) are rather strange. We dispose of these cases first.

When only two j-values are equal, the natural group for inner plethysms is $G \wr (S_2 \times S_1)$ and its irreps are reduced in

TABLE I. Multiplicity of irreps $[\mu] \times 1_G$ in representations $D_{[\lambda]j_1j_2j_3}$ of $(\text{diag } G \times G \times G)\circledS S_3$.

	$[3] \times 1_G$	$[1^3] \times 1_G$	$[21] \times 1_G$
$[3]j_1j_1j_1$	$\frac{1}{6G}\int_G \{\chi_1(u)\}^3$ $+ 3\chi_1(u^2)\chi_1(u) + 2\chi_1(u^3)\,du$	$\frac{1}{6G}\int_G \{\chi_1(u)\}^3$ $- 3\chi_1(u^2)\chi_1(u) + 2\chi_1(u^3)\,du$	$\frac{1}{3G}\int_G \{\chi_1(u)\}^3 - \chi_1(u^3)\,du$
$[1^3]j_1j_1j_1$	$\frac{1}{6G}\int_G \{\chi_1(u)\}^3$ $- 3\chi_1(u^2)\chi_1(u) + 2\chi_1(u^3)\,du$	$\frac{1}{6G}\int_G \{\chi_1(u)\}^3$ $+ 3\chi_1(u^2)\chi_1(u) + 2\chi_1(u^3)\,du$	$\frac{1}{3G}\int_G \{\chi_1(u)\}^3 - \chi_1(u^3)\,du$
$[21]j_1j_1j_1$	$\frac{1}{3G}\int_G \{\chi_1(u)\}^3 - \chi_1(u^3)\,du$	$\frac{1}{3G}\int_G \{\chi_1(u)\}^3 - \chi_1(u^3)\,du$	$\frac{1}{3G}\int_G 2\{\chi_1(u)\}^3 + \chi_1(u^3)\,du$
$[2]j_1j_1j_2$	$\frac{1}{2G}\int_G \{\chi_1(u)\}^2\chi_3(u)$ $+ \chi_1(u^2)\chi_3(u)\,du$	$\frac{1}{2G}\int_G \{\chi_1(u)\}^2\chi_3(u)$ $- \chi_1(u^2)\chi_3(u)\,du$	$\frac{1}{G}\int_G \{\chi_1(u)\}^2\chi_3(u)\,du$
$[1^2]j_1j_1j_2$	$\frac{1}{2G}\int_G \{\chi_1(u)\}^2\chi_3(u)$ $- \chi_1(u^2)\chi_3(u)\,du$	$\frac{1}{2G}\int_G \{\chi_1(u)\}^2\chi_3(u)$ $+ \chi_1(u^2)\chi_3(u)\,du$	$\frac{1}{G}\int_G \{\chi_1(u)\}^2\chi_3(u)\,du$
$[1]j_1j_2j_3$	$\frac{1}{G}\int_G \chi_1(u)\chi_2(u)\chi_3(u)\,du$	$\frac{1}{G}\int_G \chi_1(u)\chi_2(u)\chi_3(u)\,du$	$\frac{2}{G}\int_G \chi_1(u)\chi_2(u)\chi_3(u)\,du$

$G'\otimes(S_2\times S_1)\cong G\times S_2$. The reduction is then to $1_G\times[2]$ and $1_G\times[1^2]$ which only deals with the (12) permutations of $(j_1 j_1 j_3)$, the (13) permutations of $(j_1 j_3 j_1)$, or the (23) permutations of $(j_3 j_1 j_1)$. A (123) permutation cannot be used to relate the three tensors for it is simply not in the group. On the other hand, using Γ' is not correct either for it is not clear as to whether reductions should be to $1_G\times[3]$, $1_G\times[21]$, or $1_G\times[1^3]$. Inner plethysms are just not appropriate for discussing the $3j$ symmetries of this case. Similar considerations hold for none of the irreps equivalent.

When all three j-values are equal though, the multiplicities certainly tally and inner plethysms can be used for characterlike calculations. A closer investigation reveals that the plethysm reduction is something like a coupling coefficient to the $[\lambda]r$-$3jm$ tensor. We define this plethysm reduction by a $3jm$-$[\mu]r$ tensor:

$$jjj(uuu)^{\pi(m_1m_2m_3)}{}_{n_1n_2n_3}$$

$$= \sum_{[\mu]\in\mathrm{Irr}(S_3)} (jjj)_{[\mu]}{}^{m_1m_2m_3}{}_{rt'}\,\mu(\pi)^r{}_s(jjj)_{[\mu]}{}^{st'}{}_{n_1n_2n_3} \qquad (6.1)$$

$$\oplus \text{ other irreps of } G\times S_3.$$

If this equation is multiplied by $\lambda(\pi)^{r'}{}_{s'}$ and integrated over Γ' divided by its volume, the left-hand side becomes

$$([\lambda]\,jjj)^{r'm_1m_2m_3}{}_t([\lambda]\,jjj)^t{}_{s'n_1n_2n_3} \qquad (6.2)$$

as this is the only component transforming as $1_{\Gamma'}$. On the right the only nonvanishing component is that which transforms as $1_G\times[3]$ of $G\times S_3$ which can only occur for the [3] component of $[\lambda]\otimes[\mu]$. Thus $[\mu]$ must equal $[\lambda]$ and the direct product must be reduced by a $2jm$ tensor in S_3:

$$([\lambda]\,jjj)^{r'm_1m_2m_3}{}_t([\lambda]\,jjj)^t{}_{s'n_1n_2n_3}$$

$$=(jjj)_{[\lambda]}{}^{m_1m_2m_3}{}_{rt'}([\lambda][\lambda])^{rr'}([\lambda][\lambda])_{ss'}$$

$$\times(jjj)_{[\lambda]}{}^{st'}{}_{n_1n_2n_3}. \qquad (6.3)$$

By the orthogonality properties of all the tensors this may be recast into

$$([\lambda]\,jjj)^{r'm_1m_2m_3}{}_t = U^{t'}{}_t(jjj)_{[\lambda]}{}^{m_1m_2m_3}{}_{rt'}([\lambda][\lambda])^{rr'}, \qquad (6.4)$$

where U is a unitary tensor relating bases in the multiplicity spaces. By transforming one of the $[\lambda]\ r$-$3jm$ or $3jm$-$[\mu]\ r$ tensors this may be taken as diagonal so that one tensor may be found directly from the other.

The $2jm$ tensor in Eq. (6.4) is, in general, *not* trivial. However, if the irreps of S_3 are chosen to be *orthogonal* it reduces to a tensor which merely changes columns into rows. (This is because the $1jm$ tensor is diagonal.) For this special case, which is nevertheless the most common one, we may write

$$([\lambda]\,jjj)^{r'm_1m_2m_3}{}_t = |\lambda|^{-1/2}\delta^{rr'}(jjj)_{[\lambda]}{}^{m_1m_2m_3}{}_{rt}$$

to give one tensor from the other.

7. A POSSIBLE GENERALIZATION

It has been shown that the permutation properties of the $3jm$ tensor may be found by reducing irreps of $\Gamma = G \wr S_3$ to the trivial irrep $1_{\Gamma'}$ in the subgroup Γ', or when all j-values are equal by reducing certain irreps of Γ to certain others in Γ'. In Table I this means we have used the first row and the first column only. A quite natural question is to ask what tensors correspond to the other entries. A formal answer is to define a tensor which reduces $D_{[\lambda]j_1j_2j_3}$ to $D_{[\mu]1_G1_G1_G}$, which for consistent terminology must go under the title of a $[\lambda]r$-$3jm$-$[\mu]r$ tensor. Such a tensor must always exist for any group even if trivially by setting $[\lambda]=[\mu], j_1=j_2=j_3=1_G$. It is unlikely that this tensor will prove of much importance as it can be expressed in terms of $3jm$ tensors for G and S_3, but if any sufficiently interesting results are discovered they will be reported.

APPENDIX

In this appendix we state the terminology which appears to be most appropriate in discussing the Wigner–Racah algebra. For more details the reader is referred to Derome and Sharp[1,2] or better the article by Butler.[3]

The *1jm tensor or Wigner tensor*. For any calculations with irreps of a group, it is assumed that the matrices of each irrep are fixed. To each irrep j there is a conjugate irrep j^* (which may of course equal j). If the complex conjugate of the matrices of j is taken then this matrix irrep will be *equivalent* to the matrices of j^*. The $1jm$ tensor is the matrix of equivalence. The "m" in the notation signifies that it is basis dependent.

The *1j phase*. By Schur's lemma, the product of the $1jm$ tensors for $j\rightarrow j^*$ and $j^*\rightarrow j$ is a scalar matrix λI. λ is the $1j$ phase. It is independent of basis so no "m" labels are included.

The *2jm tensor* is the tensor which reduces the inner product $j\otimes j^*$ to 1_G, the trivial irrep.

The *2j phase* is the phase factor arising on permuting the j- and m-values in the $2jm$ tensor.

The *coupling coefficient* is the tensor which reduces the inner product of two irreps $j_1\otimes j_2$.

The *3jm tensor* is the tensor which reduces the inner product of three irreps $j_1\otimes j_2\otimes j_3$ to 1_G.

The *3j tensor* is the permutation tensor in the multiplicity labels relating one $3jm$ tensor to another (permuted) one. It is independent of basis labels m but does depend on the three irrep labels.

The *Clebsch–Gordan series* gives the multiplicity of each irrep in the inner product $j_1\otimes j_2$. This is based on the original series for SO(3), and is *not* the coupling coefficient.

[1] J-R. Derome and W. T. Sharp, J. Math. Phys. **6**, 1584 (1965).

[2] J-R. Derome, J. Math. Phys. **7**, 612 (1966).

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

[4] P. H. Butler and B. G. Wybourne, Int. J. Quantum Chem. **10**, 581 (1976); **10**, 599 (1976); **10**, 615 (1976).

[5] P. H. Butler, R. W. Haase, and B. G. Wybourne, Aust. J. Phys. **31**, 131 (1978).

[6] P. H. Butler and M. F. Reid, J. Phys. A **12**, 1655 (1979).

[7] P. H. Butler and A. M. Ford, J. Phys. A **12**, 1357 (1979).

[8] P. H. Butler and M. F. Reid, J. Phys. A **13**, 2889 (1980).

[9] E. König and S. Kremer, Z. Naturforsch. A **29**, 1179 (1974).

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

[11] J. D. Newmarch and R. M. Golding, J. Math. Phys. **22**, 2113 (1981).

[12] J. D. Newmarch and R. M. Golding, "The Racah Algebra for Groups

with Time Reversal Symmetry. III, J. Math. Phys. **24**, 441 (1983).

[13]J. D. Newmarch, "On the Symmetries of the 6*j* symbol," J. Math. Phys. **24**, 451 (1983).

[14]D. R. Pooler, J. Phys. A **13**, 1197 (1980).

[15]L. S. R. K. Prasad and K. Bharathi, J. Phys. A **13**, 781 (1980).

[16]G. E. Stedman and P. H. Butler, J. Phys. A **13**, 3125 (1980).

[17]R. C. King, J. Math. Phys. **15**, 258 (1974).

[18]P. H. Butler and R. C. King, Can. J. Math. **26**, 328 (1974).

[19]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).

[20]J. D. Newmarch, "Some Character Theory for Groups of Linear and Anti-linear Operators," J. Math. Phys. **24**, 742 (1983).

[21]L. Jansen and M. Boon, *Theory of Finite Groups. Applications in Physics* (North-Holland, Amsterdam, 1967).

[22]A. Kerber, *Representations of Permutation Groups II, Lecture Notes in Mathematics*, Vol. 495 (Springer, Berlin, 1975).

[23]A. J. van Zanten and E. de Vries, J. Algebra **25**, 475 (1973).

[24]A. J. van Zanten and E. de Vries, Can. J. Math. **26**, 1090 (1974).

[25]A. J. van Zanten and E. de Vries, Can. J. Math. **27**, 528 (1975).

[26]S. Schindler and R. Mirman, J. Math. Phys. **18**, 1678 (1977); **18**, 1697 (1977).

A theorem on orbit structures (strata) of compact linear Lie groups

G. Sartori

Istituto di fisica dell "Universita" di Padova and I.N.F.N., Sezione di Padova, 1 35100 Padova, Italy

(Received 24 November 1981; accepted for publication 15 October 1982)

We present a comprehensive constructive proof of a theorem characterizing the tangent space to a stratum (orbit structure) of the Euclidean space $\mathbf{R}^n$, seat of an orthogonal representation of a compact group G. The characterization is made in terms of gradients of a complete set (integrity basis) of G-invariant polynomials. In a recent paper [M. Abud and G. Sartori, Phys. Lett. B **104**, 147 (1981)], the theorem, which may be considered a generalization of a theorem by Michel [C. R. Acad. Sci. Ser. A **272**, 433 (1971)], has been shown to be effective in the determination of the equations of the strata and in the determination of natural extrema of G-invariant functions.

PACS numbers: 02.20.Qs

I. INTRODUCTION

After the pioneer papers by Michel and Radicati[1] it became evident that some information on the geometry of group representations may help[2] in determining the location of extrema (stationary points) of scalar functions, a basic problem in many physical problems.

In a recent letter by M. Abud and the present author[3] a new theorem on this subject was stated (without proof), and exploited essentially in order to characterize possible directions of spontaneous symmetry breaking and phase transitions. The purpose of the present paper is to give a comprehensive proof of that theorem, whose content will be recalled in this Introduction, after a few definitions.

Let G be a compact r-dimensional Lie group of orthogonal transformations in the Euclidean space $\mathbf{R}^n$. We shall denote by $\mathbf{R}^n/G$ the orbit space associated to the G-space $(\mathbf{R}^n, G)$. As is well known the collection of isotropy subgroups of G (little groups), at points lying on the same orbit $\Omega(\phi)$ as ϕ, constitutes a class of subgroups conjugated in G to the stability subgroup G_ϕ at ϕ. This class is called a G-orbit type and will be denoted equivalently by $\{G_\phi\}$ and $[\Omega(\phi)]$.

Following Michel,[1] we shall call "stratum" of $\mathbf{R}^n$ through ϕ [of $\mathbf{R}^n/G$ through $\Omega(\phi)$] the union [the set] of all G-orbits of $\mathbf{R}^n$ [of $\mathbf{R}^n/G$] that are of type $[\Omega(\phi)]$.

The results we shall prove in this paper have been shown[3] to be quite effective in the characterization of strata.

As is well known,[4] the tangent space to a stratum of $\mathbf{R}^n$ at one of its points ϕ is the direct sum of the tangent space to the G-orbit $\Omega(\phi)$ at ϕ and the vector space $N_\phi^{(0)}$ formed by all the G_ϕ-invariant vectors that are orthogonal to $\Omega(\phi)$ at ϕ. It is also well known that the gradient at ϕ of any differentiable G-invariant function $F(\phi)$ belongs to $N_\phi^{(0)}$.

Our final result in Sec. III will prove that, conversely,[5] every vector of $N_\phi^{(0)}$ can be expressed as a linear combination of gradients at ϕ of a fixed finite set of G-invariant polynomial functions of ϕ (integrity basis for the ring of of G-invariant polynomials of ϕ). In particular, when $N_\phi^{(0)}$ is one dimensional, $\Omega(\phi)$ turns out to be an isolated point in its orbit-space stratum and, according to our theorem, the gradients at ϕ of all differentiable G-invariant functions of ϕ must be (null or) parallel vectors and vice versa. Thus our result may also be considered a generalization of Michel's theorem[1] for linear G-actions.

The plan of the paper is the following. In Sec. II we shall collect some relevant definitions and, in an elementary way, prove some results concerning the geometry of group representations which we shall need in Sec. III. These results are already known in mathematical literature. In Sec. III, after recalling a few fundamental results in the theory of G-invariant functions, we shall restate and prove the theorem which is the main result of this paper.

In three appendices we have confined elementary proofs of some known results used in the paper. In Appendix A a convenient local parametrization of a G-orbit is defined. It will be exploited in Appendix B in order to derive some properties of the Hessian of the function $f(g) = \delta(g{\cdot}\xi,\phi)$.

The results of Appendix B are needed in Appendix C, which is devoted to a proof of Proposition 1, stated in Sec. II.

II. SOME GEOMETRY OF GROUP REPRESENTATIONS

In this section we shall complete the necessary collection of definitions and results concerning the geometry of compact group representations.

An open square □ will denote the end of a proof.

$\phi\in\mathbf{R}^n$ will be identified, in a standard way, as an n-dimensional vector and a point of an n-dimensional manifold.

We shall use the module notations $g{\cdot}\phi$ and $t{\cdot}\phi$ for the linear action on the vector $\phi\in\mathbf{R}^n$ of $g\in G$ and $t\in\mathscr{G}$, the Lie algebra of G.

The scalar product in $\mathbf{R}^n$ will be denoted by $\langle\,,\,\rangle$, and the squared Euclidean distance by $\delta(\phi,\xi) = \langle\xi-\phi,\xi-\phi\rangle$. Both are G-invariant quantities, since

$$\langle g{\cdot}\phi, g{\cdot}\xi\rangle = \langle\phi,\xi\rangle \quad \forall\phi,\xi\in\mathbf{R}^n, g\in G.$$

The compact C^∞-manifold structure of G induces an analogous structure on each G-orbit Ω. Thus, at every $\phi\in\Omega$ a tangent space $T_\phi(\Omega)$ and a normal space to Ω can be defined. As a vector space $T_\phi(\Omega)$ is isomorphic[4] to the space T_ϕ of tangent displacements to Ω at ϕ, we have:

Definition 1: $T_\phi = \{t{\cdot}\phi;\ \text{all}\ t\in\mathscr{G}\}$.

The orthogonal complement N_ϕ to T_ϕ in $\mathbf{R}^n$ can be identified with the normal space to Ω at ϕ.

Definition 2: The subspace $N_\phi^{(0)}$ formed by all the G_ϕ-invariant vectors of $N_\phi = \{\xi\in\mathbf{R}^n;\ \langle\xi,t{\cdot}\phi\rangle = 0, \text{all}\ t\in\mathscr{G}\}$ will be called the invariant normal space at ϕ.

Clearly $\phi\in N_\phi^{(0)}$, since every $t\in\mathscr{G}$ is skew symmetric with respect to $\langle\,,\,\rangle$.

The squared distance of a point ϕ from an orbit Ω will be defined as the following G-invariant function of ϕ:

Definition 3: $\delta(\phi,\Omega) = \min_{\xi\in\Omega}\delta(\phi,\xi)$.

Analogously the squared distance $\delta(\Omega,\Omega')$ between two orbits Ω and Ω' will be defined by

Definition 4:
$\delta(\Omega,\Omega') = \min_{\phi\in\Omega}\delta(\phi,\Omega') = \min_{\phi'\in\Omega'}\delta(\Omega,\phi')$.

The minima involved in Defs. 3 and 4 certainly exist since G-orbits are compact sets.

Definition 5: If the point on an orbit Ω at minimal distance from a given point ϕ is unique, it will be called[6] the retraction of ϕ on Ω and denoted by $\rho_\Omega(\phi)$.

The function $\rho_\Omega(\phi)$ will play an essential role in the proof of our main result. In the following Lemmas 1–3 and Proposition 1 we shall state some of the its fundamental properties.

Lemma 1: Points on an orbit Ω at minimal and maximal distance from a given point ϕ lie on N_ϕ.

Proof: Let ξ be a point on Ω at minimal or maximal distance from ϕ. The real function on G, $f(g)\equiv\langle g\cdot\xi,\phi\rangle$, consequently, has an extremum (stationary point) at $g = 1$, since G is open. Therefore,

$$0 = \left.\frac{\partial f(g(\omega))}{\partial\omega}\right|_{g=1} = \langle t\cdot\xi,\phi\rangle = -\langle\xi,t\cdot\phi\rangle \quad \forall t\in\mathscr{G},$$

where ω is any parameter of G.□

In Appendix C we shall prove:

Proposition 1[6]: The retraction on an orbit Ω is a C^∞-function at least in an open tubular ϵ-neighborhood of Ω, $\mathscr{T}_\epsilon(\Omega) = \{\phi\in\mathbf{R}^n; \delta(\phi,\Omega) < \epsilon\}$.

Definition 6: The map $\phi\to f(\phi)$ from a G-invariant subset $\mathscr{I}\subset\mathbf{R}^n$ into $\mathbf{R}^n$ is said to be equivariant if

$$f(g\cdot\phi) = g\cdot f(\phi) \quad \forall\phi\in\mathscr{I} \quad \text{and} \quad g\in G. \tag{1}$$

As a trivial consequence of the G-invariance of the distance of a point from an orbit (Def. 3) one finds:

Lemma 2[6]: The map $\phi\to\rho_\Omega(\phi)$ is an equivalent function of ϕ.

The existence of an equivariant function mapping one orbit onto another relates the associated orbit types as we shall show in the following Lemma 4.

In this aim let us introduce a partial ordering in the collection C of all orbit types:

Definition 7: $[\Omega]\leqslant[\Omega']$ if $G_\phi\subseteq G_{\phi'}$ for at least one couple (ϕ,ϕ'), $\phi\in\Omega$, $\phi'\in\Omega'$.

Lemma 3[6]: Let f be an equivariant map from the G-orbit onto the G-orbit Ω'; then $[\Omega']\geqslant[\Omega]$ and the equality holds if and only if f is one-to-one.

Proof: From the equivariance of f, Eq. (1), one immediately obtains

$$G_{f(\phi)}\supseteq G_\phi. \tag{2}$$

Then $[\Omega']\geqslant[\Omega]$ follows from Def. 7. If moreover f is one-to-one, $f^{-1}:\Omega'\to\Omega$ is also equivariant. Thus, using the first part of the lemma we have just proved, we obtain $[\Omega']\leqslant[\Omega]$, which added to $[\Omega]\geqslant[\Omega']$ implies $[\Omega] = [\Omega']$. Conversely let us assume $[\Omega] = [\Omega']$ and prove that if g satisfies

$$f(\phi) = f(g\cdot\phi) \tag{3}$$

then, necessarily $g\cdot\phi = \phi$.

In fact, if we assume Eq. (3), from the equivariance of f we get $g\cdot f(\phi) = f(\phi)$ which implies $g\in G_{f(\phi)}$. But Eq. (2) and the assumption $[\Omega] = [\Omega']$ assure that $G_{f(\phi)} = G_\phi$. Therefore, Eq.(3) implies $g\cdot\phi = \phi$.□

III. A CHARACTERIZATION OF THE INVARIANT NORMAL SPACE

All the functions we shall deal with in this section will be C^∞-functions.

Definition 8: the map F: $\mathbf{R}^n\to\mathbf{R}^1$ is said to be G-invariant if

$$F(g\cdot\phi) = F(\phi), \quad \text{all} \quad \phi\in\mathbf{R}^n \quad \text{and} \quad g\in G. \tag{4}$$

Definition 9: A finite set $\theta = (\theta_1,\ldots\theta_q)$ of homogeneous polynomials $\theta_i(\phi)$ will be called an integrity basis (for the ring of G-invariant polynomials of ϕ), if every G-invariant polynomial $F(\phi)$ can be written as a polynomial $\widehat{F}$ in the θ's:

$$F(\phi) = \widehat{F}(\theta(\phi)), \quad \text{for all } \phi\in\mathbf{R}^n. \tag{5}$$

As shown by Hilbert,[7] all finite dimensional representations of compact Lie groups admit integrity bases.

Schwarz[8] has extended Hilbert's results, showing that for each integrity basis $\theta = (\theta_1,\ldots,\theta_q)$ and each G-invariant C^∞-function $F(\phi)$, $\phi\in\mathbf{R}^n$, there exists a C^∞-function $\widehat{F}(\theta)$, defined in a convenient subset of $\mathbf{R}^q$, so that Eq. (5) holds identically in ϕ.

Lemma 4[4]: The gradient of ϕ at every G-invariant differentiable function lies on the invariant normal space at ϕ.

Proof: Let $\partial F(\phi)$ denote the gradient of $F(\phi)$ at ϕ. By differentiating Eqs. (4), respectively, with respect to ϕ and the parameters of G at $g = 1$, one gets

$$g^{-1}\cdot\partial F(g\cdot\phi) = \partial F(\phi), \quad \text{all } g\in G \tag{6}$$

and

$$\langle t\cdot\phi,\partial F(\phi)\rangle = 0 \quad \text{for all } g\in\mathscr{G}. \tag{7}$$

Equations (7) assure that $\partial F(\phi)\in N_\phi$ and Eqs. (6) that $\partial F(\phi)$ is a G_ϕ-invariant vector.□

The converse statement of Lemma 4 represents the main result of this paper and will be proposed as

Theorem: Let G be a compact Lie group acting orthogonally on $\mathbf{R}^n$. The invariant normal space at $\phi\in\mathbf{R}^n$ is spanned by the gradients at ϕ of the elements of an integrity basis for the ring of all G-invariant polynomials of ϕ.

Proof: Let $\bar{\phi}\in\mathbf{R}^n$ and $\bar{\eta}$ be an arbitrarily chosen vector of $N_{\bar{\phi}}^{(0)}$. We shall define a G-invariant C^∞-function $F_{\bar{\eta}}(\phi)$ whose gradient at $\bar{\phi}$ is proportional to $\bar{\eta}$.

For $a\in\mathbf{R}^1$ let us define

$$\bar{\xi} = \bar{\phi} + a\bar{\eta}. \tag{8}$$

Then $\bar{\xi}\in N_{\bar{\phi}}^{(0)}$ and

$$G_{\bar{\xi}}\supseteq(G_{\bar{\phi}}\cap G_{\bar{\eta}})\supseteq G_{\bar{\phi}}. \tag{9}$$

Moreover, if a is sufficiently small, $\bar{\xi}$ will belong to an open tubular ϵ-neighborhood of $\Omega(\bar{\phi})$, $\mathscr{T}_\epsilon(\Omega(\bar{\phi}))$, where $\rho_{\Omega(\bar{\phi})}$ is defined and is a C^∞-function according to Proposition 1.

Thus, from Lemma 1,

$$\rho_{\Omega(\bar{\phi})}(\bar{\xi}) = \bar{\phi}, \tag{10}$$

which implies, according to Lemma 3, $[\Omega(\bar{\phi})] \geqslant [\Omega(\bar{\xi})]$. From this relation and Eq. (9) we obtain $[\Omega(\bar{\phi})] = [\Omega(\bar{\xi})]$, assuring that $\rho_{\Omega(\bar{\phi})}$ is a one-to-one map from $\Omega(\bar{\xi})$ onto $\Omega(\bar{\phi})$ (see Lemma 3.). As a consequence $\rho_{\Omega}(\bar{\xi})$ will be defined on $\Omega(\bar{\phi})$ (Ref. 9) and, according to Proposition 1, an open tubular neighborhood $\mathscr{T}_{\epsilon'}(\Omega(\bar{\xi})) \ni \bar{\phi}$ will exist, where $\rho_{\Omega(\bar{\xi})}$ is a C^{∞}-function. Moreover, evidently

$$\rho_{\Omega(\bar{\xi})}(\bar{\phi}) = \bar{\xi}. \tag{11}$$

Let us now define the following G-invariant function:

$$F(\phi) = \begin{cases} (1/a)\exp\{\delta(\phi,\Omega(\bar{\xi})) - \epsilon'\}^{-1} & \text{for } \phi \in \mathscr{T}_{\epsilon'}(\Omega(\bar{\xi})) \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

For $\phi \in \mathscr{T}_{\epsilon'}(\Omega(\bar{\xi}))$, one gets from Defs. 3 and 5 and the orthogonality of every matrix $g \in G$,

$$\delta(\phi,\Omega(\bar{\xi})) = \langle \phi,\phi \rangle + \langle \bar{\xi},\bar{\xi} \rangle - 2\langle \phi, \rho_{\Omega(\bar{\xi})}(\phi) \rangle, \tag{13}$$

so that, denoting $\partial/\partial\phi_j$ by ∂_j,

$$\tfrac{1}{2}\partial_j \delta(\phi,\Omega(\bar{\xi})) = \phi_j - (\rho_{\Omega(\bar{\xi})}(\phi))_j - \langle \phi, \partial_j \rho_{\Omega(\bar{\xi})}(\phi) \rangle. \tag{14}$$

Therefore $F(\phi)$ is C^{∞} and its gradient at ϕ can be calculated using Eqs. (14).

The last term in the rhs of Eqs. (14) vanishes owing to Eqs. (C7), (11), (10), amd Lemma 1. Thus, using once again Eq. (11) to evaluate $\rho_{\Omega(\bar{\xi})}(\bar{\phi}) = \bar{\xi}$, and recalling Eq. (8), we get from Eqs. (12) and (14),

$$\partial F(\phi)|_{\phi=\bar{\phi}} = B\bar{\eta}, \tag{15}$$

where B is a nontrivial real constant.

To conclude the proof of the theorem it suffices to recall Schwarz' theorem. In fact, from Eqs. (5) and (15) we get

$$\bar{\eta} = B^{-1}\partial F(\phi)|_{\phi=\bar{\phi}} = B^{-1}\sum_{k=1}^{q} \frac{\partial \hat{F}(\theta)}{\partial \theta_k}\bigg|_{\theta=\theta(\bar{\phi})} \partial\theta_k(\bar{\phi}),$$

so that every vector $\bar{\eta} \in N_{\bar{\phi}}^{(0)}$ can be expressed as a linear combination of the gradients at $\bar{\phi}$ of the elements of an integrity basis.□

ACKNOWLEDGMENTS

I would like to thank Professor R. R. Gatto for his kind hospitality at the Department of Theoretical Physics of the University of Geneva where most of this paper was prepared, and Mrs. M. Prosperi Flaviani for going over the English.

APPENDIX A: LOCAL PARAMETRIZATION OF *G*-ORBITS

A local parametrization for a given G-orbit Ω can be defined in the following way.[10]

Let $\mathscr{G}_{\phi}$ denote the Lie algebra of G_{ϕ}. For each $\xi \in \Omega$ let us choose a basis $B^{(\xi)} = \{t_{\alpha}^{(\xi)}\}_{\alpha=1,\ldots,r}$ for $\mathscr{G}$ so that the following conditions (i) and (ii) are satisfied

(i) $\operatorname{Tr} t_{\alpha}^{(\xi)} t_{\beta}^{(\xi)} = -\delta_{\alpha\beta}$.

(ii) The last $r_{\Omega} = \dim \mathscr{G}_{\xi}$ elements of $B^{(\xi)}$ yield a basis for $\mathscr{G}_{\xi}$.

Conditions (i) and (ii) imply that the adjoint matrix representation of G is orthogonal in the basis $\{t_{\alpha}^{(\xi)}\}_{\alpha=1,\ldots,r}$ and

$$t_{\alpha}^{(\xi)}\cdot\xi = 0 \quad \text{for } \alpha = r - r_{\Omega} + 1,\ldots,r, \tag{A1}$$

while $\{t_{\alpha}^{(\xi)}\cdot\xi\}_{\alpha=1,\ldots,r-r_{\Omega}}$ is a set of independent vectors forming a basis for the vector space T_{ξ}.

As is well known,[10] a $d_{\Omega} > 0$ can be chosen so that for every $\bar{\xi} \in \Omega$ the set of points $\{\xi \in \Omega; \delta(\xi,\bar{\xi}) < d_{\Omega}\}$ is connected and $r - r_{\Omega}$ local coordinates x_{α} can be introduced in the following way:

$$\xi(x) = g_{\bar{\xi}}(x)\cdot\bar{\xi}, \tag{A2a}$$

where $x = (x_1,\ldots,x_{r-r_{\Omega}})$ is in a convenient neighborhood of O and

$$G \ni g_{\bar{\xi}} = \exp\left(\sum_{\alpha=1}^{r-r_{\Omega}} x_{\alpha} t_{\alpha}^{(\bar{\xi})}\right). \tag{A2b}$$

Moreover, d_{Ω} does not depend on $\bar{\xi}$, since $\bar{\xi} \rightarrow g\cdot\bar{\xi}$ is an isometric map from Ω onto Ω for all $g \in G$.

APPENDIX B: THE HESSIAN OF $f(g) = \delta(g\cdot\xi,\phi)$

Let $\hat{K}(\xi,\phi)$ be the $(r - r_{\Omega(\xi)}) \times (r - r_{\Omega(\xi)})$ real symmetric matrix defined by

$$\hat{K}_{\alpha\beta}(\xi,\phi) = \tfrac{1}{2}\langle t_{\alpha}^{(\xi)}\cdot\xi, t_{\beta}^{(\xi)}\cdot\phi \rangle + (\alpha \leftrightarrow \beta), \tag{B1}$$

$$\alpha,\beta = 1,\ldots,r - r_{\Omega(\xi)}, \quad \xi,\phi \in \mathbf{R}^n,$$

where $t_{\alpha}^{(\xi)}$ is defined in Appendix A. $\hat{K}(\xi,\phi)$ is a continuous function of ϕ and a positive definite matrix for $\phi = \xi$, owing to condition (ii) of Appendix A. Therefore, when ϕ is in a convenient neighborhood of ξ, $\hat{K}(\xi,\phi) > 0$. Below we shall also prove that for all $g \in G$:

$$\hat{K}(g\cdot\xi, g\cdot\phi) = \hat{O}(g,\xi)\hat{K}(\xi,\phi)\hat{O}^{T}(g,\xi), \tag{B2}$$

where $\hat{O}(g,\xi)$ is a $(r - r_{\Omega(\xi)}) \times (r - r_{\Omega(\xi)})$ real orthogonal matrix and $\hat{O}^{T}$ denotes its transpose. As a consequence the rank of $\hat{K}(\xi,\phi)$ will be a constant along the G-orbit in $\mathbf{R}^n \times \mathbf{R}^n$ defined by $\{g\cdot(\xi,\phi) = (g\cdot\xi, g\cdot\phi)\}_{g\in G}$.

Thus we have proved

Lemma B 1: For each orbit Ω a positive number d'_{Ω} exists, so that $\hat{K}(\xi,\phi) > 0$ whenever $\xi \in \Omega$ and $\delta(\xi,\phi) < d'_{\Omega}$.

It remains for us to prove Eq. (B2).

Proof of Eq. (B 2): From Eq. (B1) and the G invariance of the scalar product in $\mathbf{R}^n$ one immediately gets

$$\hat{K}_{\alpha\beta}(g\cdot\xi, g\cdot\phi) = \tfrac{1}{2}\langle \{g^{-1}t_{\alpha}^{(g\cdot\xi)}g\}\cdot\xi, \{g^{-1}t_{\beta}^{(g\cdot\xi)}g\}\cdot\phi \rangle. \tag{B3}$$

The r matrices $\{g^{-1}t_{\alpha}^{(g\cdot\xi)}g\}$, $\alpha = 1,\ldots,r$ yield an orthonormal basis for $\mathscr{G}$, satisfying conditions (i) and (ii) of Appendix A. Therefore,

$$g^{-1}t_{\alpha}^{(g\cdot\xi)}g = \sum_{\beta=1}^{r-r_{\Omega(\xi)}} \hat{O}_{\alpha\beta}(g,\xi)t_{\beta}^{(\xi)},$$

$$\alpha = 1,\ldots,r - r_{\Omega(\xi)}, \tag{B4}$$

where $\hat{O}(g,\xi)$ is a real orthogonal matrix. Equation (B2) is an immediate consequene of Eqs. (B3) and (B4).□

APPENDIX C: PROOF OF PROPOSITION 1

We must show that if ϕ is near enough to an orbit Ω, there is only one point on Ω at minimal distance from ϕ and its location is a C^{∞}-function of ϕ.

Let d_{Ω} and d'_{Ω} be as defined in Appendix A and

Lemma B1 and call D_Ω the smaller of the numbers $\frac{1}{2}d_\Omega$ and d'_Ω,

$$D_\Omega = \min\{\tfrac{1}{2}d_\Omega, d'_\Omega\}. \tag{C1}$$

Thus, whenever $\bar{\phi}\in\mathscr{T}_{D_\Omega}(\Omega)$, all the points of Ω at minimal distance from $\bar{\phi}$ belong to the set $\mathscr{D}$:

$$\mathscr{D} = \{\,\xi\in\Omega;\ \delta(\xi,\bar{\phi}) < D_\Omega\,\}. \tag{C2}$$

Let $\bar{\xi}$ be one of these points. Then $\mathscr{D}$ can be parametrized as in Eq. (A2) and the points of Ω at minimal distance from $\bar{\phi}$ can be determined as the points where the following function of x,

$$\begin{aligned} F(x;\bar{\phi},\bar{\xi}) &= 2\langle\,\xi(x),\bar{\phi}\,\rangle \\ &= \langle\bar{\phi},\bar{\phi}\,\rangle + \langle\,\xi(x),\xi(x)\rangle - \delta(\,\xi(x),\bar{\phi}\,) \\ &= \langle\bar{\phi},\bar{\phi}\,\rangle + \langle\,\bar{\xi},\bar{\xi}\,\rangle - \delta(\,\xi(x),\bar{\phi}\,), \end{aligned} \tag{C3}$$

is maximal, for x sufficiently near the origin and $\xi(x)$ defined in Eq. (A2a).

From Eq. (C3), the skew symmetry of $t\in\mathscr{G}$, and Eq. (B1) we get

$$\frac{\partial^2 F(x;\bar{\phi},\bar{\xi})}{\partial x_\alpha \partial x_\beta} = -\hat{K}_{\alpha\beta}(\,\xi(x),\bar{\phi}\,), \tag{C4}$$

so that, according to Lemma B1, the Hessian of $F(x;\bar{\phi},\bar{\xi})$ is negative definite for $\xi(x)\in\mathscr{D}$. As a consequence there will only be one extremum of $F(x;\bar{\phi},\bar{\xi})$ in $\mathscr{D}$, necessarily a maximum.

The above arguments show that for every $\bar{\phi}\in\mathscr{T}_{D_\Omega}(\Omega)$, and for ϕ in a convenient neighborhood of $\bar{\phi}$, there is a unique point on Ω at minimal distance from ϕ. It can be written in the form of Eq. (A2), where x is determined as the unique solution of the extremum conditions:

$$F_\alpha^{(\bar{\xi})}(x;\phi) = \frac{\partial F(x;\phi,\bar{\xi})}{\partial x_\alpha} = 0, \quad \alpha = 1,\ldots,r - r_\Omega. \tag{C5}$$

As

(a) $F_\alpha^{(\bar{\xi})}(0;\bar{\phi}) = 0$, by assumption;

(b) $F_\alpha^{(\bar{\xi})}(x;\phi)$ is C^∞ both in x and ϕ in a neighborhood of $(0,\bar{\phi})$;

(c) the matrix $\partial_\beta F_\alpha^{(\bar{\xi})}(0;\bar{\phi}) = -\hat{K}_{\alpha\beta}(\,\bar{\xi},\bar{\phi}\,)$, $\alpha,\beta = 1,\ldots,r - r_\Omega$, is positive definite; from the implicit function theorem, Eqs. (C5) will admit a solution $x = x(\phi)$ which is a C^∞-function in a neighborhood of $\bar{\phi}$ and satisfies $x(0) = 0$.

As previously argued this solution is unique and, from Def. 5, in a convenient neighborhood of $\bar{\phi}$ it satisfies

$$\rho_\Omega(\phi) = \xi\,(x(\phi)) = \exp\left(\sum_{\alpha=1}^{r-r_\Omega} x_\alpha(\phi)\,t_\alpha^{(\bar{\xi})}\right)\cdot\bar{\xi}, \tag{C6}$$

and $x(\bar{\phi}) = 0.\square$

By differentiating Eqs.(C6) we also get

$$\frac{\partial\rho_{\Omega(\bar{\xi})}(\phi)}{\partial\phi_j}\bigg|_{\phi=\bar{\phi}} = \sum_{\alpha=1}^{r-r_{\Omega(\bar{\xi})}} \frac{\partial x_\alpha(\phi)}{\partial\phi_j}\bigg|_{\phi=\bar{\phi}} t_\alpha^{(\bar{\xi})}\cdot\bar{\xi},$$

which implies

$$\partial\rho_\Omega(\phi)|_{\phi=\bar{\phi}}\in\mathscr{T}_{\rho_\Omega(\bar{\phi})}. \tag{C7}$$

---

[1] L. Michel, C. R. Acad. Sci. Ser. A **272**, 433 (1971); L. Michel and L. A. Radicati, Ann. Phys. **66**, 758 (1971).

[2] Ling Fong Li, Phys. Rev. D **9**, 1723 (1973).

[3] M. Abud and G. Sartori, Phys. Lett. B **104**, 147 (1981).

[4] See for instance L. Michel, CERN Preprint TH 2716 (1979).

[5] This is the original part of Theorem 1 in Ref. 3.

[6] R. S. Palais, Mem. Amer. Math. Soc., No. 36 (1960).

[7] See for instance: H. Weyl, *The Classical Groups*, 2nd ed. (Princeton U.P., Princeton, 1946) or J. A. Dieudonné and J. B. Carrel, Adv. Math. **4**, 1 (1970).

[8] G. W. Schwarz, Topology **14**, 63 (1975).

[9] In fact $\rho_{\Omega(\bar{\xi})}(\phi) = \rho_{\Omega(\bar{\phi})}^{-1}(\phi)\cap\Omega(\bar{\xi})$.

[10] J. L. Koszul, *Lectures on Groups of Transformations*, (Tata Inst. for Fundamental Research) *Lectures on Mathematics*, No. 32 (Tata Inst., Bombay, 1965).

Internal labels of degenerate representations[a]

T. H. Seligman[b]
Division de Physique Théorique[c], Institut de Physique Nucléaire, 91406 Orsay Cedex, France
R. T. Sharp[d]
Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France

(Received 3 June 1982; accepted for publication 12 November 1982)

An expression for the number of internal labels of degenerate irreducible representations of compact semisimple Lie groups is given in the same spirit as Racah's formula for the nondegenerate case.

PACS numbers: 02.20.Qs

I. INTRODUCTION

Years ago Racah[1] gave the useful formula

$$b = \tfrac{1}{2}(r - l) \tag{1}$$

for the number b of internal labels needed to specify the states of a general irreducible representation (IR) of a compact semisimple group; here r is the order of the group (dimension of the group manifold, or number of generators) and l is its rank (dimension of its weight space).

We point out a simple and direct interpretation of Racah's result. An arbitrary state can be constructed by the successive application of lowering generators (those corresponding to negative roots) to the highest weight state of an IR; we may fix their order of application, since different orderings are related by commutators. Then the exponents of the lowering generators, of which there are $\frac{1}{2}(r - l)$, provide the necessary labels.

Our purpose in this paper is to obtain a result for degenerate representations similar to Racah's. A degenerate representation is one for which one or more Cartan (or Dynkin) representation labels are zero. From the above interpretation of Racah's result we readily conjecture that the number of labels needed is just the number of lowering generators which yield a nonzero result when applied to the highest weight state of the degenerate representation. In Sec. II, we prove this conjecture and give a simple rule for calculating the number of labels.

There are many reasons why this result is of interest. We often encounter a group chain which has missing labels for general representations, but may or may not define states completely for degenerate representations of physical interest. Counting labels for the representations involved will decide that and similar questions.

A convenient way of presenting branching rules is provided by generating functions,[2] and their structure is directly related to the number of labels involved. If we are interested only in particular degenerate representations, knowledge of the number of labels provides important clues for the construction of the relevant generating functions. In Sec. III, we provide new examples of generating functions for branching rules, as well as of generating functions for polynomial tensors where a similar situation holds.

Known branching rules for complete chains of groups permit the counting of internal labels for degenerate representations of the classical groups and of the exceptional group G_2. However, this procedure is tedious and is not available for the higher exceptional groups.

II. COUNTING INTERNAL LABELS

In Sec. I we conjectured that by rejecting lowering generators which annihilate the highest state of a degenerate IR and counting the rest we get the correct number of internal labels. The conjecture requires proof because the polynomial independence of the remaining lowering generators, acting on states of a degenerate IR, is not obvious. The proof is based on Weyl's character formula.[3] First, we recall a few needed definitions.

To label an IR, we use Dynkin, or Cartan, labels λ_i defined by

$$\lambda_i = 2\langle M_\lambda | \alpha_i \rangle / \langle \alpha_i | \alpha_i \rangle \,, \tag{2}$$

where M_λ is the highest weight of the IR and α_i are the simple roots, in terms of which any positive root can be expressed as a nonnegative linear combination. An alternative definition of λ_i is through the expansion of M_λ as a nonnegative linear combination of fundamental weights W_i:

$$M_\lambda = \sum_i \lambda_i W_i \,. \tag{3}$$

In order that Eqs. (2) and (3) be consistent we must have

$$\langle W_j | \alpha_i \rangle = \tfrac{1}{2}\delta_{ij} \langle \alpha_i | \alpha_i \rangle \,. \tag{4}$$

Weyl's formula for the character of an IR λ may be written[3]

$$\chi_\lambda = \xi_\lambda / \Delta \,, \tag{5}$$

where the numerator is the characteristic function

$$\xi_\lambda = \sum_S (-1)^S x^{S(M_\lambda + R)} \,. \tag{6}$$

The sum in Eq. (6) is over Weyl reflections S; M_λ is the highest weight of the IR and R is half the sum of the positive roots; x^A means $\Pi_i x_i^{A_i}$, where the x_i are dummy variables which carry as exponents the components A_i of a weight A.

---

[a] Work supported by the Natural Science and Engineering Research Council of Canada and by the Ministère de L'Education du Québec.

[b] Permanent address: Instituto de Física, Universidad Nacional Autónoma de México, México DF, Mexico.

[c] Laboratoire associé au C.N.R.S.

[d] On leave from Physics Department, McGill University, Montréal, Canada.

The denominator Δ of (5) is the characteristic of the scalar IR,

$$\Delta = \sum_S (-1)^S x^{SR} = \prod_\epsilon (x^{\epsilon/2} - x^{-\epsilon/2}) . \tag{7}$$

The product in the factored form of Δ is over positive roots ϵ. We may remove a factor $\Pi_\epsilon x^{\epsilon/2} = x^R$ from numerator and denominator of Eq. (5) with the result

$$\chi_\lambda = \frac{x^{-R}\Sigma_S(-1)^S x^{S(M_\lambda + R)}}{\Pi_\epsilon(1 - x^{-\epsilon})} . \tag{8}$$

First consider general IR's and suppose that all the representation labels λ_i are large. For the purpose of counting labels we need to consider only states in the neighborhood of the highest one. Then we may ignore all S but the identity in the numerator of (8), getting

$$\chi_\lambda \cong x^{M_\lambda} \Big/ \prod_\epsilon (1 - x^{-\epsilon}) . \tag{9}$$

The expansion of (9) correctly counts states in and on the boundary of the hyperparallelopiped defined by the highest state of λ and the other l vertices of the weight diagram nearest to it, confirming our statement that internal states are correctly counted by labeling them with exponents of negative generators applied to the highest state. Outside the parallelopiped cancellations with the neglected numerator terms reduces the number of states, implying that those generated by (9) are no longer independent there.

Now let us deal with a degenerate IR $(\lambda_1, \lambda_2, ..., \lambda_l)$ for which one or more λ_i vanish.

The negative roots which lead outside the weight diagram when applied to the highest state and must therefore be rejected as carriers of internal labels are those which are orthogonal to M_λ (any root ϵ not orthogonal to M_λ implies a Weyl reflection which sends M_λ to another vertex M'_λ of the weight diagram such that the edge from M_λ to M'_λ is parallel to ϵ). The roots orthogonal to M_λ are those whose expansions $\epsilon_j = \Sigma_i n_{ij} \alpha_i$ contain only α_i corresponding to vanishing representation labels λ_i [use (3) and (4)]. This affords a direct method of counting the rejected roots.

It is easy to see that the rejected roots are just the negative roots of the largest semisimple subgroup H of G which leaves the highest state invariant. Patera[4] has pointed out to us that H is just the subgroup corresponding to the Dynkin diagram obtained from that of G by retaining only the vertices (and lines joining them) corresponding to vanishing labels of the degenerate IR; see Figs. 1–4. We give illustrative examples in the next section. For the number of internal labels we get

$$b_\lambda = \tfrac{1}{2}(r_G - l_G) - \tfrac{1}{2}(r_H - l_H) . \tag{10}$$

In particular, when one representation label vanishes the number of internal labels is reduced by one.

We complete this section by showing that all internal labels other than those already rejected are actually required. We start with Weyl's character formula, Eq. (8). Again consider states in the neighborhood of the highest state and suppose that the nonvanishing representation labels are sufficiently large. In the numerator of (3) we need to retain only those Weyl reflections S corresponding to roots orthogonal to M_λ, and those generated by them, i.e., the Weyl reflections S' of the subgroup H of the preceding paragraph; those reflections must be retained because M_λ is near the corresponding reflection planes. We get

$$\chi_\lambda = x^{-R} \sum_{S'} (-1)^{S'} x^{S'(M_\lambda + R)} \Big/ \prod_\epsilon (1 - x^{-\epsilon}) . \tag{11}$$

The remarks following Eq. (9) concerning its region of validity apply equally to Eq. (11). Write $R = R_1 + R_2$ where R_1 is half the sum of the positive roots of H and R_2 is half the sum of the other positive roots of G. Then $S'M_\lambda = M_\lambda$ and $S'R_2 = R_2$. We find

$$\chi_\lambda \cong x^{M_\lambda - R_1} \sum_{S'} (-1)^{S'} x^{S'R_1} \Big/ \prod_\epsilon (1 - x^{-\epsilon}) . \tag{12}$$

Now according to Eq. (7) we may write

$$\sum_{S'} (-1)^{S'} x^{S'R_1} = x^{R_1} \prod_{\epsilon'} (1 - x^{-\epsilon'}) , \tag{13}$$

where the product is over positive roots ϵ' of the subgroup H; the negative roots to be rejected are $-\epsilon'$. We get the desired result

$$\chi_\lambda \cong \frac{x^{M_\lambda}}{\Pi'(1 - x^{-\epsilon})} , \tag{14}$$

where now the product excludes negative roots which lead outside the weight diagram when applied to the highest weight.

III. ILLUSTRATIVE EXAMPLES

In this section we give a few (new) generating functions which illustrate the utility of our results.

The first two deal with tensors whose components are polynomials in the components of the basic spin tensor of SO(8) or SO(12).

For SO(8) the spinor transforms by the IR (0001), and for polynomial tensors we find, by looking at low degrees, the generating function

$$F(U,D) = [(1 - UD)(1 - U^2)]^{-1} . \tag{15}$$

When F is expanded, $F = \Sigma_{ud} U^u D^d C_{ud}$, the coefficient C_{ud} is the multiplicity of the IR $(000d)$ among tensors of degree u.

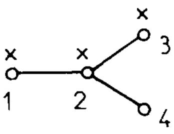


FIG. 1. Dynkin diagram of SO(8); the labels 1,2,3 pick out an SU(4) subgroup.

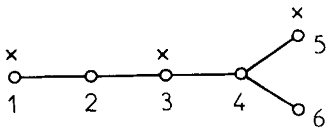


FIG. 2. Dynkin diagram of SO(12); the labels 1,3,5 pick out an SU(2)×SU(2)×SU(2) subgroup.

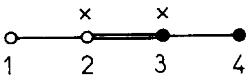


FIG. 3. Dynkin diagram of F_4; the labels 2,3 pick out an SO(5) subgroup.

Tensors other than the degenerate ones of type (000*d*) do not arise.

Figure 1 shows the Dynkin diagram for SO(8). The vertices marked *X*, which correspond to vanishing labels, constitute the Dynkin diagram for SU(4). Since SU(4) requires $\frac{1}{2}(15-3)=6$ internal labels, the generating function (15) should have $8-6=2$ denominator factors. Eight is the dimension of the original (0001) tensor; six is the number of internal labels for (000*d*) [12 internal labels for SO(8) minus 6 for SU(4)].

As a second example we consider polynomial tensors in the components of a (000001) tensor of SO(12). For the generating function we find

$$F(U,B,D,G)=[(1-UG)(1-U^2B)(1-U^3G)\times(1-U^4D)(1-U^4)]^{-1}. \qquad (16)$$

When *F* is expanded, the coefficient of $U^uB^bD^dG^g$ gives the multiplicity of tensors of degree *u* transforming by the degenerate representation (0*b* 0*d* 0*g*); no others occur as polynomials in (000001). The Dynkin diagram of SO(12) is shown in Fig. 2. The vanishing labels, marked *X*, correspond to SU(2)×SU(2)×SU(2). Considerations similar to those of the preceding paragraph show that five is here the correct number of denominator factors.

In the preceding two examples, since classical groups were involved, the counting of internal labels of degenerate IR's could have been done with the help of known branching rules for subgroup chains. Our remaining examples involve F_4 and E_7, for which complete branching rules are not known for any subgroup.

Consider branching rules of $F_4\supset\mathrm{SO}(9)$ for the degenerate IR's (*a*00*d*) of F_4. By looking at low IR's of F_4 we find the generating function

$$F(A,D;E,G,H,J)=[(1-AG)(1-AJ)(1-DE)(1-DJ)\times(1-D)(1-ADH)]^{-1}. \qquad (17)$$

The coefficient of $A^aD^dE^eG^gH^hJ^j$ in the expansion of (17) is the multiplicity of the SO(9) IR (*eghj*) in the F_4 IR (*a*00*d*). The Dynkin diagram for F_4 is shown in Fig. 3. The vanishing labels, marked *X*, imply the subgroup SO(5) with 4 internal labels. Hence the number of internal labels for these degenerate IR's of F_4 is $\frac{1}{2}(52-4)-4=20$. To this we must add 2, the number of nonzero F_4 representation labels and subtract $\frac{1}{2}(36-4)=16$, the number of internal SO(9) labels, to get the number of denominator factors in (17), namely 6.

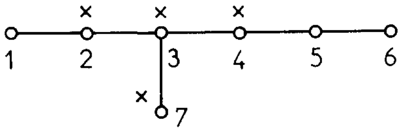


FIG. 4. Dynkin diagram of E_7; the labels 2,3,4,7 pick out an SO(8) subgroup.

As a final example we consider polynomial tensors in the 56-dimensional (0000010) tensor of E_7. We find, heuristically,

$$F(U,A,E,G)=[(1-UG)(1-U^2A)(1-U^3G)\times(1-U^4E)(1-U^4)]^{-1}, \qquad (18)$$

whose expansion gives, as the coefficient of $U^uA^aE^eG^g$, the number of (*a*000*eg*0) tensors of degree *u*; tensors other than these degenerate ones do not arise. The Dynkin diagrams of E_7 is shown in Fig. 4; the vanishing labels, marked *X*, pick out the SO(8) subgroup. Considerations like those of the preceding paragraph show that five is here the correct number of denominator factors.

ACKNOWLEDGMENTS

The authors are grateful for helpful discussions with P. Deligne and J. Patera. The branching tables of McKay and Patera[5] were of great assistance with the examples of Sec. III.

[1] G. Racah, Ergeb. Exacten. Naturwiss. **37**, 28 (1965).

[2] R. Gaskell, A. Peccia, and R. T. Sharp, J. Math. Phys. **19**, 127 (1978); J. Patera and R. T. Sharp, "Generating functions for characters of group representations" in *Lecture Notes in Physics* (Springer, New York, 1979), Vol. 94.

[3] H. Weyl, Math. Z. **23**, 271 (1925); **24**, 328, 377 (1926).

[4] J. Patera, private communication.

[5] W. McKay and J. Patera, *Tables of Dimensions, Second and Fourth Indices and Branching Rules of Representations of Simple Lie Algebras* (Dekker, New York, 1981).

Analysis of the outer product for the symmetric group

L. J. Somers
Institute for Theoretical Physics, University of Nijmegen, Nijmegen, The Netherlands

(Received 23 August 1982; accepted for publication 24 September 1982)

Expressions are derived to write the basis vectors for an irreducible representation μ of the symmetric group in terms of basis vectors for irreducible representations whose outer product yields μ.

PACS numbers: 02.20.Qs

I. INTRODUCTION

It has been noticed[1,2] that the symmetric group can be used to calculate recoupling coefficients for special unitary groups SU(N). The most obvious approach is to study the properties of the representations of the symmetric group in a tensor space. For this it is necessary to consider the outer product of the symmetric group in some detail. In particular, one must know how to express the basis vectors for an irreducible representation (irrep) μ in basis vectors belonging to irreps whose outer product gives μ. The factors which give these relations are called outer coefficients. These outer coefficients are very important because the recoupling coefficients for SU(N) can be written[3] as products of outer coefficients and Clebsch–Gordan coefficients for the symmetric group independent of N.

The outer coefficients can be calculated in a number of ways. The first possibility is to use projection operators and the matrix form of the representations of the symmetric group. This is done in Sec. II. The second method generates the outer coefficients for S_p recursively from the outer coefficients for S_{p-1}. Sections IV and V deal with this method. Section VI gives a graphical rule for a few special cases. Our notation for the representations of the symmetric group is given in Appendix A.

II. OUTER COEFFICIENTS

Suppose μ is an irreducible representation (irrep) of S_p. It is defined in a vector space $V(\mu)$ with an orthonormal basis $e_M^{(\mu)}$. The matrix elements of μ are written as

$$D^{(\mu)}(s)e_M^{(\mu)} = \sum_{M'} e_{M'}^{(\mu)} D^{(\mu)}_{M'M}(s) \tag{1}$$

for all elements s of S_p. We choose the standard form for the vectors $e_M^{(\mu)}$. Standard means in this context that the basis vectors are labeled with Yamanouchi symbols M and that the matrix elements of μ are in the "Young's orthogonal form" [see Appendix A, Eq. (40)]. We restrict ourselves to those elements s of S_p that are also contained in the subgroup $S_{p_1} \times S_{p_2}$ with $p_1 + p_2 = p$. Then we may write $s = s_1 s_2$, where s_1 and s_2 are elements of S_{p_1} and S_{p_2}. The operators $D^{(\mu)}(s_1s_2)$ for all $s_1 \in S_{p_1}$ and $s_2 \in S_{p_2}$ form a representation of the subgroup $S_{p_1} \times S_{p_2}$. This representation is in general reducible. It can be reduced completely in irreps $\kappa \times \lambda$ of $S_{p_1} \times S_{p_2}$. This means that we can construct subspaces $V(\kappa \times \lambda; \mu\gamma)$ of $V(\mu)$ that are invariant under μ. The restriction of μ to such a subspace $V(\kappa \times \lambda; \mu\gamma)$ is equivalent to $\kappa \times \lambda$. We need the extra index γ to distinguish the different equivalent subspaces.

In each of the invariant subspaces we choose a properly adapted basis for the product $\kappa \times \lambda$. These basis vectors $e_{KL}^{(\kappa\times\lambda;\mu\gamma)}$ are also orthonormal. The two orthonormal bases in $V(\mu)$ are connected by a unitary transformation. The matrix elements of this transformation we call outer coefficients. The relation is written as

$$e_{KL}^{(\kappa\times\lambda;\mu\gamma)} = \sum_M S^{\kappa\lambda\mu_\gamma}_{KLM} e_M^{(\mu)} \tag{2a}$$

or

$$e_M^{(\mu)} = \sum_{\kappa\lambda KL\gamma} S^{\kappa\lambda\mu_\gamma *}_{KLM} e_{KL}^{(\kappa\times\lambda;\mu\gamma)}, \tag{2b}$$

where $S^{\kappa\lambda\mu_\gamma}_{KLM}$ is an outer coefficient for $S_{p_1} \times S_{p_2} \subset S_p$. It appears that the phases of the basis vectors $e_{KL}^{(\kappa\times\lambda;\mu\gamma)}$ can be chosen in such a way that the outer coefficients are real. Since the outer coefficients are elements of a unitary matrix, they satisfy the following orthogonality relations:

$$\sum_M S^{\kappa\lambda\mu_\gamma}_{KLM} S^{\kappa'\lambda'\mu_{\gamma'}*}_{K'L'M} = \delta(\kappa,\kappa')\delta(\lambda,\lambda')\delta(K,K')\delta(L,L')\delta(\gamma,\gamma')$$

and

$$\sum_{\kappa\lambda KL\gamma} S^{\kappa\lambda\mu_\gamma}_{KLM} S^{\kappa\lambda\mu_\gamma *}_{KLM'} = \delta(M,M'). \tag{3}$$

The asterisk, denoting complex conjugation, is superfluous when the coefficients are real, as is the case for the symmetric group. From now on we will omit this asterisk everywhere.

The problem is: How to calculate these outer coefficients? Or to state it differently: How to construct the basis vectors $e_{KL}^{(\kappa\times\lambda;\mu\gamma)}$? For this we use the projection and shift operators defined in Appendix B. They are equal to

$$P^{(\kappa\times\lambda)}_{KL,K'L'} = \frac{f(\kappa)f(\lambda)}{p_1!\,p_2!} \sum_{s_1s_2} D^{(\kappa\times\lambda)}_{KL,K'L'}(s_1s_2) D^{(\mu)}(s_1s_2), \tag{4}$$

where s_1 and s_2 are elements of S_{p_1} and S_{p_2} respectively. $D^{(\mu)}(s_1s_2)$ is the representation of $S_{p_1} \times S_{p_2}$ subduced from μ. We use only real matrix elements for the irreps of the symmetric group. Therefore, we have omitted the complex conjugation. The dimensions of the representations κ and λ are written as $f(\kappa)$ and $f(\lambda)$. Applying the shift operator to a basis vector $e_{M'}^{(\mu)}$ yields

$$
\begin{aligned}
&P^{(\kappa\times\lambda)}_{KL,\,11}\,e^{(\mu)}_{M'}\\
&\quad=\frac{f(\lambda)}{p_2!}\sum_{s_2,\,M} D^{(\lambda)}_{L,\,1}(s_2)D^{(\mu)}_{M,\,KM'(p_2)}(s_2)\\
&\qquad\times\delta(\kappa,\mu/M'(p_2))\delta(1,M'(p_1))e^{(\mu)}_{M}\,. \qquad (5)
\end{aligned}
$$

In this equation the labels $K', L' = 1, 1$ correspond to the first Yamanouchi symbols in the standard ordering for κ and λ (see Appendix A). We have used the fact that the permutations s_1 and s_2 commute. Furthermore, the following general orthogonality relation has been used for the matrix elements of irreps of a group G of order $f(G)$.

$$
\sum_{g\in G} D^{(\mu)}_{MN}(g)D^{(\mu')}_{M'N'}(g)^* = \frac{f(G)}{f(\mu)}\delta(\mu,\mu')\delta(M,M')\delta(N,N')\,, \qquad (6)
$$

where $f(\mu)$ is the dimension of the irrep μ.

According to the prescription given in Appendix B all we have to do is:

—Apply $P^{(\kappa\times\lambda)}_{11,11}$ to all vectors $e^{(\mu)}_{M'}$.

—Orthonormalize the result. This means that the orthonormal vectors $e^{(\kappa\times\lambda;\,\mu\gamma)}_{11}$ are the result of the action of the projection operator upon a certain linear combination of vectors $e^{(\mu)}_{M'}$. They can be expressed as

$$
e^{(\kappa\times\lambda;\,\mu\gamma)}_{11} = P^{(\kappa\times\lambda)}_{11,11}\sum_{M'}\alpha(\gamma,M')e^{(\mu)}_{M'} = \sum_{M} S^{\kappa\lambda\mu_\gamma}_{11M}e^{(\mu)}_{M}\,. \qquad (7)
$$

—Let the other shift operators act upon the same linear combination of vectors $e^{(\mu)}_{M'}$. Again the resulting vectors $e^{(\kappa\times\lambda;\,\mu\gamma)}_{KL}$ are expressed as a linear combination of vectors $e^{(\mu)}_{M}$

$$
e^{(\kappa\times\lambda;\,\mu\gamma)}_{KL} = P^{(\kappa\times\lambda)}_{KL,11}\sum_{M'}\alpha(\gamma,M')e^{(\mu)}_{M'} = \sum_{M} S^{\kappa\lambda\mu_\gamma}_{KLM}e^{(\mu)}_{M}\,. \qquad (8)
$$

It is possible to simplify the outer coefficients. In Appendix A we show that for elements s_2 of S_{p_2} the matrix elements of the representation μ only depend upon the part of the Young diagram associated with $M(p_2)$. So if we define κ as being the diagram belonging to the tableau K, the following relation holds for the matrix elements appearing in (5):

$$
D^{(\mu)}_{M,\,KM'(p_2)}(s_2) = \delta(K, M(p_1))D^{(\mu/\kappa)}_{M(p_2),\,M'(p_2)}(s_2)\,. \qquad (9)
$$

The number $\delta(K, M(p_1))$ can always be factored out of $S^{\kappa\lambda\mu_\gamma}_{KLM}$ in a trivial fashion. This means that the outer coefficients do not really depend upon K. Therefore, we will represent the outer coefficient by the notation $\left(\begin{smallmatrix}\lambda & \mu/\kappa & \gamma\\ L & M(p_2) & \end{smallmatrix}\right)$

$$
S^{\kappa\lambda\mu_\gamma}_{KLM} = \delta(K, M(p_1))\begin{pmatrix}\lambda & \mu/\kappa & \gamma\\ L & M(p_2) & \end{pmatrix}. \qquad (10)
$$

We will now study the case in which there is no degeneracy γ present and derive an expression for the outer coefficients. When the product is not degenerate, it is sufficient to choose one vector $e^{(\mu)}_{M'}$ for which the result of the projection operator $P^{(\kappa\times\lambda)}_{11,11}$ is unequal to zero. The normalization is then carried out by dividing by the norm N of the result. The square of this norm is equal to

$$
\begin{aligned}
N^2 &= \frac{f(\lambda)}{p_2!}\sum_{s_2} D^{(\lambda)}_{1,1}(s_2)D^{(\mu)}_{1M'(p_2),1M'(p_2)}(s_2)\\
&= \frac{f(\lambda)}{p_2!}\sum_{s_2} D^{(\lambda)}_{1,1}(s_2)D^{(\mu/\kappa)}_{M'(p_2),\,M'(p_2)}(s_2)\,. \qquad (11)
\end{aligned}
$$

The result of the other shift operators must be divided by the same norm. So the outer coefficient will be

$$
S^{\kappa\lambda\mu}_{KLM} = \frac{1}{N}\frac{f(\lambda)}{p_2!}\sum_{s_2} D^{(\lambda)}_{L,1}(s_2)D^{(\mu)}_{M,\,KM'(p_2)}(s_2)\,. \qquad (12)
$$

For the shorthand outer coefficients defined in (10) a similar formula holds:

$$
\begin{pmatrix}\lambda & \mu/\kappa\\ L & M(p_2)\end{pmatrix} = \frac{1}{N}\frac{f(\lambda)}{p_2!}\sum_{s_2} D^{(\lambda)}_{L,1}(s_2)D^{(\mu/\kappa)}_{M(p_2),\,M'(p_2)}(s_2)\,. \qquad (13)
$$

We have fixed an overall phase by choosing some particular M' and dividing out the norm (instead of the norm times some phase factor). It turns out that in this particular case the sign of the result is independent of the choice of M' (provided, of course, that the result is unequal to zero).

Consider now the degenerate case. We adopt the following phase convention: any nonzero outer coefficient which has the following properties is positive:

—it must contain the first L in the standard ordering of the different L's belonging to λ;

—it has the first possible M for μ (that means the outer coefficient is nonzero).

III. SOME PROPERTIES OF THE OUTER COEFFICIENTS

We will derive some useful equations for the outer coefficients. Apply $D^{(\mu)}(s_1s_2)$ to (2b), where s_1 is an element of S_{p_1} and s_2 of S_{p_2}. For the left-hand side of the equation this results in

$$
\sum_{M'} e^{(\mu)}_{M'}D^{(\mu)}_{M'M}(s_1s_2) = \sum_{\substack{\kappa\lambda\gamma\\ KLM'}} S^{\kappa\lambda\mu_\gamma}_{KLM'}\,e^{(\kappa\times\lambda;\,\mu\gamma)}_{KL}D^{(\mu)}_{M'M}(s_1s_2)\,; \qquad (14)
$$

for the right-hand side we find

$$
\sum_{\substack{\kappa\lambda\gamma\\ KL}} S^{\kappa\lambda\mu_\gamma}_{KLM}\sum_{K'L'} e^{(\kappa\times\lambda;\,\mu\gamma)}_{K'L'}D^{(\kappa)}_{K'K}(s_1)D^{(\lambda)}_{L'L}(s_2)\,. \qquad (15)
$$

Putting (14) and (15) together, we find, after removing the vector from the equation, interchanging the left- and right-hand side and choosing $s_1 = e$,

$$
\sum_{L'} S^{\kappa\lambda\mu_\gamma}_{KL'M}D^{(\lambda)}_{LL'}(s_2) = \sum_{M'} S^{\kappa\lambda\mu_\gamma}_{KLM'}D^{(\mu)}_{M'M}(s_2)\,. \qquad (16)
$$

From now one we will use the shorthand notation given in (10) for the outer coefficients. We also introduce the abbreviation:

$$
\nu \equiv \mu/\kappa \qquad (17)
$$

for the skew-symmetric Young diagram found by subtracting κ from μ. The label N is used to denote the corresponding part of the Yamanouchi symbol M [we used to write this in the form $M(p_2)$]. Equation (16) will then look like

$$
\begin{aligned}
&\sum_{L'}\begin{pmatrix}\lambda & \nu & \gamma\\ L' & N & \end{pmatrix}D^{(\lambda)}_{LL'}(s_2)\\
&\qquad= \sum_{N'}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N' & \end{pmatrix}D^{(\nu)}_{N'N}(s_2)\,. \qquad (18)
\end{aligned}
$$

In the following we will also omit the argument s_2 from the representation matrices D. Shifting the outer coefficient to the right yields

$$D^{(\lambda)}_{LL'}\delta(\lambda,\lambda')\delta(\gamma,\gamma')$$

$$=\sum_{NN'}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N' & \end{pmatrix}D^{(\nu)}_{N'N}\begin{pmatrix}\lambda' & \nu & \gamma'\\ L' & N & \end{pmatrix}. \tag{19}$$

We can shift the outer coefficient in (18) to the left:

$$\sum_{\lambda LL'\gamma}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N' & \end{pmatrix}D^{(\lambda)}_{LL'}\begin{pmatrix}\lambda & \nu & \gamma\\ L' & N & \end{pmatrix}=D^{(\nu)}_{N'N}\,. \tag{20}$$

IV. RECURSION COEFFICIENTS

The method described in Sec. II to calculate the outer coefficients has the disadvantage that for larger values of p_2 the work becomes extremely time-consuming. To solve this problem, we show that the outer coefficients for a given p_2 can be calculated recursively from coefficients for p_2-1. To obtain the recursion coefficients which relate the outer coefficients for p_2 and p_2-1, one has to solve a simple set of coupled linear equations.

Consider the elements s_2 of S_{p_2} which leave p invariant. They form a subgroup S_{p_2-1} of S_{p_2}. For these elements we may write

$$D^{(\lambda)}_{LL'}(s_2)=\delta(L_p,L'_p)D^{(\lambda/L_p)}_{L_*L'_*}(s_2)$$

and

$$D^{(\nu)}_{NN'}(s_2)=\delta(N_p,N'_p)D^{(\nu/N_p)}_{N_*N'_*}(s_2)\,. \tag{21}$$

We have introduced here the subscript asterisk to denote that the last number of a Yamanouchi symbol M has been omitted: $M_*\equiv M_1\cdots M_{p-1}$. Inserting the restriction (21) into (19), one finds

$$D^{(\lambda/L_p)}_{L_*L'_*}\delta(L_p,L'_p)\delta(\lambda,\lambda')\delta(\gamma,\gamma')$$

$$=\sum_{NN'}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N' & \end{pmatrix}D^{(\nu/N_p)}_{N'_*N_*}\begin{pmatrix}\lambda' & \nu & \gamma'\\ L' & N & \end{pmatrix}\delta(N_p,N'_p)\,. \tag{22}$$

We apply now Eq. (20) to representations λ/L_p or ρ/R_p of S_{p_2-1} and ν/N_p of S_{p-1} (limited to the last p_2-1 objects). The corresponding Yamanouchi symbols are L_* or R_* and N_*. We find

$$D^{(\nu/N_p)}_{N'_*N_*}=\sum_{\substack{\rho/R_p,\beta\\ R_*,R'_*}}\begin{pmatrix}\rho/R_p & \nu/N_p & \beta\\ R_* & N'_* & \end{pmatrix}$$

$$\times D^{(\rho/R_p)}_{R_*R'_*}\begin{pmatrix}\rho/R_p & \nu/N_p & \beta\\ R'_* & N_* & \end{pmatrix}. \tag{23}$$

One now inserts (23) in (22) and shifts two outer coefficients to the left-hand side to find

$$\sum_{L'_*,N_*}\begin{pmatrix}\lambda & \nu & \gamma\\ L' & N & \end{pmatrix}\begin{pmatrix}\rho/R_p & \nu/N_p & \beta\\ R'_* & N_* & \end{pmatrix}D^{(\lambda/L_p)}_{L_*L'_*}$$

$$=\sum_{R_*,N'_*}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N' & \end{pmatrix}\begin{pmatrix}\rho/R_p & \nu/N_p & \beta\\ R_* & N'_* & \end{pmatrix}D^{(\rho/R_p)}_{R_*R'_*}, \tag{24}$$

where $L'_p=L_p$ and $N'_p=N_p$ is assumed. The above equation can also be written in the following form:

$$\sum_{L'_*}D^{(\lambda/L_p)}_{L_*L'_*}Z(\lambda,L_p,\nu,N_p,\gamma,\rho/R_p,\beta)_{L'_*R'_*}$$

$$=\sum_{R_*}Z(\lambda,L_p,\nu,N_p,\gamma,\rho/R_p,\beta)_{L_*R_*}D^{(\rho/R_p)}_{R_*R'_*}, \tag{25}$$

where we have defined

$$Z(\lambda,L_p,\nu,N_p,\gamma,\rho/R_p,\beta)_{L_*R_*}$$

$$\equiv\sum_{N_*}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N & \end{pmatrix}\begin{pmatrix}\rho/R_p & \nu/N_p & \beta\\ R_* & N_* & \end{pmatrix}. \tag{26}$$

Now we suppress all indices which are the same in the left- and right-hand side of (25). The simplified notation is

$$\sum_{L'_*}D^{(\lambda/L_p)}_{L_*L'_*}Z_{L'_*R'_*}=\sum_{R_*}Z_{L_*R_*}D^{(\rho/R_p)}_{R_*R'_*}\,. \tag{27}$$

The above equation is in fact nothing else than a matrix equation for D and Z. We will apply Schur's lemma to (27). This lemma says that from (27) follows that either Z is zero when $\rho/R_p\neq\lambda/L_p$ or else Z is a multiple of the unit matrix when $\rho/R_p=\lambda/L_p$. Therefore,

$$Z(\lambda,L_p,\nu,N_p,\gamma,\rho/R_p,\beta)_{L_*R_*}$$

$$=\delta(\lambda/L_p,\rho/R_p)\delta(L_*,R_*)R^{\lambda\nu\gamma}_{L_pN_p\beta}\,. \tag{28}$$

Now we fill in the definition (26) of Z:

$$\sum_{N_*}\begin{pmatrix}\lambda & \nu & \gamma\\ L & N & \end{pmatrix}\begin{pmatrix}\rho/R_p & \nu/N_p & \beta\\ R_* & N_* & \end{pmatrix}$$

$$=\delta(\lambda/L_p,\rho/R_p)\delta(L_*,R_*)R^{\lambda\nu\gamma}_{L_pN_p\beta}\,. \tag{29}$$

Shifting one outer coefficient to the right-hand side of the equation yields the recursion relation we were looking for:

$$\begin{pmatrix}\lambda & \nu & \gamma\\ L & N & \end{pmatrix}=\sum_{\beta}R^{\lambda\nu\gamma}_{L_pN_p\beta}\begin{pmatrix}\lambda/L_p & \nu/N_p & \beta\\ L_* & N_* & \end{pmatrix} \tag{30a}$$

or

$$S^{\kappa\lambda\mu_\gamma}_{KLM}=\sum_{\beta}R^{\lambda\ \mu/\kappa\ \gamma}_{L_pM_p\beta}S^{\kappa\ \lambda/L_p\ \mu/M_p\ \beta}_{KL_*M_*}\,. \tag{30b}$$

With the help of (30a) or (30b) we are able to calculate all outer coefficients for p_2 when the outer coefficients for p_2-1 and the recursion coefficients $R^{\lambda\nu\gamma}_{L_pN_p\beta}$ are known. For $p_2=1$ we have

$$\begin{pmatrix}[1] & [1]\\ 1 & 1\end{pmatrix}=1\,. \tag{31}$$

It is straightforward to prove that the recursion coefficients satisfy the following orthogonality relations:

$$\sum_{N_p\beta}R^{\lambda\nu\gamma}_{L_pN_p\beta}R^{\lambda'\nu\gamma'}_{L'_pN_p\beta}\delta(\lambda/L_p,\lambda'/L'_p)$$

$$=\delta(\lambda,\lambda')\delta(L_p,L'_p)\delta(\gamma,\gamma') \tag{32a}$$

and

$$\sum_{\lambda L_p\gamma}R^{\lambda\nu\gamma}_{L_pN_p\beta}R^{\lambda\nu\gamma}_{L_pN'_p\beta'}\delta(\lambda/L_p,\lambda'/L'_p)$$

$$=\delta(N_p,N'_p)\delta(\beta,\beta')\,. \tag{32b}$$

In (32b) the factor $\delta(\lambda/L_p,\lambda'/L'_p)$ means that one has to sum all λ and L_p for which λ/L_p is equal to some given λ'/L'_p.

V. RELATIONS FOR THE RECURSION COEFFICIENTS

In this section we derive a set of equations that can be used to calculate the recursion coefficients. Consider the transposition $(p-1,p)$. According to Eq. (40) the matrix element for this element of the symmetric group is given by

$$\begin{aligned}&D^{(\lambda)}_{L'L}(p-1,p)\\&=\sigma(\lambda,L_p,L_{p-1})\delta(L'_p,L_p)\delta(L'_{p-1},L_{p-1})\delta(L'_{\bullet\bullet},L_{\bullet\bullet})\\&\quad+\tau(\lambda,L_p,L_{p-1})\delta(L'_p,L_{p-1})\delta(L'_{p-1},L_p)\delta(L'_{\bullet\bullet},L_{\bullet\bullet})\end{aligned}\tag{33}$$

and the same with ν, N instead of λ, L. We have used the notation $M_{\bullet\bullet}\equiv M_1\cdots M_{p-2}$ for a Yamanouchi symbol M with the last two numbers omitted. Furthermore, σ is the inverse of the axial distance ρ defined in (A2) and $\tau=(1-\sigma^2)^{1/2}$. Inserting (33) in (18) yields

$$\begin{aligned}&\{\sigma(\lambda,L_p,L_{p-1})-\sigma(\nu,N_p,N_{p-1})\}\begin{pmatrix}\lambda&\nu&\gamma\\L&N&\end{pmatrix}\\&\quad+\tau(\lambda,L_p,L_{p-1})\begin{pmatrix}\lambda&\nu&\gamma\\L_{\bullet\bullet}L_pL_{p-1}&N&\end{pmatrix}\\&=\tau(\nu,N_p,N_{p-1})\begin{pmatrix}\lambda&\nu&\gamma\\L&N_{\bullet\bullet}N_pN_{p-1}&\end{pmatrix}.\end{aligned}\tag{34}$$

We have used $\tau(\lambda,L_p,L_{p-1})=\tau(\lambda,L_{p-1},L_p)$. For the recursion coefficients we find

$$\begin{aligned}&\{\sigma(\lambda,L_p,L_{p-1})-\sigma(\nu,N_p,N_{p-1})\}\sum_\beta R^{\lambda\nu\gamma}_{L_pN_p\beta}\begin{pmatrix}\lambda/L_p&\nu/N_p&\beta\\L_\bullet&N_\bullet&\end{pmatrix}\\&\quad+\tau(\lambda,L_p,L_{p-1})\sum_\beta R^{\lambda\nu\gamma}_{L_{p-1}N_p\beta}\begin{pmatrix}\lambda/L_{p-1}&\nu/N_p&\beta\\L_{\bullet\bullet}L_p&N_\bullet&\end{pmatrix}\\&=\tau(\nu,N_p,N_{p-1})\sum_\beta R^{\lambda\nu\gamma}_{L_pN_{p-1}\beta}\begin{pmatrix}\lambda/L_p&\nu/N_{p-1}&\beta\\L_\bullet&N_{\bullet\bullet}N_p&\end{pmatrix}.\end{aligned}\tag{35}$$

The above set of linear equations can be solved using also the orthogonality relations (32). For each λ and ν it will have as many independent solutions as there are equivalent subspaces $V(\kappa\times\lambda;\mu\gamma)$ of $V(\mu)$. These solutions are distinguished from each other by the label γ.

VI. GRAPHICAL RULES

A graphical rule to calculate the recursion coefficients for the case that $\lambda=[p]$ or $\lambda=[1^p]$ can be given.

Consider first the case $\lambda=[p]$. Suppose one wants to calculate the recursion coefficient $R^{\lambda\nu}_{L_pN_p}$. For $\lambda=[p]$ (and also for $\lambda=[1^p]$) there are no degeneracy labels γ and β present.

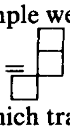
Choose the first possible $N_\bullet$ in the standard ordering to form a Yamanouchi symbol N with N_p. Then the above recursion coefficient is equal to

$$R^{\lambda\nu}_{L_pN_p}=(p)^{-1/2}\prod_{q\neq p}[1+\sigma(\nu,N_p,N_q)]^{1/2}.\tag{36}$$

For $\lambda=[1^p]$ the formula is

$$R^{\lambda\nu}_{L_pN_p}=\epsilon(s)(p)^{-1/2}\prod_{q\neq p}[1-\sigma(\nu,N_p,N_q)]^{1/2},\tag{37}$$

where $\epsilon(s)$ is the sign of the permutation s which transforms the first Yamanouchi symbol in the standard ordering into the Yamanouchi symbol N. The label q runs from p_1+1 to $p-1$.

As an example we consider $p=5$, $p_2=3$, $\nu=[221]/[11]=$  and $N_p=2$. Then $N_\bullet=13$. So the permutation s which transforms the first Yamanouchi symbol for ν (which is equal to 123) into $N=132$ is equal to $s=(23)$. Therefore, s is odd. The inverse axial distances involved are

$$\sigma(2,1)=-1\quad\text{and}\quad\sigma(2,3)=\tfrac{1}{2}.$$

Therefore, for $\lambda=[1^3]$ and $L_p=3$,

$$R^{\lambda\nu}_{L_pN_p}=-(3)^{-1/2}(1+1)^{1/2}(1-\tfrac{1}{2})^{1/2}=-(\tfrac{1}{3})^{1/2}.$$

ACKNOWLEDGMENTS

The author would like to thank Dr. T. A. Rijken and Professor J. J. de Swart for giving useful suggestions. Part of this work was included in the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM) with financial support from the Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO).

APPENDIX A: THE SYMMETRIC GROUP

1. General remarks

A Young diagram $\mu=[\mu_1\cdots\mu_p]$ is a figure containing p boxes ordered in p rows of length μ_i, with the properties:

$$\mu_1\geqslant\cdots\geqslant\mu_p>0\quad\text{and}\quad\mu_1+\cdots+\mu_p=p.\tag{A1}$$

A standard Young tableau is a diagram which contains the numbers 1 to p in such a way that the numbers in each row increase from left to right and in each column increase from top to bottom.

Each standard tableau can be written in a compact way by a Yamanouchi symbol $M=M_1\cdots M_p$. This is an array of p numbers, the M_i being the rows in the standard tableau in which the number i appears. For example, the standard Young tableaux and Yamanouchi symbols for the diagram [31] are given by

$$\begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4\\\cline{1-1}\end{array}=1112,\quad\begin{array}{|c|c|c|}\hline 1&2&4\\\hline 3\\\cline{1-1}\end{array}=1121$$

and

$$\begin{array}{|c|c|c|}\hline 1 & 3 & 4 \\ \hline 2 \\ \cline{1-1} \end{array} = 1211 \,.$$

Note that our notation differs from the one used by Hamermesh.[4] The Yamanouchi symbols belonging to some diagram (and therefore the corresponding Young tableaux) can be ordered. The symbol M comes before the symbol N if $M < N$ when the symbols are regarded as composite numbers (lexicographic ordering).

Consider two boxes x and y in a Young diagram μ. Box x is at the position (a,b), where a denotes row and b column, and y at (c,d). The axial distance $\rho(\mu; x, y)$ between these boxes is equal to

$$\rho(\mu; x, y) = (b-a) - (d-c) \,. \tag{A2}$$

It is the number of steps (horizontal or vertical) from x to y. The steps are counted positive going down or to the left and negative when going up or to the right.

The different irreducible representations (irreps) of S_p can be represented by Young diagrams. Let the irrep μ of S_p be defined in a vector space $V(\mu)$. The orthonormal basis vectors $e_M^{(\mu)}$ in this space can be labeled by the Yamanouchi symbols $M = M_1 \cdots M_p$. The matrices of the transpositions $(i, i+1)$ in the "Young's orthogonal form"[5,6] are given by

$$D^{(\mu)}(i,i+1) e^{(\mu)}_{M_1 \cdots M_p} = \sigma e^{(\mu)}_{M_1 \cdots M_p} + (1-\sigma^2)^{1/2} e^{(\mu)}_{M_1 \cdots M_{i+1} M_i \cdots M_p} \,, \tag{A3}$$

where $\sigma \equiv 1/\rho(\mu; M_{i+1}, M_i)$ is the inverse axial distance between the boxes corresponding to M_{i+1} and M_i.

2. Subgroup representations

Consider subgroups S_{p_1} and S_{p_2} of S_p, where $p_1 + p_2 = p$. Here S_{p_1} and S_{p_2} are the permutation groups of the first p_1 objects and the last p_2 objects. The Yamanouchi symbol M for the symmetric group S_p is adapted for the subgroups S_{p_1} and S_{p_2}

$$M \equiv M_1 \cdots M_{p_1} M_{p_1+1} \cdots M_p \equiv M(p_1) M(p_2) \,. \tag{A4}$$

The part $M(p_1)$ of the Yamanouchi symbol M forms again a valid Yamanouchi symbol for the group S_{p_1}. It belongs to a Young tableau for S_{p_1} which can be obtained from the tableau M of S_p by removing the boxes with the numbers $p_1 + 1, \ldots, p$. This new tableau for S_{p_1} belongs to a Young diagram that is denoted as $\mu/M(p_2)$ or $\mu/M_{p_1+1} \cdots M_p$. From (A3) it follows immediately that the matrix elements of an irreducible representation of any transposition of the subgroup S_{p_1} depend only upon that part of the Young diagram where the numbers $1, \ldots, p_1$ have been put. Hence the same holds for general permutations in the subgroup S_{p_1}. So the elements s_1 of S_{p_1} leave the last p_2 numbers in the Yamanouchi symbol invariant. We have

$$\begin{aligned} D^{(\mu)}(s_1) e^{(\mu)}_{M(p_1)M(p_2)} &= \sum_{M'(p_1)} e^{(\mu)}_{M'(p_1)M(p_2)} D^{(\mu)}_{M'(p_1)M(p_2),\, M(p_1)M(p_2)}(s_1) \\ &= \sum_{M'(p_1)} e^{(\mu)}_{M'(p_1)M(p_2)} D^{(\mu/M(p_2))}_{M'(p_1),\, M(p_1)}(s_1) \,. \end{aligned} \tag{A5}$$

For transpositions (and also for general elements) of the subgroup S_{p_2} holds analogously that the matrix elements of the irreducible representations only depend upon the form of that part of the diagram where the numbers $p_1 + 1, \ldots, p$ have been placed. These elements s_2 of S_{p_2} leave the first p_1 numbers in the Yamanouchi symbol invariant:

$$D^{(\mu)}(s_2) e^{(\mu)}_{M(p_1)M(p_2)} = \sum_{M'(p_2)} e^{(\mu)}_{M(p_1)M'(p_2)} D^{(\mu/M(p_1))}_{M'(p_2),\, M(p_2)}(s_2) \,. \tag{A6}$$

The skew-symmetric diagram obtained from the diagram μ by omitting the boxes $M_1, \ldots, M_{p_1}$ will be written down as

$$\mu/M(p_1) \quad \text{or} \quad \mu/M_1 \cdots M_{p_1} \quad \text{or} \quad \mu/\kappa \,, \tag{A7}$$

where κ is the Young diagram which corresponds to the boxes which contain the numbers $1, \ldots, p_1$.

APPENDIX B: PROJECTION AND SHIFT OPERATORS

Consider a vector space V where a representation D of a group G is defined. Suppose D contains the irrep μ $\Gamma(\mu)$ times. The shift operator[7] is defined by

$$P^{(\mu)}_{MM'} = \frac{f(\mu)}{f(G)} \sum_g D^{(\mu)}_{MM'}(g)^* D(g) \,, \tag{B1}$$

where $f(\mu)$ is the dimension of μ and $f(G)$ is the number of elements g of G. P is a projection operator if $M = M'$. When the basis vectors of the irreducible subspaces are given by $e_N^{(\nu\gamma)}$, the following relations hold:

$$\begin{aligned} P^{(\mu)}_{MM'} e^{(\nu\gamma)}_N &= \delta(\mu, \nu) \delta(M', N) e^{(\nu\gamma)}_M \,, \\ P^{(\mu)}_{MM'} P^{(\nu)}_{NN'} &= \delta(\mu, \nu) \delta(M', N) P^{(\mu)}_{MN'} \,. \end{aligned} \tag{B2}$$

The procedure for constructing a basis for the irreducible subspaces is as follows:
—apply $P^{(\mu)}_{1,1}$ to all vectors of V (1 means the first basis vector;
—orthonormalize the result; this will lead to vectors $e_1^{(\mu\gamma)}$, where γ runs from 1 to $\Gamma(\mu)$;
—determine the vectors

$$e_M^{(\mu\gamma)} = P^{(\mu)}_{M,1} e_1^{(\mu\gamma)} \,; \tag{B3}$$

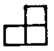
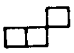
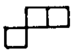
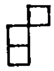
—the vector spaces $V^{(\mu\gamma)}$ spanned by the vectors $e_M^{(\mu\gamma)}$ will form invariant subspaces of V, such that the restriction of D to $V^{(\mu\gamma)}$ is equivalent to μ.

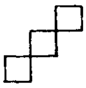
APPENDIX C: TABLE OF RECURSION COEFFICIENTS

We have tabulated the recursion coefficients $R^{\lambda\nu\gamma}_{L_pN_p\beta}$ for $p_2 = 2$ and $p_2 = 3$. In these cases there is no degeneracy label β present. In Table I we denote the inverse of the axial distance between the box in the lowest left corner of a diagram ν and the box in the upper right corner by x. For example, ν_1 (see Fig. 1) could stand for μ_1 where x is equal to $-\frac{1}{3}$. One sees here that when boxes of a diagram ν in the table touch only at the corners they can be shifted with respect to each other. For the diagram ν_2 (see Fig. 1) the situation is somewhat more complicated. Number the boxes according to the first Yamanouchi symbol. This means that 1 is the upper

TABLE I. Table of recursion coefficients.

$p_2 = 2$		λ	[2]	[11]
ν	N_2	L_2	1	2
[2]	1		1	0
[11]	2		0	1
	2		$(1+x)/2$	$(1-x)/2$
	1		$(1-x)/2$	$-(1+x)/2$

$p_2 = 3$		λ	[3]	[21]	[21]	[111]
ν	N_3	L_3	1	2	1	3
[3]	1		1	0	0	0
	2		0	1	1	0
	2		$\frac{2(1+2x)}{3(1+x)}$	$\frac{(1-x)}{3(1+x)}$	1	0
	1		$\frac{(1-x)}{3(1+x)}$	$-\frac{2(1+2x)}{3(1+x)}$	0	0
[21]	2		0	1	0	0
	1		0	0	1	0
	2		$(1+2x)/3$	$2(1-x)/3$	0	0
	1		$2(1-x)/3$	$-(1+2x)/3$	1	0
[111]	3		0	0	0	1
	3		0	1	$-(1+2x)/3$	$2(1-x)/3$
	1		0	0	$2(1-x)/3$	$(1+2x)/3$
	3		0	0	$\frac{2(1+2x)}{3(1+x)}$	$\frac{(1-x)}{3(1+x)}$
	2		0	1	$\frac{(1-x)}{3(1+x)}$	$-\frac{2(1+2x)}{3(1+x)}$

$\nu =$ 

λ	L_3	γ	N_3	$R^{\lambda\nu\gamma}_{L_3N_3}$
[3]	1	·	3	$(1+y)(1+x)/3$
			2	$(1+z)(1-y)/3$
			1	$(1-z)(1-x)/3$
[111]	3	·	3	$(1-y)(1-x)/3$
			2	$-(1-z)(1+y)/3$
			1	$(1+z)(1+x)/3$
[21]	2	1	3	$a/(3(y+z))$
			2	$-(1+z)(1+y)(1-y)(1+x)(y+z)/(3a)$
			1	$-(1-z)(1+y)(1+x)(1-x)(y+z)/(3a)$
	1	1	3	$-(1+z)(1-z)y^4/(a(y+z))$
			2	$(1+z)(1+y)(1-y)(1-x)(y+z)/a$
			1	$(1-z)(1-y)(1+x)(1-x)(y+z)/a$
	2	2	3	0
			2	$(1-z)(1-x)(y+z)/a$
			1	$-(1+z)(1-y)(y+z)/a$
	1	2	3	$4(1+y)(1-y)(1+x)(1-x)(y+z)/(3a)$
			2	$(1-z)(1+x)(1-2y)^2(y+z)/(3a)$
			1	$-(1+z)(1+y)(1-2x)^2(y+z)/(3a)$

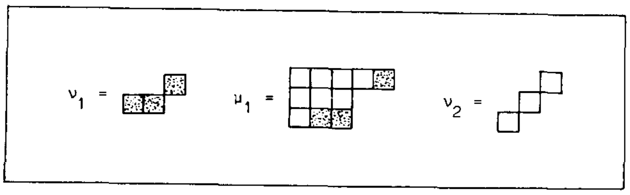


FIG. 1. Graphical representation of the diagrams ν_1, μ_1, and ν_2.

box, 2 is the middle box, and 3 is the lower box. Then x is the inverse axial distance from 3 to 1, y from 3 to 2, and z from 2 to 1. The variables x, y, and z are related via

$$1/x = 1/y + 1/z\,.$$

Throughout the table we have used the abbreviation

$$a \equiv -zy^2 - 2zy + 2z - y^2 + 2y\,.$$

The recursion coefficients in the table yield outer coefficients with phases according to the convention of Sec. II. A $\surd$ must be added over each entry in the table. For example, $-(1+x)/2$ means $-[(1+x)/2]^{1/2}$.

---

[1] John J. Sullivan, J. Math. Phys. **19**, 1674, 1681 (1978).

[2] Jin-Quan Chen, J. Math. Phys. **22**, 1 (1981).

[3] L. J. Somers, Nijmegen Report THEF-NYM-82.13, 1982, to be submitted to J. Math. Phys.

[4] M. Hamermesh, *Group Theory* (Addison-Wesley, Reading, Mass., 1964).

[5] G. de B. Robinson, *Representation Theory of the Symmetric Group* (University of Toronto Press, Toronto, 1961).

[6] H. Boerner, *Representations of Groups* (North-Holland, Amsterdam, 1970).

[7] W. Miller, *Symmetry Groups and Their Applications* (Academic, New York, 1972).

Complementary group with respect to SO(n)

G. Couvreur, J. Deenen, and C. Quesne[a]
Physique Théorique et Mathématique CP 229, Université Libre de Bruxelles, Bd. du Triomphe, B 1050 Brussels, Belgium

(Received 20 July 1982; accepted for publication 12 November 1982)

We look for a complementary group with respect to SO(n) within either irreducible representation $\langle \frac{1}{2}^{dn} \rangle$ or $\langle \frac{1}{2}^{dn-1} \frac{3}{2} \rangle$ of the group Sp($2dn,R$) of linear canonical transformations in a $2dn$-dimensional phase space. We prove that: (i) such a group is Sp($2d,R$) when $n = 2q + 1$ or $n = 2q > 2d$; (ii) it is SU(d,d) when $n = 2$ and $d > 1$; (iii) it does not exist when $n = 2$, $d = 1$, or $2 < n = 2q \leqslant 2d$.

PACS numbers: 02.20.Qs, 02.20.Rt

I. INTRODUCTION

The notion of complementary groups, introduced by Moshinsky and Quesne some years ago,[1] plays an important part in physical applications. Two groups G_1 and G_2, whose direct product is a subgroup of a larger group H, are referred to as complementary within a definite irreducible representation μ of H, if there is a one-to-one correspondence between all the irreducible representations λ_1 and λ_2 of G_1 and G_2 contained in this irreducible representation of H. Then all the basis states of the irreducible representation μ of H which transform irreducibly in the same way under G_1, i.e., belong to the same row of equivalent irreducible representations of G_1 characterized by λ_1, form a basis for the irreducible representation λ_2 of G_2, and vice versa. In other words, the multiplicity of any irreducible representation of one of the subgroups, contained in the irreducible representation μ of H, is equal to the dimensionality of the corresponding irreducible representation of the complementary subgroup.

Various complementary subgroups of the real symplectic group Sp($2N,R$)[2] are known when N factorizes into a product of two integers d and n.[2-5] Such are the full orthogonal group O(n) and the real symplectic group Sp($2d,R$), corresponding to the group chain Sp($2dn,R$)$\supset$Sp($2d,R$)$\times$O(n).[2,4] However, the complementarity relationship with respect to Sp($2d,R$) does not remain in general valid when O(n) is restricted to its rotation subgroup SO(n). This property, which was already implicitly contained in Ref. 4, was recently stressed by some of the authors.[6] The purpose of the present paper is to further investigate this point and to look for a subgroup of Sp($2dn,R$), complementary with respect to SO(n), in those cases where Sp($2d,R$) does not fulfill the complementarity requirements.

In Sec. II, we begin by establishing that the complementarity relationship between SO(n) and Sp($2d,R$) does not hold if and only if n is even and not larger than $2d$. Sections III–V are devoted to the construction of a complementary group with respect to SO(2). In Sec. III, we show that SU(d,d)$\times$SO(2) is a subgroup of Sp($4d,R$). We build the Lie algebra of SU(d,d) in Sec. IV and prove the complementarity of SO(2) and SU(d,d) for $d > 1$ in Sec. V. By generalizing the procedure followed in Sec. IV, we demonstrate in Sec. VI that no subgroup of Sp($2dn,R$) is complementary with respect to SO(n) for even values of n such that $2 < n \leqslant 2d$. Finally, Sec. VIII summarizes the conclusions.

II. DISCUSSION OF THE COMPLEMENTARY RELATIONSHIP BETWEEN SO(n) AND Sp($2d,R$)

The real symplectic group Sp($2dn,R$) is the group of linear canonical transformations in a $2dn$-dimensional phase space.[2] In this space, the coordinates and momenta are denoted by x_{is} and p_{is}, $i = 1,...,d$ and $s = 1,...,n$. Let us define $\mathbf{x}_s$ and $\mathbf{x}$ by

$$\mathbf{x}_s = \begin{pmatrix} x_{1s} \\ x_{2s} \\ \vdots \\ x_{ds} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}, \tag{2.1}$$

respectively. Similar relations hold for $\mathbf{p}_s$ and $\mathbf{p}$. A linear canonical transformation from the coordinates and momenta x_{is}, p_{is} to new coordinates and momenta $\bar{x}_{is}, \bar{p}_{is}$,

$$\begin{pmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{p}} \end{pmatrix} = \mathbf{S} \begin{pmatrix} \mathbf{x} \\ \mathbf{p} \end{pmatrix}, \tag{2.2}$$

is represented by a $2dn \times 2dn$ real symplectic matrix

$$\mathbf{S} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{2.3}$$

i.e., a real matrix such that

$$\mathbf{S}\mathbf{K}\tilde{\mathbf{S}} = \mathbf{K}, \tag{2.4}$$

where the matrix $\mathbf{K}$ is defined by

$$\mathbf{K} = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix} \tag{2.5}$$

and the tilde stands for transposed. On the right-hand side of Eqs. (2.3) and (2.5), all the submatrices are of dimension $dn \times dn$. The condition (2.4) implies the following restrictions on the real matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$:

$$\mathbf{B}\tilde{\mathbf{A}} = \mathbf{A}\tilde{\mathbf{B}}, \quad \mathbf{C}\tilde{\mathbf{D}} = \mathbf{D}\tilde{\mathbf{C}}, \tag{2.6}$$

and

$$\mathbf{D}\tilde{\mathbf{A}} - \mathbf{C}\tilde{\mathbf{B}} = \mathbf{I}.$$

The generators of Sp($2dn,R$) are more easily written in terms of the boson creation and annihilation operators, defined as usual by

$$\eta_{is} = 2^{-1/2}(x_{is} - ip_{is}), \quad \xi_{is} = 2^{-1/2}(x_{is} + ip_{is}),$$
$$i = 1,...,d, \quad s = 1,...,n, \tag{2.7}$$

whose commutation relations are

$$[\eta_{is},\eta_{jt}] = [\xi_{is},\xi_{jt}] = 0, \quad [\xi_{is},\eta_{jt}] = \delta_{ij}\delta_{st}. \tag{2.8}$$

They are given by the expressions[2]

$$D^{\dagger}_{is,jt} = \eta_{is}\eta_{jt},$$
$$D_{is,jt} = \xi_{is}\xi_{jt}, \tag{2.9}$$

and

$$E_{is,jt} = \tfrac{1}{2}(\eta_{is}\xi_{jt} + \xi_{jt}\eta_{is}) = C_{is,jt} + \tfrac{1}{2}\delta_{ij}\delta_{st},$$

where

$$C_{is,jt} = \eta_{is}\xi_{jt} \tag{2.10}$$

denotes the generators of the U(dn) subgroup. Their commutation relations are given by

$$\begin{aligned}
&[D^{\dagger}_{is,jt},D^{\dagger}_{i's',j't'}] = [D_{is,jt},D_{i's',j't'}] = 0,\\
&[D_{is,jt},D^{\dagger}_{i's',j't'}] = \delta_{ii'}\delta_{ss'}E_{j't',jt} + \delta_{ij'}\delta_{st'}E_{i's',jt}\\
&\qquad\qquad + \delta_{ji'}\delta_{ts'}E_{j't',is} + \delta_{jj'}\delta_{tt'}E_{i's',is},\\
&[E_{is,jt},D^{\dagger}_{i's',j't'}] = \delta_{ji'}\delta_{ts'}D^{\dagger}_{is,j't'} + \delta_{jj'}\delta_{tt'}D^{\dagger}_{is,i's'},\\
&[E_{is,jt},D_{i's',j't'}] = -\delta_{ii'}\delta_{ss'}D_{jt,j't'} - \delta_{ij'}\delta_{st'}D_{jt,i's'},\\
&[E_{is,jt},E_{i's',j't'}] = \delta_{ji'}\delta_{ts'}E_{is,j't'} - \delta_{ij'}\delta_{st'}E_{i's',jt}.
\end{aligned} \tag{2.11}$$

In addition, they satisfy the following symmetry properties:

$$D^{\dagger}_{is,jt} = D^{\dagger}_{jt,is} = [D_{is,jt}]^{\dagger}$$

and

$$E_{is,jt} = [E_{jt,is}]^{\dagger}, \tag{2.12}$$

where the dagger stands for the Hermitian conjugate. The set of all boson states belongs to one of two irreducible representations of the group Sp($2dn,R$), which can be characterized by their lowest weight $\langle(\frac{1}{2})^{dn}\rangle$ or $\langle(\frac{1}{2})^{dn-1}\frac{3}{2}\rangle$, according as the boson number is even or odd.[2]

Let us consider the group chain[2]

$$\mathrm{Sp}(2dn,R) \supset \mathrm{Sp}(2d,R) \times \mathrm{O}(n). \tag{2.13}$$

Here O(n) denotes as usual the full orthogonal group in n dimensions, whose connected piece, the rotation group SO(n), is generated by the operators

$$\Lambda_{st} = -i(C_{st} - C_{ts}), \tag{2.14}$$

where

$$C_{st} = \sum_i C_{is,it} \tag{2.15}$$

is a generator of the U(n) subgroup of U(dn). The generators of Sp($2d,R$) are obtained by contracting those of Sp($2dn,R$) with respect to index s, and are given by

$$\begin{aligned}
D^{\dagger}_{ij} &= D^{\dagger}_{ji} = \sum_s \eta_{is}\eta_{js},\\
D_{ij} &= D_{ji} = (D^{\dagger}_{ij})^{\dagger} = \sum_s \xi_{is}\xi_{js},\\
E_{ij} &= (E_{ji})^{\dagger} = \sum_s (\eta_{is}\xi_{js} + \xi_{js}\eta_{is}) = C_{ij} + \tfrac{1}{2}n\delta_{ij},
\end{aligned} \tag{2.16}$$

where

$$C_{ij} = \sum_s C_{is,js} \tag{2.17}$$

is a generator of the U(d) subgroup of U(dn).

It was shown by Chacón[4] that the irreducible representations of O(n) and Sp($2d,R$), contained in either irreducible representations $\langle(\frac{1}{2})^{dn}\rangle$ or $\langle(\frac{1}{2})^{dn-1}\frac{3}{2}\rangle$ of Sp($2dn,R$), are essentially characterized by the same partition $[\lambda_1,\lambda_2\cdots\lambda_r]$ into r parts, where $r=\min(d,q)$ and $q=[n/2]$ is the rank of O(n). By essentially, we mean that if $[\lambda_1,\lambda_2\cdots\lambda_r]$ specifies an O(n) irreducible representation, then the Sp($2d,R$) corresponding one is characterized by the lowest weight $\langle\lambda_d + n/2,\ldots,\lambda_2 + n/2, \lambda_1 + n/2\rangle$ or $\langle(n/2)^{d-q}, \lambda_q + n/2,\ldots,\lambda_1 + n/2\rangle$ according as $d\leqslant q$ or $d>q$. The groups Sp($2d,R$) and O(n) are therefore complementary within either irreducible representation of Sp($2dn,R$).[2]

When O(n) is restricted to its rotation subgroup SO(n), the irreducible representation $[\lambda_1,\ldots,\lambda_r]$ decomposes according to the following branching rule[7]: $[\lambda_1\cdots\lambda_r]$ remains irreducible if $n=2q+1$ or if $n=2q$ and $\lambda_q=0$, and separates into two irreducible representations, respectively characterized by $[\lambda_1\cdots\lambda_{q-1}\,\lambda_q]$ and $[\lambda_1\cdots\lambda_{q-1}\,-\lambda_q]$, if $n=2q$ and $\lambda_q\neq 0$. Consequently, whenever $n=2q+1$ or $n=2q\rangle 2d$, SO(n) and Sp($2d,R$) are complementary within either irreducible representation of Sp($2dn,R$), and the relation between their associated irreducible representations remains the same as for O(n) and Sp($2d,R$). However, when $n=2q\leqslant 2d$, there is a two-to-one correspondence between the irreducible representations of SO(n) and Sp($2d,R$), which are therefore not complementary.

We are then faced with the following problem: Does a complementary group with respect to SO(n) within either irreducible representation of Sp($2d,R$) exist when $n=2q\leqslant 2d$? We shall answer this question in the following sections, starting with the simplest case corresponding to $n=2$.

III. THE GROUP CHAIN Sp ($4d,R$)⊃SU(d,d)×SO(2)

In the present section, we are going to construct the subgroup of linear canonical transformations in a $4d$-dimensional phase space, which are invariant under SO(2). For this purpose, it is convenient to use creation and annihilation operators in polar coordinates, $\eta_{i\sigma}$ and $\xi_{i\sigma}$, $i=1,\ldots,d$, $\sigma=+,-$, defined in terms of η_{is}, ξ_{is}, $i=1,\ldots,d$, $s=1,2$, by

$$\begin{aligned}
\eta_{i\pm} &= \mp 2^{-1/2}(\eta_{i1} \pm i\eta_{i2}),\\
\xi_{i\pm} &= \mp 2^{-1/2}(\xi_{i1} \mp i\xi_{i2}) = (\eta_{i\pm})^{+}.
\end{aligned} \tag{3.1}$$

Let us introduce

$$\boldsymbol{\eta}_{\sigma} = \begin{pmatrix}\eta_{1\sigma}\\ \eta_{2\sigma}\\ \eta_{d\sigma}\end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix}\boldsymbol{\eta}_{+}\\ \boldsymbol{\eta}_{-}\end{pmatrix}, \tag{3.2}$$

and $\boldsymbol{\xi}_{\sigma}$, $\boldsymbol{\xi}$, defined by similar relations. Note that for simplicity's sake, we use lower indices to denote both the covariant components of $\boldsymbol{\eta}$ and the contravariant ones of $\boldsymbol{\xi}$. By combining Eqs. (2.7) and (3.1), we obtain

$$\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\xi}\end{pmatrix} = \mathbf{U}\begin{pmatrix}\mathbf{x}\\ \mathbf{p}\end{pmatrix}, \tag{3.3}$$

where U is a $4d \times 4d$ unitary matrix, defined by

$$\mathbf{U}=2^{-1/2}\begin{pmatrix}\mathbf{V} & -i\mathbf{V}\\ \mathbf{V}^* & i\mathbf{V}^\dagger\end{pmatrix}, \tag{3.4}$$

in terms of the $2d\times 2d$ unitary matrix

$$\mathbf{V}=2^{-1/2}\begin{pmatrix}-\mathbf{I} & -i\mathbf{I}\\ \mathbf{I} & -i\mathbf{I}\end{pmatrix}. \tag{3.5}$$

Here the asterisk stands for complex conjugate.

The linear canonical transformation (2.2) induces a transformation from the creation and annihilation operators $\eta_{i\sigma}$ and $\xi_{i\sigma}$ to new creation and annihilation operators $\bar{\eta}_{i\sigma}$ and $\bar{\xi}_{i\sigma}$,

$$\begin{pmatrix}\bar{\boldsymbol{\eta}}\\ \bar{\boldsymbol{\xi}}\end{pmatrix}=\mathscr{S}\begin{pmatrix}\boldsymbol{\eta}\\ \boldsymbol{\xi}\end{pmatrix}, \tag{3.6}$$

where

$$\mathscr{S}=\mathbf{U}\mathbf{S}\mathbf{U}^\dagger=\begin{pmatrix}\mathscr{A} & \mathscr{B}\\ \mathscr{C} & \mathscr{D}\end{pmatrix}, \tag{3.7}$$

and $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, $\mathscr{D}$ are $2d\times 2d$ complex matrices. The latter can be easily expressed in terms of the submatrices **A**, **B**, **C**, **D** of Eq. (2.3). We shall, however, not proceed this way, and instead write the conditions to be fulfilled by $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, $\mathscr{D}$. From Eq. (2.4) and the relation

$$\mathbf{U}\mathbf{K}\tilde{\mathbf{U}}=i\mathbf{K}, \tag{3.8}$$

$\mathscr{S}$ must satisfy the same condition as **S**, i.e.,

$$\mathscr{S}\mathbf{K}\tilde{\mathscr{S}}=\mathbf{K}. \tag{3.9}$$

Consequently, $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, $\mathscr{D}$ are restricted by conditions similar to Eq. (2.6). In addition, $\mathscr{S}$ must preserve the Hermiticity properties of the creation and annihilation operators, so that

$$\mathscr{D}=\mathscr{A}^*$$

and

$$\mathscr{C}=\mathscr{B}^*. \tag{3.10}$$

Therefore, $\mathscr{S}$ can be written as

$$\mathscr{S}=\begin{pmatrix}\mathscr{A} & \mathscr{B}\\ \mathscr{B}^* & \mathscr{A}^*\end{pmatrix}, \tag{3.11}$$

where

$$\mathscr{B}\tilde{\mathscr{A}}=\mathscr{A}\tilde{\mathscr{B}},$$

$$\mathscr{A}\mathscr{A}^\dagger-\mathscr{B}\mathscr{B}^\dagger=\mathbf{I}. \tag{3.12}$$

Let us consider now the canonical transformations corresponding to rotations in the two-dimensional space associated with index σ. They are represented by the $4d\times 4d$ matrices

$$\mathscr{R}=\begin{pmatrix}\boldsymbol{\Theta} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Theta}^*\end{pmatrix}, \tag{3.13}$$

where

$$\boldsymbol{\Theta}=\begin{pmatrix}e^{-i\theta}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & e^{i\theta}\mathbf{I}\end{pmatrix}, \tag{3.14}$$

and θ is some real number such that $0\leqslant\theta\leqslant 2\pi$. The canonical transformations invariant under SO(2) are represented by those matrices (3.11) which commute with all the $\mathscr{R}$ matrices. Let us decompose $\mathscr{A}$ and $\mathscr{B}$ into $d\times d$ submatrices $\mathscr{A}_i$, $\mathscr{B}_i$, $i=1,\dots,4$, such that

$$\mathscr{A}=\begin{pmatrix}\mathscr{A}_1 & \mathscr{A}_2\\ \mathscr{A}_3 & \mathscr{A}_4\end{pmatrix},\quad \mathscr{B}=\begin{pmatrix}\mathscr{B}_1 & \mathscr{B}_2\\ \mathscr{B}_3 & \mathscr{B}_4\end{pmatrix}. \tag{3.15}$$

The condition $[\mathscr{R},\mathscr{S}]=0$ leads to the following form for $\mathscr{S}$:

$$\mathscr{S}=\begin{pmatrix}\mathscr{A}_1 & \mathbf{0} & \mathbf{0} & \mathscr{B}_2\\ \mathbf{0} & \mathscr{A}_4 & \mathscr{B}_3 & \mathbf{0}\\ \mathbf{0} & \mathscr{B}_2^* & \mathscr{A}_1^* & \mathbf{0}\\ \mathscr{B}_3^* & \mathbf{0} & \mathbf{0} & \mathscr{A}_4^*\end{pmatrix}, \tag{3.16}$$

where, from Eq. (3.12), $\mathscr{A}_1$, $\mathscr{B}_2$, $\mathscr{B}_3^*$, and $\mathscr{A}_4^*$ are restricted by the conditions

$$\begin{aligned}&\mathscr{A}_1\mathscr{A}_1^\dagger-\mathscr{B}_2\mathscr{B}_2^\dagger=\mathbf{I},\\ &\mathscr{B}_3^*\tilde{\mathscr{B}}_3-\mathscr{A}_4^*\tilde{\mathscr{A}}_4=-\mathbf{I},\\ &\mathscr{A}_1\tilde{\mathscr{B}}_3-\mathscr{B}_2\tilde{\mathscr{A}}_4=\mathbf{0}.\end{aligned} \tag{3.17}$$

The matrices (3.16) form a $4d\times 4d$ matrix representation of the group U(d,d), whose elements are the $2d\times 2d$ matrices

$$\mathscr{T}=\begin{pmatrix}\mathscr{A}_1 & \mathscr{B}_2\\ \mathscr{B}_3^* & \mathscr{A}_4^*\end{pmatrix}, \tag{3.18}$$

satisfying Eq. (3.17). We have therefore established that U(d,d), or more exactly a $4d$-dimensional representation thereof, is a subgroup of Sp$(4d,R)$.

Since SO(2) is an abelian group, any matrix $\mathscr{R}$ belongs to the $4d$-dimensional representation of U(d,d), and corresponds to $\mathscr{T}=e^{-i\theta}\mathbf{I}$. We are thus led to split U(d,d) into the direct product of SO(2) and the unimodular subgroup SU(d,d). We then obtain the following group chain:

$$\mathrm{Sp}(4d,R)\supset\mathrm{U}(d,d)\simeq\mathrm{SU}(d,d)\times\mathrm{SO}(2). \tag{3.19}$$

The Sp$(2d,R)$ group of Eq. (2.13) is a subgroup of SU(d,d) since it is made of those matrices $\mathscr{T}$ for which $\mathscr{A}_1=\mathscr{A}_4$ and $\mathscr{B}_2=\mathscr{B}_3$. Equation (3.19) may therefore be supplemented by the following group chains:

$$\mathrm{SU}(d,d)\supset\mathrm{Sp}(2d,R),\quad \mathrm{SO}(2)\subset\mathrm{O}(2). \tag{3.20}$$

In Sec. V, we shall prove that SU(d,d) solves the complementary problem for SO(2) when $d>1$. Before coming to that point, we shall derive the Lie algebra of SU(d,d) in the next section.

IV. LIE ALGEBRA OF SU(*d*,*d*)

In polar coordinates, the Sp$(4d,R)$ generators are denoted by $D^\dagger_{i\sigma,j\tau}$, $D_{i\sigma,j\tau}$, and $E_{i\sigma,j\tau}$ and are defined by relations similar to Eq. (2.9) with s and t replaced by σ and τ, respectively. Equations (2.11) and (2.12) remain valid for the new generators provided the same substitution is carried out. In the same way, the single generator of SO(2), defined in Eq. (2.14), becomes

$$\Lambda=\Lambda_{12}=\sum_{i\sigma}\epsilon_\sigma\eta_{i\sigma}\xi_{i\sigma}, \tag{4.1}$$

where

$$\begin{aligned}\epsilon_\sigma&=+1\quad\text{if}\quad\sigma=+,\\ &=-1\quad\text{if}\quad\sigma=-.\end{aligned} \tag{4.2}$$

Let us look for those linear combinations of the

Sp(4*d*,*R*) generators which commute with Λ. A straightforward calculation shows that the following operators,

$$P_{i+,j+} = E_{i+,j+}, \quad P_{i+,j-} = D^{+}_{i+,j-},$$
$$P_{i-,j-} = E_{j-,i-}, \quad P_{i-,j+} = D_{j+,i-}, \tag{4.3}$$

fulfill the condition

$$[\Lambda, P_{i\sigma,j\tau}] = 0, \tag{4.4}$$

and that no additional operator commuting with Λ can be formed in the Lie algebra of Sp(4*d*,*R*). From the commutation relations and symmetry properties of the Sp(4*d*,*R*) generators, those of the operators $P_{i\sigma,j\tau}$ are, respectively, obtained as

$$[P_{i\sigma,j\tau}, P_{i'\sigma',j'\tau'}] = g_{j\tau,i'\sigma'} P_{i\sigma,j'\tau'} - g_{j'\tau',i\sigma} P_{i'\sigma',j\tau}, \tag{4.5}$$

and

$$P_{i\sigma,j\tau} = (P_{j\tau,i\sigma})^{\dagger}, \tag{4.6}$$

where the metric $g_{i\sigma,j\tau}$ is equal to

$$g_{i\sigma,j\tau} = \epsilon_{\sigma} \delta_{ij} \delta_{\sigma\tau}. \tag{4.7}$$

The operators $P_{i\sigma,j\tau}$ are therefore the generators of the U(*d*,*d*) subgroup of Sp(4*d*,*R*). The SO(2) generator Λ is just the first-order Casimir operator of U(*d*,*d*),

$$\Lambda = \sum_{i\sigma} g_{i\sigma,i\sigma} P_{i\sigma,i\sigma}, \tag{4.8}$$

and the traceless operators

$$P'_{i\sigma,j\tau} = P_{i\sigma,j\tau} - (2d)^{-1} g_{i\sigma,j\tau} \sum_{k\rho} g_{k\rho,k\rho} P_{k\rho,k\rho} \tag{4.9}$$

generate the SU(*d*,*d*) subgroup.

To express the Sp(2*d*,*R*) generators in terms of those of SU(*d*,*d*), it is most convenient to go back to Cartesian coordinates. In such coordinates, the operators $P_{i\sigma,j\tau}$ can be expressed as

$$\begin{aligned} P_{i+,j+} &= \tfrac{1}{2}(E_{ij} - \Gamma_{ij}), \\ P_{i-,j-} &= \tfrac{1}{2}(E_{ji} + \Gamma_{ji}), \\ P_{i+,j-} &= -\tfrac{1}{2}(D^{\dagger}_{ij} - \Delta^{\dagger}_{ij}), \\ P_{i-,j+} &= -\tfrac{1}{2}(D_{ij} + \Delta_{ij}), \end{aligned} \tag{4.10}$$

in terms of the Sp(2*d*,*R*) generators $D^{\dagger}_{ij}$, D_{ij}, E_{ij} and of some additional operators $\Delta^{\dagger}_{ij}$, Δ_{ij}, Γ_{ij}, defined by

$$\begin{aligned} \Delta^{\dagger}_{ij} &= i \sum_{st} \epsilon_{st} \eta_{is} \eta_{jt}, \\ \Delta_{ij} &= -i \sum_{st} \epsilon_{st} \xi_{is} \xi_{jt}, \\ \Gamma_{ij} &= i \sum_{st} \epsilon_{st} \eta_{is} \xi_{jt}, \end{aligned} \tag{4.11}$$

and satisfying the following symmetry properties:

$$\Delta^{\dagger}_{ij} = -\Delta^{\dagger}_{ji} = (\Delta_{ij})^{\dagger},$$
$$\Gamma_{ij} = (\Gamma_{ji})^{\dagger}. \tag{4.12}$$

In Eq. (4.11), ϵ_{st} is the antisymmetric tensor. From Eqs. (4.10), (2.16), (4.12), and (4.9) the Sp(2*d*,*R*) generators are given by

$$\begin{aligned} D^{\dagger}_{ij} &= -P_{i+,j-} - P_{j+,i-} = -P'_{i+,j-} - P'_{j+,i-}, \\ D_{ij} &= -P_{i-,j+} - P_{j-,i+} = -P'_{i-,j+} - P'_{j-,i+}, \\ E_{ij} &= P_{i+,j+} + P_{j-,i-} = P'_{i+,j+} + P'_{j-,i-}. \end{aligned} \tag{4.13}$$

We know that the Sp(2*d*,*R*) generators are the scalars with respect to O(2) which can be formed from the Sp(4*d*,*R*) generators. In the same way, the operators $\Delta^{\dagger}_{ij}$, Δ_{ij}, and Γ_{ij} are the pseudoscalars which can be built from the same generators. Both types of operators become scalars when restricting to unimodular transformations.[8] It is then clear that to get the Lie algebra of SU(*d*,*d*), we could start from that of Sp(2*d*,*R*) and add to it the pseudoscalars built from creation and annihilation operators. We shall show in Sec. VI that such a procedure can be generalized to SO(*n*) for arbitrary *n* values.

V. COMPLEMENTARY RELATIONSHIP BETWEEN SO(2) AND SU(*d*,*d*)

In the present section, we are going to prove that, whenever $d > 1$, SO(2) and SU(*d*,*d*) are complementary within either irreducible representation $\langle (\frac{1}{2})^{2d} \rangle$ or $\langle (\frac{1}{2})^{2d-1} \frac{3}{2} \rangle$ of Sp(4*d*,*R*). For such a purpose, we shall determine the lowest weight state of the irreducible representations of SU(*d*,*d*) contained in either irreducible representation of Sp(4*d*,*R*), and show that the irreducible representation of SO(2) to which it belongs is in one-to-one correspondence with that of SU(*d*,*d*).

From Eq. (4.5), it is clear that the weight generators of U(*d*,*d*) are the operators $P_{i\sigma,i\sigma}$, where $i = 1,...,d$ and $\sigma = +, -$. If, in accordance with Eq. (3.2), we enumerate the values of the double index $i\sigma$ in the order

$$i\sigma \Rightarrow 1+,\ 2+,...,d+,\ 1-,\ 2-,...,d-, \tag{5.1}$$

the lowering generators of U(*d*,*d*) are the operators $P_{i+,j+}$ $(i > j)$, $P_{j-,i-}$ $(i > j)$, and $P_{i-,j+}$. The lowest weight state $|F\rangle$ of an irreducible representation $[h_{1+},...,h_{d+}, h_{1-},...,h_{d-}]$ of U(*d*,*d*) is then obtained by solving the following system of equations:

$$C_{i\sigma,i\sigma}|F\rangle = (h_{i\sigma} - \tfrac{1}{2})|F\rangle, \tag{5.2a}$$

$$C_{i\sigma,j\sigma}|F\rangle = 0, \ i > j, \tag{5.2b}$$

$$D_{j+,i-}|F\rangle = 0. \tag{5.2c}$$

Let us look for the solution of Eqs. (5.2a), (5.2b), and (5.2c) in a Bargmann space of analytic functions $F(z_{i\sigma})$ in 2*d* complex variables $z_{i\sigma}$, $i = 1,...,d$, $\sigma = +, -$.[9] In such a space, the boson operators $\eta_{i\sigma}$ and $\xi_{i\sigma}$ are represented by $z_{i\sigma}$ and $\partial/\partial z_{i\sigma}$, respectively. Equations (5.2a), (5.2b), and (5.2c) can be rewritten as

$$z_{i\sigma} \frac{\partial}{\partial z_{i\sigma}} F(z_{k\rho}) = (h_{i\sigma} - \tfrac{1}{2}) F(z_{k\rho}), \tag{5.3a}$$

$$z_{i\sigma} \frac{\partial}{\partial z_{j\sigma}} F(z_{k\rho}) = 0, \quad i > j, \tag{5.3b}$$

$$\frac{\partial^2}{\partial z_{j+} \partial z_{i-}} F(z_{k\rho}) = 0. \tag{5.3c}$$

Let us successively consider the conditions imposed on $F(z_{k\rho})$ by Eqs. (5.3a), (5.3b), and (5.3c). Equation (5.3a) implies that F is a monomial of degree $h_{i\sigma} - \frac{1}{2}$ in $z_{i\sigma}$, $i = 1,...,d$, $\sigma = +, -$. All the U(d,d) irreducible representation labels must therefore be half-integer. Equation (5.3b) means that F does not depend upon $z_{i\sigma}$, $i = 1,...,d-1$, $\sigma = +, -$. Then

$$F(z_{i\sigma}) \propto z_{d+}^{h_{d+} - 1/2} z_{d-}^{h_{d-} - 1/2}, \tag{5.4}$$

and the U(d,d) irreducible representation labels must satisfy the condition $h_{i\sigma} = \frac{1}{2}$ for $i = 1,...,d-1$ and $\sigma = +, -$. Finally, Eq. (5.3c) is satisfied if F does not depend upon z_{d+} or $z_{\mathrm{d}-}$ or both. We therefore obtain three different types of solutions according as $h_{d+} = m + \frac{1}{2}$ and $h_{d-} = \frac{1}{2}$, $h_{d+} = \frac{1}{2}$, and $h_{d-} = m + \frac{1}{2}$, or $h_{d+} = h_{d-} = \frac{1}{2}$, where m is any positive integer. We shall denote them by A, B, and C, respectively. They are listed in Table I, together with the corresponding U(d,d) and SU(d,d) irreducible representation labels. We note that, for all d values except $d = 1$, the three classes of SU(d,d) irreducible representations are specified by different labels.

Let us now turn to the SO(2) irreducible representation label characterizing $F(z_{i\sigma})$. In Bargmann space, the single SO(2) generator can be written as

$$\Lambda = \sum_{i\sigma} \epsilon_\sigma z_{i\sigma} \frac{\partial}{\partial z_{i\sigma}}. \tag{5.5}$$

The lowest weight functions $F(z_{i\sigma})$ of SU(d,d) irreducible representations belonging to class A, B, or C are eigenfunctions of Λ corresponding to the eigenvalues m, $-m$, or 0 respectively. When $d > 1$, the SO(2)-irreducible representations characterized by $[m]$ and $[-m]$ where m is any positive integer, are therefore associated with two inequivalent SU(d,d)-irreducible representations, respectively specified by

$$[(\tfrac{1}{2} - \delta)^{d-1}, m + \tfrac{1}{2} - \delta, (\tfrac{1}{2} + \delta)^d]$$

and

$$[(\tfrac{1}{2} + \delta)^d, (\tfrac{1}{2} - \delta)^{d-1}, m + \tfrac{1}{2} - \delta],$$

where $\delta = m(2d)^{-1}$. This completes the proof of the complementarity between SO(2) and SU(d,d) when $d > 1$. When $d = 1$, SO(2) and SU(1,1) are not complementary. This is not surprising since SU(1,1) and Sp(2,R) have isomorphic Lie algebras.

In conclusion, we have shown that the Sp(4d,R) irreducible representation $\langle(\frac{1}{2})^{2d}\rangle$ ($\langle(\frac{1}{2})^{2d-1}\frac{3}{2}\rangle$) contains the class A and B irreducible representations corresponding to even (odd) m values just once. In addition, the irreducible representation $\langle(\frac{1}{2})^{2d}\rangle$ also contains the single class C irreducible representation. When restricting SU(d,d) to Sp(2d,R), all irreducible representations remain irreducible, but those belonging to class A and B and corresponding to equal m values are characterized by the same label $\langle 1^{d-1}, m+1\rangle$, while the class C irreducible representation is specified by $\langle 1^d\rangle$.

The first significant case of complementarity between SO(2) and SU(d,d) occurs for $d = 2$. The unitary irreducible representations of SU(2,2) were classified by Yao.[10] It is straightforward to show that those encountered in the present work correspond to exceptional degenerate discrete series or E^+ series in Yao's classification. The occurence of SU(2,2) in connection with SO(2) was already noted, although in an indirect way, by Moshinsky and his collaborators.[11] In their study of the microscopic collective model in two dimensions, they showed that boson states invariant under SO(2) can be classified according to some SO(4,2) group. Since the latter is locally isomorphic to SU(2,2), this result follows from the complementarity between SO(2) and SU(2,2). For arbitrary d values, the SU(d,d) group was also considered by Perelomov as a group of canonical transformations in the problem of boson pair creation in an alternating external field.[12] However, its relation with SO(2) was not noted in Ref. 12.

VI. GENERALIZATION TO SO(n) FOR $2 < n = 2q \leqslant 2d$

In the present section, we are going to generalize to higher n values the procedure used in Sec. IV to derive the Lie algebra of the complementary group with respect to SO(2). This will enable us to show that no such group exists within Sp(2dn,R) for $2 < n = 2q \leqslant 2d$.

If a complementary group with respect to SO(n) does exist, its Lie algebra is made of invariants under rotations. The latter separates into scalars and pseudoscalars. From the complementarity between O(2) and Sp(2d,R), we know that the simplest scalars which can be built from boson creation and annihilation operators are the Sp(2d,R) generators. All the other scalars are then polynomials in these generators, i.e., belong to the enveloping algebra of Sp(2d,R).

In the same way, let us construct the simplest pseudoscalars which can be formed from the η_{is} and ξ_{is} operators. They are the $n \times n$ determinants[8]

$$\Pi^{(p)}_{i_1\cdots i_n} = i^{n/2} \sum_{s_1\cdots s_n} \epsilon_{s_1\cdots s_n} \eta_{i_1 s_1}\cdots\eta_{i_p s_p}\xi_{i_{p+1}s_{p+1}}\cdots\xi_{i_n s_n},$$
$$p = 0,1,...,n, \quad i_1,...,i_n = 1,...,d, \tag{6.1}$$

where $\epsilon_{s_1\cdots s_n}$ is the antisymmetric tensor. The operators $\Pi^{(p)}_{i_1\cdots i_n}$ satisfy the following relations:

TABLE I. Lowest weight function $F(z_{i\sigma})$ and U(d, d), SU(d, d), SO(2) labels of class A, B, and C irreducible representations for the $n = 2$ case. The number m may take any positive integer value and δ is defined by $\delta = m(2d)^{-1}$.

Class	$F(z_{i\sigma})$	U(d, d)	SU(d, d)	SO(2)
A	z_{d+}^m	$[(\frac{1}{2})^{d-1}, m + \frac{1}{2}, (\frac{1}{2})^d]$	$[(\frac{1}{2} - \delta)^{d-1}, m + \frac{1}{2} - \delta, (\frac{1}{2} + \delta)^d]$	$[m]$
B	z_{d-}^m	$[(\frac{1}{2})^{2d-1}, m + \frac{1}{2}]$	$[(\frac{1}{2} + \delta)^d, (\frac{1}{2} - \delta)^{d-1}, m + \frac{1}{2} - \delta]$	$[-m]$
C	1	$[(\frac{1}{2})^{2d}]$	$[(\frac{1}{2})^{2d}]$	[0]

$$\Pi^{(p)}_{i_1\cdots i_r\cdots i_t\cdots i_n} = -\Pi^{(p)}_{i_1\cdots i_t\cdots i_r\cdots i_n} \quad \text{if } r,t \leqslant p \text{ or } r,t > p, \tag{6.2}$$

and

$$\Pi^{(p)}_{i_1\cdots i_n} = (-1)^{n/2+p}\left[\Pi^{(n-p)}_{i_{p+1}\cdots i_n i_1\cdots i_p}\right]^{\dagger}. \tag{6.3}$$

For $n = 2$ and $p = 2, 1, 0$, they reduce respectively to the operators $\Delta^{\dagger}_{i_1 i_2}$, $\Gamma_{i_1 i_2}$, and $-\Delta_{i_1 i_2}$, defined in Eq. (4.11). Starting from the basic pseudoscalars $\Pi^{(p)}_{i_1\cdots i_n}$, all the remaining ones can then be obtained by multiplication with arbitrary polynomials in the Sp($2d,R$) generators.

We note that, except in the $n = 2$ case, no pseudoscalar can be found in the Lie algebra of Sp($2dn,R$). The only rotational invariants belonging to the latter are therefore the Sp($2d,R$) generators. We have thus proved that no complementary group with respect to SO(n) exists within Sp($4d,R$) when $2 < n = 2q \leqslant 2d$.

VII. CONCLUSION

In the present paper, we have looked for a complementary group with respect to SO(n) within either irreducible representation $\langle(\frac{1}{2})^{nd}\rangle$ or $\langle(\frac{1}{2})^{nd-1}\frac{3}{2}\rangle$ of Sp($2dn,R$). We have proved that:

(i) this group is Sp($2d,R$) when $n = 2q + 1$ or $n = 2q > 2d$;

(ii) it is SU(d,d) when $n = 2$ and $d > 1$;

(iii) it does not exist when $n = 2$, $d = 1$, or $2 < n = 2q \leqslant 2d$.

Eventually, when $2 < n = 2q \leqslant 2d$, we might try to extend the present work by looking for an algebra $\mathscr{G}$ containing both the Sp($2d,R$) generators and the basic pseudoscalars $\Pi^{(p)}_{i_1\cdots i_n}$, defined in Eq. (6.1). Unfortunately, it can be shown that:

(i) For $2 < n = 2q < 2d$, $\mathscr{G}$ does not close since the commutator of two basic pseudoscalars belongs to the enveloping algebra of Sp($2d,R$);

(ii) for $n = 2d > 2$, $\mathscr{G}$ does close since there is only one basic pseudoscalar $\Pi^{(d)}_{1\cdots d\,1\cdots d}$, but no larger algebra containing both $\mathscr{G}$ and the SO(2) algebra is known to exist. Pending the finding of such a larger algebra, further investigation along this very line looks useless.

[1] M. Moshinsky and C. Quesne, J. Math. Phys. **11**, 1631 (1970).

[2] M. Moshinsky and C. Quesne, J. Math. Phys. **12**, 1772 (1971).

[3] M. Moshinsky, J. Math. Phys. **4**, 1128 (1963).

[4] E. Chacón, Ph.D. thesis, UNAM, Mexico, 1969.

[5] C. Quesne, J. Math. Phys. **14**, 366 (1973).

[6] J. Deenen and C. Quesne, J. Math. Phys. **23**, 878 (1982).

[7] D. E. Littlewood, Proc. London Math. Soc. **50**, 349 (1948).

[8] H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J., 1946), p. 52.

[9] V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).

[10] T. Yao, J. Math. Phys. **8**, 1931 (1967); **9**, 1615 (1968); **13**, 315 (1971).

[11] E. Chacón, M. Moshinsky, and V. Vanagas, J. Math. Phys. **22**, 605 (1981); M. Moshinsky and T. H. Seligman, J. Math. Phys. **22**, 1528 (1981);E. Chacón and M. Moshinsky, Kinam **3**, 3 (1981).

[12] A. M. Perelomov, Phys. Lett. A **39**, 165 (1972); Teor. Mat. Fiz. **16**, 303 (1973) [Theor. Math. Phys. **16**, 852 (1973)].

Some special SU(3)$\supset$R(3) Wigner coefficients and their application[a)]

K. T. Hecht and Y. Suzuki[b)]
Physics Department, University of Michigan, Ann Arbor, Michigan 48109

(Received 20 July 1982; accepted for publication 22 October 1982)

Bargmann space expansions of oscillator functions are used to derive analytic expressions for SU(3)$\supset$R(3) Wigner coefficients for the couplings $(\lambda_1 0)\times(0\mu_2)\to(\lambda_3\mu_3)L_3 = 0$ and $(\lambda_1 0)\times(\lambda_2 0)\to(\lambda_3\mu_3)L_3 = 0$, with arbitrary $(\lambda_3\mu_3)$. These lead to expansions useful in nuclear cluster problems and are used to give a simple form for the SU(3)-irreducible tensor expansion of a scalar two-body interaction, an application which motivated this investigation.

PACS numbers: 02.20.Rt, 21.60.Fw

I. INTRODUCTION

The widespread usefulness of the group SU(3) has led to many applications of SU(3) Wigner and recoupling coefficients, and efficient computer codes for their calculation are available.[1,2] In many applications analytic expressions for certain special coefficients are very useful. In the nuclear physics applications reduced Wigner coefficients in the SU(3)$\supset$R(3) basis are needed. In this basis the natural subgroup labels *LM* are in general insufficient to label the states, and the resultant inner multiplicity leads to complicated analytic expressions. Despite this difficulty, algebraic expressions have been tabulated by Vergados[3] for many SU(3)$\supset$R(3) Wigner coefficients useful in nuclear shell model applications in an orthonormal basis which is closely tied to the physically relevant Elliott *KLM* labeling scheme.[4] In addition, many algebraic expressions have been given by Sharp, von Baeyer, and Pieper[5,6] for the reduced Wigner coefficients in the SU(3)$\supset$R(3) basis involving (1) representations free of inner multiplicity or (2) special states which are labeled completely by $(\lambda\mu)$, L, and M. These include (1) the couplings involving only representations $(\lambda\ 0)$, (0μ), $(\lambda\ 1)$, or (1μ) and (2) the SU(3)$\supset$R(3) coefficients for the "stretched" coupling $(\lambda\ 0)\times(0\mu)\to(\lambda\mu)$ and special states of $(\lambda\mu)$ such as the states $L = 0$ (λ and μ both even), or $L = 1$ (λ,μ = even/odd). In recent applications to problems in nuclear collective motion exploiting Sp(3,*R*) symmetry,[7–9] and in applications to nuclear cluster problems,[10–12] it has proved useful to expand the rotationally invariant nucleon–nucleon interaction in terms of SU(3)-irreducible tensor components. For this purpose an algebraic expression is needed for the SU(3)$\supset$R(3) Wigner coefficients for the coupling $(\lambda_1 0)\times(0\mu_2)$ to states $(\lambda_3\mu_3)L_3 = 0$ of arbitrary λ_3 and μ_3 (λ_3, μ_3 both even, but $\lambda_3\leqslant\lambda_1$, $\mu_3\leqslant\mu_2$). Similarly, the SU(3)$\supset$R(3) Wigner coefficients for the coupling $(\lambda_1 0)\times(\lambda_2 0)$ to states $(\lambda_3\mu_3)L_3 = 0$ with $\lambda_3\leqslant\lambda_1+\lambda_2$, $\mu_3\geqslant 0$ have useful applications.

It is the purpose of this note to exhibit analytic expressions for these SU(3)$\supset$R(3) Wigner coefficients. For ready reference the results are given in Tables I–III. The method of calculation is presented in Sec. II. It makes use of some of the methods of Sharp *et al.*[5,6] but is based on an expansion of the SU(3)-states in terms of Bargmann space polynomials.[13–15] The decomposition of an effective two-body interaction into SU(3)-irreducible tensor components can also be achieved most efficiently through the Bargmann transform of this operator. The process is illustrated in detail in Sec. III for a scalar interaction of Gaussian radial form as an illustration of the usefulness of the results of Sec. II.

II. METHOD OF CALCULATION

The notation and phase conventions will adhere to those of Ref. 1. (The latter are based on the canonical definitions of Biedenharn *et al.*[16]) For ready reference the results for SU(3)$\supset$R(3) reduced Wigner coefficients are collected in tabular form (Tables I-III). The method of calculation makes use of the specific construction of the state $(\lambda\mu)$ with $L = 0$ by techniques similar to those used by Sharp *et al.*[5,6] However, it has proved useful to give all expansions in terms of Bargmann space variables.[13–15] With the one-dimensional real space variable x, we associate the complex Bargmann space variable K_x. With the three-dimensional real space variable $\mathbf{r}$ we associate the three-dimensional Bargmann space variable $\mathbf{K}$. Transformations from real-space square integrable functions $\phi(x)$ to the analytic Bargmann space functions $f(K_x)$ are effected by the transform

$$f(K_x) = \int dx\, A(K_x,x)\phi(x), \tag{1}$$

where

$$A(K_x,x) = \pi^{-1/4}\exp\left[-\tfrac{1}{2}K_x^2 - \tfrac{1}{2}x^2 + \sqrt{2}K_x x\right]. \tag{2}$$

The kernel $A(K_x,x)$ is a generating function of harmonic oscillator functions

$$A(K_x,x) = \sum_{n=0}^{\infty}\phi_n(x)^*(K_x^n/\sqrt{n!}), \tag{3}$$

that is, $K_x^n/\sqrt{n!}$ is the Bargmann transform of a normalized one-dimensional harmonic oscillator function. For single three-dimensional variables $\mathbf{r}$ and $\mathbf{K}$

$$A(\mathbf{K},\mathbf{r}) = \prod_{i=x,y,z} A(K_i,x_i) = \sum_{QLM}\phi^{(Q0)}_{LM}(\mathbf{r})^* P^{(Q0)}_{LM}(\mathbf{K}). \tag{4}$$

The three-dimensional oscillator functions have been expressed in terms of SU(3) representation labels $(\lambda\mu) = (Q\,0)$, Q = total number of oscillator quanta. The ex-

---

[a)] Supported by the U. S. National Science Foundation.

[b)] Nishina Memorial Foundation Fellow, on leave of absence from Physics Department, Niigata University, Niigata 950-21, Japan.

0022-2488/83/040785-08$02.50

TABLE I. The coefficients $\langle(2n0)L;(02m)L\,\|(2n-2\nu,2m-2\nu)0\rangle$.

$$(-1)^{\min(m-\nu,n-\nu)}\left[\frac{(2n+2m-4\nu+2)(2L+1)}{F(2n,L)F(2m,L)(2\nu)!(2n+2m+2-2\nu)!}\right]^{1/2}$$

$$\times\sum_{l=0}^{\min(m-\nu,n-\nu)}\frac{(-1)^l(n+m-2\nu-l)!(2\nu+2l)!F(2\nu+2l,L)}{l!(n-\nu-l)!(m-\nu-l)!};$$

$$(-1)^{\min(m-\nu,n-\nu)}\left[\frac{(2n+2m-4\nu+2)(2L+1)(2\nu)!}{F(2n,L)F(2m,L)(2n+2m+2-2\nu)!}\right]^{1/2}\frac{F(2\nu,L)(n+m-2\nu)!}{(n-\nu)!(m-\nu)!}$$

$$\times{}_4F_3(\nu+1,\nu+\tfrac{1}{2},-(n-\nu),-(m-\nu);(\nu-\tfrac{1}{2}L+1),(\nu+\tfrac{1}{2}L+\tfrac{3}{2}),-(n+m-2\nu);1);$$

$$(-1)^{\min(m-\nu,n-\nu)}[(2n+2m-4\nu+2)(2L+1)F(2n,L)F(2m,L)(2\nu)!(2n+2m+2-2\nu)!]^{1/2}$$

$$\times\sum_{\alpha=\max(0,L/2-\nu)}^{[L/2]}\frac{(-1)^\alpha(2L-2\alpha)!(n-\tfrac{1}{2}L+\alpha)!(m-\tfrac{1}{2}L+\alpha)!}{\alpha!(L-\alpha)!(L-2\alpha)!(\nu-\tfrac{1}{2}L+\alpha)!(n+m-\nu-\tfrac{1}{2}L+\alpha)!(2n+2m+2-2\nu-L+2\alpha)};$$

with

$$\langle(2n0)L;(02m)L\,\|(2n-2\nu,2m-2\nu)0\rangle=f(n,m,\nu,L)$$

$$\langle(2n+1,0)L';(0,2m+1)L'\|(2n-2\nu,2m-2\nu)0\rangle=f(n+\tfrac{1}{2},m+\tfrac{1}{2},\nu+\tfrac{1}{2},L')\times(-1)^{L'}$$

$$F(a,L)\equiv(\tfrac{1}{2}a+\tfrac{1}{2}L)!/(\tfrac{1}{2}a-\tfrac{1}{2}L)!(a+L+1)!$$

pansion could have been given in terms of any convenient set of subgroup labels. For the totally symmetric representation $(Q\,0)$ the angular momentum labels LM give a complete labeling. In this SU(3)⊃R(3) basis the Bargmann transform of a normalized harmonic oscillator function of a single three-dimensional variable is given in terms of solid spherical harmonics, $\mathscr{Y}_{LM}(\mathbf{K})$, by

$$P^{(Q0)}_{LM}(\mathbf{K})=[4\pi 2^L F(Q,L)]^{1/2}(\mathbf{K}\cdot\mathbf{K})^{(Q-L)/2}\mathscr{Y}_{LM}(\mathbf{K}),\quad\text{(5a)}$$

where

$$F(Q,L)=\frac{[\tfrac{1}{2}(Q+L)]!}{[\tfrac{1}{2}(Q-L)]!(Q+L+1)!}\qquad\text{(5b)}$$

(see, e.g., Ref. 15, but note that the phase has been adjusted to be in agreement with Ref. 1). The solid harmonics $\mathscr{Y}_{1M}(\mathbf{K})$, e.g., are given in terms of the complex Bargmann space variables K_x,K_y,K_z by $-1/\sqrt{2}(K_x+iK_y),K_z,1/\sqrt{2}(K_x-iK_y)$ and have SU(3) transformation properties identical with those of three-dimensional SU(3) basis vectors $|(1\,0)L=1M\rangle$. Note also that the Bargmann space functions $(P_{LM}{}^{(Q0)}(\mathbf{K}))^*$ have SU(3) transformation properties identical with those of basis vectors $(-1)^M|(0Q)L-M\rangle$. The SU(3)⊃R(3) Wigner coefficients for the coupling $(\lambda_1 0)\times(0\mu_2)$ to states $(\lambda\mu)L=0$ are obtained from a construction of the Bargmann space functions $P^{(\lambda\mu)}_{L=0}(\overline{\mathbf{K}},\mathbf{K}^*)$ in terms of two independent three-dimensional Bargmann space variables $\overline{\mathbf{K}}$ and $\mathbf{K}$. Those for the coupling of $(\lambda_1 0)\times(\lambda_2 0)$ to states $(\lambda\mu)L=0$ are obtained from a construction $P^{(\lambda\mu)}_{L=0}(\mathbf{K}_1,\mathbf{K}_2)$ in terms of the two independent three-dimensional Bargmann space variables $\mathbf{K}_1$ and $\mathbf{K}_2$.

The first construction will be illustrated by the special case λ_1,μ_1 both even, with $(\lambda_1 0)=(2n0)$, $(0\mu_2)=(02m)$, and with $(\lambda\mu)=(2n-2\nu,2m-2\nu)$. Without loss of generality we shall assume $m\leqslant n$. (The construction for states with $m>n$ is trivially similar. Wigner coefficients for the case $m>n$ can also be obtained from those with $m\leqslant n$ by simple symmetry properties.[1]) The state with $(\lambda\mu)=(2n-2\nu,2m-2\nu)$ and with $L=0$ is constructed in terms of the expansion

$$\begin{aligned}P^{(\lambda,\mu)}_{L=0}(\overline{\mathbf{K}},\mathbf{K}^*)&=[P^{(2n,0)}(\overline{\mathbf{K}})\times P^{(0,2m)}(\mathbf{K}^*)]^{(\lambda\mu)=(2n-2\nu,2m-2\nu)}_{L=0}\\&=\sum_{k=0}^{m-\nu}c_k(\overline{\mathbf{K}}\cdot\overline{\mathbf{K}})^{n-m+k}(\mathbf{K}^*\cdot\mathbf{K}^*)^k(\overline{\mathbf{K}}\cdot\mathbf{K}^*)^{2m-2k}.\end{aligned}\qquad(6)$$

The square bracket denotes SU(3) coupling. From the fact

TABLE IIA. The coefficients $\langle 2n0)L;(2m0)L\,\|(2n+2m-4\nu,2\nu)0\rangle$.

$$\frac{(-1)^{L/2+\min(\nu,n+m-2\nu)}L!}{2^L[(\tfrac{1}{2}L)!]^2}\left[\frac{(2n+2m-4\nu+1)(2L+1)}{F(2n,L)F(2m,L)(2n-2\nu)!(2m-2\nu)!}\right]^{1/2}$$

$$\times\frac{(n-\nu)!(m-\nu)!(n+m-\nu)!}{\nu!(2n+2m+1-2\nu)!}\sum_{k=\nu}^{\min(n,m)}\frac{(-1)^k2^{4k-2\nu}(2n+2m-2\nu-2k)!(k!)^2F(2k,L)}{(k-\nu)!(n-k)!(m-k)!(n+m-\nu-k)!};$$

$$\frac{(-1)^{L/2+\nu+\min(\nu,n+m-2\nu)}L!}{2^L[(\tfrac{1}{2}L)!]^2}\left[\frac{(2n+2m-4\nu+1)(2L+1)}{F(2n,L)F(2m,L)(2n-2\nu)!(2m-2\nu)!}\right]^{1/2}$$

$$\times\frac{2^{2\nu+1}\nu!(n+m-\nu)!(2n+2m-4\nu)!(\nu+\tfrac{1}{2}L+1)!}{(n+m-2\nu)!(2n+2m+1-2\nu)!(\nu-\tfrac{1}{2}L)!(2\nu+L+2)!}$$

$$\times{}_4F_3(\nu+1,\nu+1,-(n-\nu),-(m-\nu);(\nu-\tfrac{1}{2}L+1),(\nu+\tfrac{1}{2}L+\tfrac{3}{2},-(n+m-2\nu-\tfrac{1}{2});1);$$

$$(-1)^{L/2+\nu+\min(\nu,n+m-2\nu)}[(2n+2m-4\nu+1)(2L+1)F(2n,L)F(2m,L)(2n-2\nu)!(2m-2\nu)!]^{1/2}$$

$$\times\sum_{\alpha=\max(0,L/2-\nu)}^{L/2}\frac{(-1)^\alpha2^{2\nu+4\alpha-L}(2L-2\alpha)![(\tfrac{1}{2}L)!]^2}{\alpha!L!(L-\alpha)![(\tfrac{1}{2}L-\alpha)!]^2}$$

$$\times\frac{(2n+2m-2\nu+1)!(n-\tfrac{1}{2}L+\alpha)!(m-\tfrac{1}{2}L+\alpha)!(n+m-\nu-\tfrac{1}{2}L+\alpha)!\nu!}{(2n+2m-2\nu-L+2\alpha+1)!(n-\nu)!(m-\nu)!(n+m-\nu)!(\nu-\tfrac{1}{2}L+\alpha)!}$$

TABLE IIB. The coefficients $\langle(2n+1,0)L';(2m+1,0)L'\|(2n+2m+2-4\nu,2\nu)0\rangle$.

$$\frac{(-1)^{(L'+1)/2+\min(\nu,n+m+1-2\nu)}L'!}{2^{L'-1}[(\frac{1}{2}(L'-1))!]^2}\left[\frac{(2n+2m-4\nu+3)(2L'+1)}{F(2n+1,L')F(2m+1,L')(2n+1-2\nu)!(2m+1-2\nu)!}\right]^{1/2}$$

$$\times\frac{(n-\nu)!(m-\nu)!(n+m+1-\nu)!}{\nu!(2n+2m+3-2\nu)!}\times\sum_{k=\nu}^{\min(n,m)}\frac{(-1)^k2^{4k-2\nu}(2n+2m+2-2\nu-2k)!(k!)^2F(2k+1,L')}{(k-\nu)!(n-k)!(m-k)!(n+m+1-\nu-k)!};$$

$$\frac{(-1)^{(L'+1)/2+\nu+\min(\nu,n+m+1-2\nu)}L'!}{2^{L'-1}[(\frac{1}{2}(L'-1))!]^2}\left[\frac{(2n+2m-4\nu+3)(2L'+1)}{F(2n+1,L')F(2m+1,L')(2n+1-2\nu)!(2m+1-2\nu)!}\right]^{1/2}$$

$$\times\frac{2^{2\nu+2}\nu!(2n+2m+1-4\nu)!(n+m+1-\nu)!(\nu+\frac{1}{2}L'+\frac{3}{2})!}{(2n+2m+3-2\nu)!(n+m-2\nu)!(\nu-\frac{1}{2}L'+\frac{1}{2})!(2\nu+L'+3)!}$$

$$\times{}_4F_3(\nu+1,\nu+1,-(n-\nu),-(m-\nu);(\nu-\tfrac{1}{2}L'+\tfrac{3}{2}),(\nu+\tfrac{1}{2}L'+2),-(n+m-2\nu+\tfrac{1}{2});1);$$

$$(-1)^{(L'+1)/2+\nu+\min(\nu,n+m+1-2\nu)}[(2n+2m+3-4\nu)(2L'+1)F(2n+1,L')F(2m+1,L')]^{1/2}$$

$$\times[(2n+1-2\nu)!(2m+1-2\nu)!]^{1/2}\sum_{\alpha=\max(0,\frac{1}{2}(L'-1)-\nu)}^{(L'-1)/2}\frac{2^{2\nu+1+4\alpha-L'}(-1)^\alpha(2L'-2\alpha)![(\frac{1}{2}(L'-1))!]^2}{\alpha!\,L'!(L'-\alpha)![(\frac{1}{2}(L'-1)-\alpha)!]^2}$$

$$\times\frac{(2n+2m+3-2\nu)!(n+\frac{1}{2}-\frac{1}{2}L'+\alpha)!(m+\frac{1}{2}-\frac{1}{2}L'+\alpha)!(n+m-\nu-\frac{1}{2}L'+\frac{3}{2}+\alpha)!\nu!}{(2n+2m+4-2\nu-L'+2\alpha)!(n-\nu)!(m-\nu)!(n+m+1-\nu)!(\nu-\frac{1}{2}L'+\frac{1}{2}+\alpha)!}$$

that the scalar product $(\bar{K}\cdot K^*)$ is also an SU(3)-scalar and from the relations

$$(\bar{\mathbf{K}}\cdot\bar{\mathbf{K}})^Q=[(2Q+1)!]^{1/2}P^{(Q0)}_{L=0}(\bar{\mathbf{K}}),$$
$$(\mathbf{K}^*\cdot\mathbf{K}^*)^Q=[(2Q+1)!]^{1/2}P^{(0Q)}_{L=0}(\mathbf{K}^*),\qquad(7)$$

we see that the k th term in the expansion (6) gives a linear combination of states with $(\lambda\mu)=(2n-2m,0)$, $(2n-2m+2,2),\ldots,(2n-2m+2k,2k)$. The coefficients c_k with $k<m-\nu$ are to be chosen to eliminate the unwanted representations with $\lambda<2n-2\nu$, $\mu<2m-2\nu$. Due to the simplicity of the Bargmann space functions this can be achieved by direct action of the SU(3) Casimir operator. In

TABLE III. Special cases.

$$\langle(2n0)0;(02m)0\|(2n-2\nu,2m-2\nu)0\rangle$$
$$=\left[\frac{(2n+2m+2-4\nu)(2n+2m+1-2\nu)!(2\nu)!}{(2n+2m+2-2\nu)(2n+1)!(2m+1)!}\right]^{1/2}\frac{(-1)^{\min(n-\nu,m-\nu)}n!m!}{(n+m-\nu)!\nu!}$$

$$\langle(2n0)L;(02m)L\|(2n,2m)0\rangle$$
$$=[(2L+1)F(2n,L)F(2m,L)(2n+2m+1)!]^{1/2}\frac{(-1)^{L/2+\min(n,m)}L!n!m!}{[(\frac{1}{2}L)!]^2(n+m)!}$$

$$\langle(2n+1,0)1;(0,2m+1)1\|(2n-2\nu,2m-2\nu)0\rangle$$
$$=[3F(2n+1,1)F(2m+1,1)(2n+2m-4\nu+2)(2\nu+1)!(2n+2m+3-2\nu)!]^{1/2}$$
$$\times(-1)^{1+\min(n-\nu,m-\nu)}\frac{n!m!}{\nu!(n+m+1-\nu)!}$$

$$\langle(2n+1,0)L';(0,2m+1)L'\|(2n,2m)0\rangle$$
$$=[(2L'+1)F(2n+1,L')F(2m+1,L')(2n+2m+3)(2n+2m+1)!]^{1/2}$$
$$\times(-1)^{(L'+1)/2+\min(n,m)}\frac{(L'+1)!n!m!}{[\frac{1}{2}(L'-1)]![\frac{1}{2}(L'+1)]!(n+m)!}$$

$$\langle(2n0)(;(2m0)0\|(2n+2m-4\nu,2\nu)0\rangle$$
$$=\left[\frac{(2n+2m+1-4\nu)(2n-2\nu)!(2m-2\nu)!}{(2n+1)!(2m+1)!}\right]^{1/2}\frac{(-1)^{\nu+\min(\nu,n+m-2\nu)}2^{2\nu}n!m!}{(n-\nu)!(m-\nu)!}$$

$$\langle(2n0)L;(2m0)L\|(2n-2m,2m)0\rangle$$
$$=\left[\frac{(2L+1)F(2m,L)(2n-2m+1)!}{F(2n,L)}\right]^{1/2}\frac{(-1)^{L/2+m+\min(m,n-m)}L!2^{2m-L}n!m!}{[(\frac{1}{2}L)!]^2(n-m)!(2n+1)!}$$

$$\langle(2n+1,0)1;(2m+1,0)1\|(2n+2m+2-4\nu,2\nu)0\rangle$$
$$=\left[\frac{3(2n+2m+3-4\nu)(n+1)(m+1)(2n+1-2\nu)!(2m+1-2\nu)!}{(2n+3)!(2m+3)!}\right]^{1/2}\frac{(-1)^{1+\nu+\min(\nu,n+m+1-2\nu)}2^{2\nu+1}n!m!}{(n-\nu)!(m-\nu)!}$$

$$\langle(2n+1,0)L';(2m+1,0)L'\|(2n-2m+2,2m)0\rangle$$
$$=\left[\frac{(2L'+1)(2n-2m+3)F(2m+1,L')(2n-2m+1)!}{F(2n+1,L')}\right]^{1/2}\frac{(-1)^{(L'+1)/2+m+\min(m,n-m+1)}2^{2\nu+2-L'}L'!(n+1)!m!}{[(\frac{1}{2}(L'-1))!]^2(2n+3)!(n-m)!}$$

$$\langle(2a,2m)0;(2b,0)0\|(2a+2b,2m)0\rangle$$
$$=\left[\frac{(2a)!(2a+2b+2m+1)!}{(2b+1)(2a+2b)!(2a+2m+1)!}\right]^{1/2}\frac{(a+m)!(a+b)!}{a!(a+b+m)!}(-1)^\sigma$$

$\sigma=0$ for $m<a$, $\sigma=b$ for $m>a$

the subspace spanned by the Bargmann space vectors $\overline{\mathbf{K}}$, $\mathbf{K}^*$, the U(3) generators can be expressed by

$$A_{ij} = \overline{K}_i \frac{\partial}{\partial \overline{K}_j} - K_j^* \frac{\partial}{\partial K_i^*}, \quad i,j = x,y,z, \tag{8}$$

leading to the SU(3) Casimir operator

$$\mathbf{C}_{\mathrm{SU(3)}} = \sum_{\alpha,\beta} A_{\alpha\beta} A_{\beta\alpha} - \tfrac{1}{3}(\mathrm{Tr}A)^2, \tag{9}$$

with eigenvalue

$$C_{\mathrm{SU(3)}} = \tfrac{2}{3}[\lambda^2 + \lambda\mu + \mu^2 + 3(\lambda + \mu)]. \tag{10}$$

The action of the operator $(\mathbf{C}_{\mathrm{SU(3)}} - C_{\mathrm{SU(3)}})$ on the relation (6) leads to the recursion relation

$$c_{k-1} = - \frac{k(n-m+k)}{(n-\nu+k)(m-\nu+1-k)} c_k. \tag{11}$$

The normalization is achieved by the coefficient $c_{m-\nu}$ [see the remarks following Eq. (17)],

$$c_{m-\nu} = (-1)^{m-\nu} \left[\frac{(2n+2m-4\nu+2)}{(2\nu)!(2n+2m-2\nu+2)!} \right]^{1/2} \times \frac{(n+m-2\nu)!}{(n-\nu)!)(m-\nu)!}, \tag{12}$$

leading to

$$\begin{aligned}
&[P^{(2n,0)}(\overline{\mathbf{K}}) \times P^{(0,2m)}(\mathbf{K}^*)]^{(2n-2\nu,2m-2\nu)}_{L=0} \\
&\quad = \left[\frac{(2n+2m-4\nu+2)}{(2\nu)!(2n+2m-2\nu+2)!} \right]^{1/2} \\
&\qquad \times \sum_{k=0}^{m-\nu} \frac{(-1)^k (n-\nu+k)!}{k!(n-m+k)!(m-\nu-k)!} \\
&\qquad \times (\overline{\mathbf{K}}\cdot\overline{\mathbf{K}})^{n-m+k} (\mathbf{K}^*\cdot\mathbf{K}^*)^k (\overline{\mathbf{K}}\cdot\mathbf{K}^*)^{2m-2k}.
\end{aligned} \tag{13}$$

The relation[10,15]

$$\begin{aligned}
\frac{1}{c!}(\overline{\mathbf{K}}\cdot\mathbf{K}^*)^c &= [\dim(c0)]^{1/2} [P^{(c0)}(\overline{\mathbf{K}}) \times P^{(0c)}(\mathbf{K}^*)]^{(00)}_0 \\
&= \sum_L (P^{(c0)}_L(\overline{\mathbf{K}}) \cdot P^{(0c)}_L(\mathbf{K}^*)),
\end{aligned} \tag{14}$$

together with a double application of Eq. (5), leads to

$$\begin{aligned}
&[P^{(2n,0)}(\overline{\mathbf{K}}) \times P^{(0,2m)}(\mathbf{K}^*)]^{(2n-2\nu,2m-2\nu)}_0 \\
&\quad = \sum_L \langle (2n0)L;(02m)L \,\|\, (2n-2\nu,2m-2\nu)0 \rangle (-1)^L \\
&\qquad \times \frac{(P^{(2n,0)}_L(\overline{\mathbf{K}}) \cdot P^{(0,2m)}_L(\mathbf{K}^*))}{[2L+1]^{1/2}} \\
&\quad = \sum_L \left[\frac{(2n+2m-4\nu+2)}{(2\nu)!(2n+2m-2\nu+2)!F(2n,L)F(2m,L)} \right]^{1/2} \\
&\qquad \times \sum_{k=0}^{m-\nu} \frac{(-1)^k (n-\nu+k)!(2m-2k)!F(2m-2k,L)}{k!(n-m+k)!(m-\nu-k)!} \\
&\qquad \times (P^{(2n,0)}_L(\overline{\mathbf{K}}) \cdot P^{(0,2m)}_L(\mathbf{K}^*)).
\end{aligned} \tag{15}$$

From this relation a first expression for the SU(3)⊃R(3) reduced Wigner coefficient[17] is obtained for the coupling $(2n\ 0) \times (0\ 2m) \rightarrow (\lambda\mu)\mathrm{L} = 0$. This is listed as the first entry in Table I. In Table I the summation index has been changed from k to $l = m - \nu - k$ to gain a more symmetrical form. The summation can be expressed in terms of a Saalschutzian generalized hypergeometric function of type ${}_4F_3$ and of argument unity (see the second entry of Table I). Since no simple closed form is known for such a function, it appears that the summation cannot be carried out in closed form. For the special case $L = 0$ the hypergeometric function collapses to a Saalschutzian function of type ${}_3F_2$ for which a closed form is known; see Eq. (III.2) or Eq. (2.3.1.4) of Ref. 18. (We are indebted to Professor A. C. T. Wu for pointing out this identity to us.) This leads to a very simple expression for the coefficient $\langle (2n\ 0)0; (0\ 2m)0 \| (\lambda\mu)0 \rangle$; see the first entry of Table III. The need for this coefficient formed the starting point for this investigation.

The normalization coefficient $c_{m-\nu}$, Eqs. (6), (11), and (12), was determined with the aid of the relation

$$\begin{aligned}
&(\overline{\mathbf{K}}\cdot\overline{\mathbf{K}})^a (\mathbf{K}^*\cdot\mathbf{K}^*)^b (\overline{\mathbf{K}}\cdot\mathbf{K}^*)^c \\
&\quad = \sum_\gamma \langle (2a0)0;(02b)0 \| (2a-2\gamma,2b-2\gamma)0 \rangle \\
&\qquad \times [(2a+1)!(2b+1)!\dim(c0)]^{1/2} c! \\
&\qquad \times [[P^{(2a,0)}(\overline{\mathbf{K}}) \times P^{(0,2b)}(\mathbf{K}^*)]^{(2a-2\gamma,2b-2\gamma)} \\
&\qquad \times [P^{(c0)}(\overline{\mathbf{K}}) \times P^{(0c)}(\mathbf{K}^*)]^{(00)}]^{(2a-2\gamma,2b-2\gamma)}_{L=0} \\
&\quad = \sum_\gamma \langle (2a0)0;(02b)0 \| (2a-2\gamma,2b-2\gamma)0 \rangle \\
&\qquad \times \left[\frac{(2a+1)!(2b+1)!(2\gamma+c)!(2a+2b+c-2\gamma+2)!}{(2\gamma)!(2a+2b-2\gamma+2)!} \right]^{1/2} \\
&\qquad \times [P^{(2a+c,0)}(\overline{\mathbf{K}}) \times P^{(0,2b+c)}(\mathbf{K}^*)]^{(2a-2\gamma,2b-2\gamma)}_{L=0}.
\end{aligned} \tag{16}$$

In the last step of Eq. (16) the SU(3)-coupled K-space functions have been subjected to an SU(3)-recoupling transformation, and the renormalization relation

$$\begin{aligned}
&[P^{(\alpha 0)}(\mathbf{K}) \times P^{(\beta 0)}(\mathbf{K})]^{(\lambda\mu)}_{LM} \\
&\quad = \delta_{(\lambda\mu)(\alpha+\beta,0)} [(\alpha+\beta)!/\alpha!\beta!]^{1/2} P^{(\alpha+\beta,0)}_{LM}(\mathbf{K})
\end{aligned} \tag{17}$$

has been used. For the recoupling transformation, see in particular Eqs. (A18) and (A14) of Ref. 11. [Appendixes A and B of Ref. 11 contain many useful formulae for SU(3)-coupled Bargmann space functions.] When Eq. (16), with $a = n - m - k$, $b = k$, $c = 2m - 2k$, is substituted into the relation (6), only the single term with $k = m - \nu$ and $\gamma = 0$ survives in the summations. This relates the normalization coefficient $c_{m-\nu}$ and the coefficient $\langle (2n-2\nu, 0)0; (0, 2m-2\nu)0 \| (2n-2\nu, 2m-2\nu)0 \rangle$ for the "stretched" coupling which can be evaluated with the further use of Eq. (15) for $L = 0$. This procedure does not determine the phase of $c_{m-\nu}$. We shall adhere to the phase conventions of Ref. 1 and choose the states $|(\lambda\mu)LM\rangle$ with $\lambda \geqslant \mu$ to have phases consistent with angular momentum projection from the intrinsic state G'_{LW} of Ref. 1, whereas those with $\mu > \lambda$ have phases consistent with angular momentum projection from the state G'_{HW}. In the notation of Ref. 1 this means that our states are of type $IJ = 0\ 1$ for $\lambda \geqslant \mu$ and of type $IJ = 1\ 0$ for $\mu > \lambda$. Note also that this insures that $P^{(Q0)}_{LM}(\mathbf{K})$ and $P^{(0Q)}_{L-M}(\mathbf{K}^*)$ are related by the simple conjugation relations spelled out in Ref. 1. With this choice of phases the states (6) are defined completely for both $m \leqslant n$, $m > n$. The sign of the coefficient $< (2n0)0; (02m)0 \| (\lambda\mu)0 >$ is given by $(-1)^\phi$ with

$\phi = \min(\frac{1}{2}\lambda, \frac{1}{2}\mu)$ (see Table III). With the knowledge of this coefficient, Eq. (16) can be put in the form

$$\begin{aligned}
&(\overline{\mathbf{K}}\cdot\overline{\mathbf{K}})^a(\mathbf{K}^*\cdot\mathbf{K}^*)^b(\overline{\mathbf{K}}\cdot\mathbf{K}^*)^c \\
&= \sum_{\gamma=0}^{\min(a,b)} \frac{(-1)^{\min(a-\gamma,b-\gamma)}}{2}\,[(2a+2b-4\gamma+2) \\
&\quad\times(2\gamma+c)!(2a+2b+c-2\gamma+2)!]^{1/2} \\
&\quad\times\frac{a!b!}{\gamma!(a+b+1-\gamma)!}\,[P^{(2a+c,0)}(\overline{\mathbf{K}}) \\
&\quad\times P^{(0,2b+c)}(\mathbf{K}^*)]^{(2a-2\gamma,2b-2\gamma)}_{L=0}.
\end{aligned}\tag{18}$$

This expression will prove useful in applications to nuclear cluster problems.[10–12]

The first entry of Table I is very convenient when $m-\nu$, or $n-\nu$, is a small integer since the number of terms in the l-sum is then small. For small values of ν, or for small values of L, an alternate form may prove more convenient. To obtain this form, we start with the expansion

$$\begin{aligned}
&(P^{(2n,0)}_L(\overline{\mathbf{K}})\cdot P^{(0,2m)}_L(\mathbf{K}^*)) \\
&= \sum_\nu \langle(2n0)L;\,(02m)L\,\|(2n-2\nu,2m-2\nu)0\rangle \\
&\quad\times(-1)^L[(2L+1)]^{1/2} \\
&\quad\times[P^{(2n,0)}(\overline{\mathbf{K}})\times P^{(0,2m)}(\mathbf{K}^*)]^{(2n-2\nu,2m-2\nu)}_0 \\
&= (2L+1)[F(2n,L)F(2m,L)]^{1/2} \\
&\quad\times\sum_{\alpha=0}^{L/2}\frac{(-1)^\alpha(2L-2\alpha)!}{\alpha!(L-\alpha)!(L-2\alpha)!} \\
&\quad\times(\overline{\mathbf{K}}\cdot\overline{\mathbf{K}})^{n-L/2+\alpha}(\mathbf{K}^*\cdot\mathbf{K}^*)^{m-L/2+\alpha}(\overline{\mathbf{K}}\cdot\mathbf{K}^*)^{L-2\alpha},
\end{aligned}\tag{19}$$

where Eq. (5) has been used together with the expansion of the Legendre polynomial $P_L(\xi)$ in powers of $\xi^{L-2\alpha}$. By substituting Eq. (18) into the right-hand side of Eq. (19) the alternate form for the SU(3)$\supset$R(3) Wigner coefficient is obtained. This is given as the third entry in Table I in a form which can be generalized to the coupling $(\lambda_1 0)\times(0\mu_2)$ with λ_1 and μ_2 both odd.

Similar techniques can be used to calculate SU(3)$\supset$R(3) Wigner coefficients for the coupling $(\lambda_1 0)\times(\lambda_2 0)\rightarrow(\lambda\mu)L=0$. The SU(3)-coupled Bargmann space functions are now constructed from two independent Bargmann-space variables $\mathbf{K}_1$ and $\mathbf{K}_2$. The cases λ_1, λ_2 both even and λ_1, λ_2 both odd are now slightly different and are treated separately. With $\lambda_1 = 2n$, $\lambda_2 = 2m$, $m \leqslant n$; the coupled Bargmann space function with $(\lambda\mu) = (2n+2m-4\nu, 2\nu)$ and $L=0$ is now constructed in terms of the expansion

$$\begin{aligned}
&[P^{(2n,0)}(\mathbf{K}_1)\times P^{(2m,0)}(\mathbf{K}_2)]^{(2n+2m-4\nu,2\nu)}_{L=0} \\
&= \sum_{k=\nu}^{m} c_k(\mathbf{K}_1\cdot\mathbf{K}_1)^{n-k}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m-k} \\
&\quad\times[(\mathbf{K}_1\cdot\mathbf{K}_1)(\mathbf{K}_2\cdot\mathbf{K}_2)-(\mathbf{K}_1\cdot\mathbf{K}_2)^2]^k.
\end{aligned}\tag{20}$$

The last factor involves the square of the vector product $[\mathbf{K}_1\times\mathbf{K}_2]$, an SU(3) (01)-tensor, so that it carries the single SU(3) representation $(0, 2k)$. The kth term in the expansion (20) is therefore a linear combination of states with $(\lambda\mu) = (2n+2m-4k, 2k), (2n+2m-4k-4, 2k+2), \ldots, (2n-2m, 2m)$. The coefficients c_k with $k>\nu$ are to be chosen to eliminate the unwanted representations with $\lambda < 2n+2m-4\nu$ and $\mu > 2\nu$. This is again achieved by the action of the operator $(\mathbf{C}_{\mathrm{SU}(3)} - C_{\mathrm{SU}(3)})$, see Eqs. (9) and (10), where the U(3) generators are now given by

$$A_{ij} = K_{1i}\frac{\partial}{\partial K_{1j}} + K_{2i}\frac{\partial}{\partial K_{2j}}\tag{21}$$

in the subspace spanned by the Bargmann space vectors $\mathbf{K}_1$, $\mathbf{K}_2$. This leads to the recursion formula

$$c_{k+1} = -\frac{2(n-k)(m-k)}{(k+1-\nu)(2n+2m-2\nu-1-2k)}\,c_k.\tag{22}$$

The normalization is achieved by the coefficient c_ν

$$\begin{aligned}
c_\nu &= (-1)^{\nu+\min(\nu,n+m-2\nu)}\left[\frac{(2n+2m-4\nu+1)}{(2n-2\nu)!(2m-2\nu)!}\right]^{1/2} \\
&\quad\times\frac{(2n+2m-4\nu)!(n+m-\nu)!}{(2n+2m-2\nu+1)!\nu!(n+m-2\nu)!},
\end{aligned}\tag{23}$$

which follows from the term with $k=\nu$ and $\gamma=0$ in the expansion

$$\begin{aligned}
&(\mathbf{K}_1\cdot\mathbf{K}_1)^{n-k}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m-k}([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^k \\
&= [(2n-2k+1)!(2m-2k+1)!]^{1/2}(2k+1)!(-i)^{2k} \\
&\quad\times\sum_{\substack{\nu \\ \gamma=(\nu-k)}}\langle(2n-2k,0)0; \\
&\quad\times(2m-2k,0)0\|(2n+2m-4k-4\gamma,2\gamma)0\rangle \\
&\quad\times\langle(2n+2m-4k-4\gamma,2\gamma,)0; \\
&\quad\times(0,2k)0\|(2n+2m-4\nu,2\nu)0\rangle \\
&\quad\times[[P^{(2n-2k,0)}(\mathbf{K}_1) \\
&\quad\times P^{(2m-2k,0)}(\mathbf{K}_2)]^{(2n+2m-4k-4\gamma,2\gamma)} \\
&\quad\times[P^{(2k,0)}(\mathbf{K}_1)\times P^{(2k,0)}(\mathbf{K}_2)]^{(0,2k)}]^{(2n+2m-4\nu,2\nu)}_{L=0} \\
&= [(2n-2k+1)!(2m-2k+1)!]^{1/2}(2k+1)!(-i)^{2k} \\
&\quad\times\sum_\nu\langle(2n-2k,0)0; \\
&\quad\times(2m-2k,0)0\|(2n+2m-4\nu,2\nu-2k)0\rangle \\
&\quad\times\langle(2n+2m-4\nu,2\nu-2k)0; \\
&\quad\times(0,2k)0\|(2n+2m-4\nu,2\nu)0\rangle\,\frac{1}{(2k)!} \\
&\quad\times\left[\frac{(2\nu)!(2n+2m-2\nu+1)!}{(2k+1)(2\nu-2k)!(2n+2m-2\nu-2k+1)!}\right]^{1/2} \\
&\quad\times[P^{(2n,0)}(\mathbf{K}_1)\times P^{(2m,0)}(\mathbf{K}_2)]^{(2n+2m-4\nu,2\nu)}_{L=0},
\end{aligned}\tag{24}$$

where an SU(3)-recoupling transformation was used, together with relation (17), in the last step of Eq. (24); [see Eq. (A21) of Ref. 11]. Note also that the nature of the K_1, K_2-space function requires that the final SU(3) representation must correspond to a two-rowed tableau. As a result, γ is restricted to the value $\gamma = \nu - k$. Equation (24) has used Eq. (7) and the related equation

$$\begin{aligned}
&P^{(0,2k)}_{L=0}([\mathbf{K}_1\times\mathbf{K}_2]) \\
&= [(2k+1)!]^{1/2}(-i)^{2k}[P^{(2k,0)}(\mathbf{K}_1)\times P^{(2k,0)}(\mathbf{K}_2)]^{(0,2k)}_{L=0}.
\end{aligned}\tag{25}$$

(This relation has been given in Ref. 11; see Eqs. (B4) and

(B5); but a correction by the phase factor $(-i)^{2k}$ is needed.] In the general case, Eq. (24) can be evaluated once the coefficient $\langle(2a,0)0;(2b,0)0\|(\lambda\mu)0\rangle$ is known. The second coefficient, which is of type $\langle(2a,2b)0;(02c)0\|(2a,2b+2c)0\rangle$, is related by symmetry to $\langle(2b,2a)0;(2c,0)0\|(2b+2c,2a)0\rangle$. For $a \leqslant b$ this can be read from Eq. (13), together with Eq. (17). For $a > b$, the phase of this coefficient is $(-1)^c$; see the last entry of Table III. For the evaluation of c_ν, only the term with $k=\nu$ is needed, and in this special case Eq. (24) can be evaluated through coefficients known from our earlier discussion. This leads to

$$
\begin{aligned}
&[P^{(2n,0)}(\mathbf{K}_1)\times P^{(2m,0)}(\mathbf{K}_2)]^{(2n+2m-4\nu,2\nu)}_{L=0} \\
&\quad=(-1)^{\nu+\min(\nu,n+m-2\nu)}\left[\frac{2n+2m-4\nu+1}{(2n-2\nu)!(2m-2\nu)!}\right]^{1/2} \\
&\qquad\times\frac{(n-\nu)!(m-\nu)!(n+m-\nu)!}{\nu!(2n+2m-2\nu+1)!} \\
&\qquad\times\sum_{k=\nu}^{\min(n,m)}\frac{(-1)^{k-\nu}2^{2k-2\nu}(2n+2m-2\nu-2k)!}{(k-\nu)!(n-k)!(m-k)!(n+m-\nu-k)!} \\
&\qquad\times(\mathbf{K}_1\cdot\mathbf{K}_1)^{n-k}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m-k}([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^k.
\end{aligned}
\tag{26}
$$

We now use Eq. (25) with the simple coefficient[6] $\langle(2k\,0)0;(2k\,0)0\|(02k)0\rangle$ (with phases chosen according to Ref. 1), to obtain

$$
\begin{aligned}
&([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^k \\
&\quad=\sum_L (k!)^2\frac{(-1)^{L/2}2^{2k-L}L!}{[(\frac{1}{2}L)!]^2} \\
&\qquad\times(-1)^L(P^{(2k,0)}_L(\mathbf{K}_1)\cdot P^{(2k,0)}_L(\mathbf{K}_2)).
\end{aligned}
\tag{27}
$$

With this relation, Eq. (26) together with a double application of Eq. (5) leads to the desired SU(3)⊃R(3) Wigner coefficient. The result is given as the first entry of Table II. The single sum can again be expressed in terms of a generalized hypergeometric function of type ${}_4F_3$; see the second entry of Table II. For $L=0$ this collapses to a ${}_3F_2$ of Saalschutzian form and leads to the simple special case shown in Table III. With this result, Eq. (24) yields

$$
\begin{aligned}
&(\mathbf{K}_1\cdot\mathbf{K}_1)^{n-k}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m-k}([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^k \\
&\quad=\sum_\nu(-1)^{\nu+\min(\nu,n+m-2\nu)} \\
&\qquad\times[(2n+2m-4\nu+1)(2n-2\nu)!(2m-2\nu)!]^{1/2} \\
&\qquad\times\frac{2^{2\nu-2k}(n-k)!(m-k)!\nu!(n+m-\nu-k)!(2n+2m-2\nu+1)!}{(n-\nu)!(m-\nu)!(\nu-k)!(n+m-\nu)!(2n+2m-2\nu-2k+1)!} \\
&\qquad\times[P^{(2n,0)}(\mathbf{K}_1)\times P^{(2m,0)}(\mathbf{K}_2)]^{(2n+2m-4\nu,2\nu)}_{L=0}.
\end{aligned}
\tag{28}
$$

This result by itself may have useful applications in nuclear cluster problems.[10–12] It can also be used to derive an alternate form for the SU(3) Wigner coefficient more useful when L or ν (rather than $m-\nu$ or $n-\nu$) are small integers. The analog of Eq. (19) becomes

$$
\begin{aligned}
&(P^{(2n,0)}_L(\mathbf{K}_1)\cdot P^{(2m,0)}_L(\mathbf{K}_2)) \\
&\quad=(2L+1)[F(2n,L)F(2m,L)]^{1/2} \\
&\qquad\times\sum_{\alpha=0}^{L/2}\frac{(-1)^{L/2+\alpha}2^{2\alpha}(2L-2\alpha)![(\frac{1}{2}L)!]^2}{\alpha!L!(L-\alpha)![(\frac{1}{2}L-\alpha)!]^2} \\
&\qquad\times(\mathbf{K}_1\cdot\mathbf{K}_1)^{n-L/2+\alpha}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m-L/2+\alpha} \\
&\qquad\times([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^{L/2-\alpha}.
\end{aligned}
\tag{29}
$$

If this is combined with Eq. (28) the alternate form of the Wigner coefficient is obtained; see the third entry of Table II.

The SU(3)⊃R(3) Wigner coefficients for the coupling $(\lambda_1 0)\times(\lambda_2 0)\rightarrow(\lambda\mu)L=0$ with λ_1 and λ_2 both odd have a slightly different form. The key intermediate results are

$$
\begin{aligned}
&[P^{(2n+1,0)}(\mathbf{K}_1)\times P^{(2m+1,0)}(\mathbf{K}_2)]^{(2n+2m+2-4\nu,2\nu)}_{L=0} \\
&\quad=(-1)^{\min(\nu,n+m+1-2\nu)} \\
&\qquad\times\left[\frac{(2n+2m-4\nu+3)}{(2n+1-2\nu)!(2m+1-2\nu)!}\right]^{1/2} \\
&\qquad\times\frac{(n-\nu)!(m-\nu)!(n+m+1-\nu)!}{\nu!(2n+2m-2\nu+3)!}\sum_{k=\nu}^{\min(n,m)} \\
&\qquad\times\frac{(-1)^k2^{2k-2\nu}(2n+2m+2-2\nu-2k)!}{(k-\nu)!(n-k)!(m-k)!(n+m+1-\nu-k)!} \\
&\qquad\times(\mathbf{K}_1\cdot\mathbf{K}_1)^{n-k}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m-k}(\mathbf{K}_1\cdot\mathbf{K}_2) \\
&\qquad\times[\mathbf{K}_1\cdot\mathbf{K}_1)(\mathbf{K}_2\cdot\mathbf{K}_2)-(\mathbf{K}_1\cdot\mathbf{K}_2)^2]^k.
\end{aligned}
\tag{30}
$$

Combining Eq. (27) with $(\mathbf{K}_1\cdot\mathbf{K}_2) = (P^{(1\,0)}_{L=1}(\mathbf{K}_1)\cdot P^{(1\,0)}_{L=1}(\mathbf{K}_2))$, applying Eq. (5) and standard spherical harmonic addition theorems, followed by a further application of Eq. (5), we obtain the first form of the coefficient $\langle(2n+1,0)L';(2m+1,0)L'\|(\lambda\mu)0\rangle$ tabulated in Table II. This form can again be expressed in terms of a Saalschutzian hypergeometric function of type ${}_4F_3$; see the entry in Table II. In this case the ${}_4F_3$ collapses to a ${}_3F_2$ for the special case $L'=1$, so that the coefficient with $L'=1$ can be given in closed form. This is included among the special cases of Table III. An alternate form for this coefficient is obtained from the analog of Eq. (29), which is now

$$
\begin{aligned}
&(P^{(2n+1,0)}_{L'}(\mathbf{K}_1)\cdot P^{(2m+1,0)}_{L'}(\mathbf{K}_2))\\
&\quad=(2L'+1)[F(2n+1,L')F(2m+1,L')]^{1/2}\sum_{\alpha=0}^{(L'-1)/2}\\
&\quad\times\frac{(-1)^{(L'-1)/2+\alpha}2^{2\alpha}(2L'-2\alpha)!\{[\frac{1}{2}(L'-1)]!\}^2}{\alpha!L'!(L'-\alpha)!\{[\frac{1}{2}(L'-1)-\alpha]!\}^2}\\
&\quad\times(\mathbf{K}_1\cdot\mathbf{K}_1)^{n+1/2-L'/2+\alpha}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m+1/2-L'/2+\alpha}\\
&\quad\times(\mathbf{K}_1\cdot\mathbf{K}_2)([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^{(L'-1)/2-\alpha}.
\end{aligned}
\tag{31}
$$

Simple Wigner coefficients from Table III, together with a recoupling transformation, now give

$$
\begin{aligned}
&(\mathbf{K}_1\cdot\mathbf{K}_1)^{n+1/2-L'/2+\alpha}(\mathbf{K}_2\cdot\mathbf{K}_2)^{m+1/2-L'/2+\alpha}(\mathbf{K}_1\cdot\mathbf{K}_2)([\mathbf{K}_1\times\mathbf{K}_2]\cdot[\mathbf{K}_1\times\mathbf{K}_2])^{(L'-1)/2-\alpha}\\
&\quad=\sum_{\nu}(-1)^{\nu+\min(\nu,n+m+1-2\nu)}[(2n+2m+3-4\nu)(2n+1-2\nu)!(2m+1-2\nu)!]^{1/2}\\
&\quad\times 2^{2\nu+2\alpha-L'+1}\frac{(2n+2m+3-2\nu)!(n+\frac{1}{2}-\frac{1}{2}L'+\alpha)!(m+\frac{1}{2}-\frac{1}{2}L'+\alpha)!\nu!}{(2n+2m+4-2\nu+2\alpha-L')!(n-\nu)!(m-\nu)!(\nu+\frac{1}{2}-\frac{1}{2}L'+\alpha)!}\\
&\quad\times\frac{(n+m-\nu-\frac{1}{2}L'+\frac{3}{2}+\alpha)!}{(n+m-\nu+1)!}[P^{(2n+1,0)}(\mathbf{K}_1)\times P^{(2m+1,0)}(\mathbf{K}_2)]^{(2n+2m+2-4\nu,2\nu)}_{L=0}.
\end{aligned}
\tag{32}
$$

The combination of Eqs. (31) and (32) leads to the alternate form for the coefficient $\langle 2n+1,0)L';(2m+1,0)L'\|(\lambda\mu)0\rangle$, given as the last entry in Table II.

III. AN APPLICATION: SU(3)-IRREDUCIBLE TENSOR DECOMPOSITION OF A SCALAR INTERACTION

In recent applications to problems in nuclear collective motion exploiting Sp(3,R) symmetry[7–9] and in nuclear cluster problems[10–12] it has proved useful to expand the rotationally invariant nucleon–nucleon interaction in terms of SU(3) irreducible tensor components. If the two-body interaction with $V=\Sigma_{i<j}V_{ij}$ is given by

$$
V_{12}=\sum_{ST}V_{ST}(|\mathbf{r}_1-\mathbf{r}_2|)P_{ST},
\tag{33}
$$

where P_{ST} is a two-particle spin–isospin projection operator, it frequently proves convenient to expand the radial part of V in terms of Gaussians

$$
V(|\mathbf{r}_1-\mathbf{r}_2|)=V_0\exp(-\beta r^2),
\tag{34}
$$

where $\mathbf{r}$ is the dimensionless relative coordinate

$$
\mathbf{r}=[m\omega/2\hbar]^{1/2}(\mathbf{r}_1-\mathbf{r}_2).
\tag{35}
$$

If such a V is expanded in SU(3) irreducible tensor components

$$
V=\sum_{(\lambda_0\mu_0)}V^{(\lambda_0\mu_0)}_{L_0=0},
\tag{36}
$$

it is sufficient to specify V by its SU(3) reduced matrix elements, defined by[10]

$$
\langle(\bar\lambda\bar\mu)\bar\kappa LM\,|\,V^{(\lambda_0\mu_0)}_{L_0=0}\,|(\lambda\mu)\kappa LM\rangle=(-1)^{\lambda+\mu}\sum_{\rho_0}\frac{\langle(\bar\lambda\bar\mu)\bar\kappa L;(\mu\lambda)\kappa L\,\|(\lambda_0\mu_0)L_0=0\rangle_{\rho_0}\langle(\bar\lambda\bar\mu)\|V^{(\lambda_0\mu_0)}\|(\lambda\mu)\rangle_{\rho_0}}{[(2L+1)]^{1/2}},
\tag{37}
$$

where the SU(3) reduced matrix element is given by the last factor of Eq. (37). Note that the SU(3) reduced matrix element is defined in terms of an unconventional order of the SU(3) coupling. This definition proves convenient when the SU(3) coupling $(\bar\lambda\bar\mu)\times(\lambda\mu)\rightarrow(\lambda_0\mu_0)$ requires an outer multiplicity label ρ_0; that is, when $(\lambda_0\mu_0)$ occurs in the coupling with a multiplicity >1. For a scalar interaction of the relative coordinate $\mathbf{r}$, $V(r)$ is specified by its SU(3) reduced matrix elements in the space of oscillator functions $\phi^{(q0)}_{lm}(\mathbf{r})$ of the single three-dimensional variable $\mathbf{r}$, that is, by the numbers $\langle(\bar q0)\|V^{(\lambda_0\mu_0)}\|(q0)\rangle$. These follow at once from the Bargmann transform of $V(r)$:

$$
\begin{aligned}
\mathscr{V}(\bar{\mathbf{k}},\mathbf{k}^*)&\equiv\int d\mathbf{r}\,A(\bar{\mathbf{k}},\mathbf{r})V(r)A(\mathbf{k}^*,\mathbf{r})\\
&=\sum_{\bar q q}\sum_{(\lambda_0\mu_0)}\langle(\bar q0)\|V^{(\lambda_0\mu_0)}\|(q0)\rangle\\
&\qquad\times[P^{(\bar q0)}(\bar{\mathbf{k}})\times P^{(0q)}(\mathbf{k}^*)]^{(\lambda_0\mu_0)}_{L_0=0},
\end{aligned}
\tag{38}
$$

where we have used the expansion (4) for $A(\bar{\mathbf{k}},\mathbf{r})$, $A(\mathbf{k}^*,\mathbf{r})$, the defining Equation (37), and the orthonormality of the Wigner coefficients. To determine the needed SU(3) reduced matrix elements, it is thus only necessary to evaluate the Bargmann transform of $V(r)$, expand it in the SU(3)-coupled Bargmann space functions, and pick off the coefficient of the $\bar{q}$, q $(\lambda_0\mu_0)$ term. For the Gaussian interaction of Eq. (34)

$$\begin{aligned}\mathscr{V}(\bar{\mathbf{k}}, \mathbf{k}^*) &= \frac{V_0}{(1+\beta)^{3/2}} \\ &\quad\times\exp\left[\frac{(\bar{\mathbf{k}}\cdot\mathbf{k}^*)}{1+\beta} - \frac{\beta}{2(1+\beta)}(\bar{\mathbf{k}}\cdot\bar{\mathbf{k}} + \mathbf{k}^*\cdot\mathbf{k}^*)\right] \\ &= \sum_{a,b,c}\frac{1}{a!b!c!}\left(-\frac{\beta}{2}\right)^{a+b} V_0 \\ &\quad\times\frac{1}{(1+\beta)^{a+b+c+3/2}}(\bar{\mathbf{k}}\cdot\bar{\mathbf{k}})^a(\mathbf{k}^*\cdot\mathbf{k}^*)^b(\bar{\mathbf{k}}\cdot\mathbf{k}^*)^c.\end{aligned} \tag{39}$$

Direct application of Eq. (18) gives

$$\begin{aligned}\mathscr{V}(\bar{\mathbf{k}},\mathbf{k}^*) &= V_0\sum_{a,b,c,\nu}\left(-\frac{\beta}{2}\right)^{a+b} \\ &\times\frac{1}{(1+\beta)^{a+b+c+3/2}}\frac{(-1)^{\min(a-\nu,b-\nu)}}{2} \\ &\times\frac{[(2a+2b-4\nu+2)(2\nu+c)!(2a+2b-2\nu+c+2)!]}{c!\nu!(a+b-\nu+1)!} \\ &\times[P^{(2a+c,0)}(\bar{\mathbf{k}})\times P^{(0,2b+c)}(\mathbf{k}^*)]^{(2a-2\nu,2b-2\nu)}_{L=0}.\end{aligned} \tag{40}$$

With $2a+c=\bar{q}$, $2b+c=q$, $(\lambda_0\mu_0)=(2a-2\nu, 2b-2\nu)$, and Eq. (38) this gives the needed reduced matrix elements. It is convenient to name the summation index $b-\frac{1}{2}\mu_0=m$; note also that $q-\mu_0=\bar{q}-\lambda_0$. The result is

$$\begin{aligned}&\langle(\bar{q}0)\|V^{(\lambda_0\mu_0)}\|(q0)\rangle \\ &\quad= V_0\frac{(-1)^{\min(\lambda_0/2,\mu_0/2)}(-1)^{(\bar{q}-q)/2}(\frac{1}{2}\beta)^{(\lambda_0+\mu_0)/2}}{2(1+\beta)^{(\bar{q}+q+3)/2}} \\ &\qquad\times[(\lambda_0+\mu_0+2)(q-\mu_0)!(\lambda_0+q+2)!]^{1/2} \\ &\qquad\times\sum_{m=0}^{[(q-\mu_0)/2]}\left(\frac{\beta}{2}\right)^{2m} \\ &\qquad\times\frac{1}{m!(q-\mu_0-2m)![\frac{1}{2}(\lambda_0+\mu_0+2)+2m]!}.\end{aligned} \tag{41}$$

An SU(3)-recoupling transformation converts this reduced matrix element for the space of the relative coordinate $\mathbf{r}$ to the full two particle space. If the two- particle states are specified by two-particle relative motion functions $(q0)$ coupled with two-particle center of mass motion functions $(Q\,0)$ to resultant $(\lambda\mu)$, then the two-particle reduced matrix elements are

$$\begin{aligned}&\langle[(\bar{q}0)\times(Q\,0)](\bar{\lambda}\bar{\mu})\|V^{(\lambda_0\mu_0)}\|[(q0)\times(Q\,0)](\lambda\mu)\rangle_{\rho_0} \\ &\quad= U((\bar{\lambda}\bar{\mu})(0Q)(\lambda_0\mu_0)(0q);(\bar{q}0);\_\ \_;(\mu\lambda)\_\rho_0) \\ &\qquad\times(-1)^{\lambda+\mu+\bar{\lambda}+\bar{\mu}}\left[\frac{\dim(\bar{\lambda}\bar{\mu})}{\dim(\bar{q}0)}\right]^{1/2}\langle(\bar{q}0)\|V^{(\lambda_0\mu_0)}\|(q0)\rangle.\end{aligned} \tag{42}$$

Here the U coefficient is an SU(3) Racah coefficient in unitary form.[1,11] For some applications it may be important to make a Talmi–Moshinsky–Brody transformation from the two-particle relative and center of mass motion basis to a two-particle basis expressed in terms of single-particle oscillator functions $\phi^{(q_10)}_{\dots}(\mathbf{r}_1)$, $\phi^{(q_20)}_{\dots}(\mathbf{r}_2)$. In the SU(3)-coupled basis the needed transformation coefficients from the $[(q0)\times(Q\,0)](\lambda\mu)$ to the $[(q_10)\times(q_20)](\lambda\mu)$ basis are simple SU(2) d-functions [see Eq. (4.1.15) of Ref. 19 for our phase convention], and

$$\begin{aligned}&\langle[(\bar{q}_10)\times(\bar{q}_20)](\bar{\lambda}\bar{\mu})\|V^{(\lambda_0\mu_0)}\|[(q_10)\times(q_20)](\lambda\mu)\rangle_{\rho_0} \\ &\quad= \sum_{\bar{q},q,Q} d^{\bar{\lambda}/2}_{(\bar{q}_1-\bar{q}_2)/2,(\bar{q}-Q)/2}(\tfrac{1}{2}\pi)\, d^{\lambda/2}_{(q_1-q_2)/2,(q-Q)/2}(\tfrac{1}{2}\pi) \\ &\qquad\times\langle[(\bar{q}0)\times(Q\,0)](\bar{\lambda}\bar{\mu})\|V^{(\lambda_0\mu_0)}\|[(q0)\times(Q\,0)](\lambda\mu)\rangle_{\rho_0}.\end{aligned} \tag{43}$$

With Eqs. (41), (42), and (43) the full many-particle expression for the full interaction can then be expanded in terms of the SU(3) reduced matrix elements by

$$\begin{aligned}V &= -\tfrac{1}{2}\sum_{\bar{q}_1\bar{q}_2q_1q_2}\ \sum_{(\bar{\lambda}\bar{\mu})(\lambda\mu)}\ \sum_{(\lambda_0\mu_0)\rho_0}\ \sum_{ST}[(2S+1)(2T+1)]^{1/2} \\ &\quad\times\langle[(\bar{q}_10)\times(\bar{q}_20)](\bar{\lambda}\bar{\mu})\|V^{(\lambda_0\mu_0)}_{ST}\|[(q_10)\times(q_20)](\lambda\mu)\rangle_{\rho_0} \\ &\quad\times\big[[a^{+}_{(\bar{q}_10)}\times a^{+}_{(\bar{q}_20)}]^{(\bar{\lambda}\bar{\mu})ST} \\ &\quad\times[a_{(0q_1)}\times a_{(0q_2)}]^{(\mu\lambda)ST}\big]^{(\lambda_0\mu_0)\rho_0S_0=0\ T_0=0}_{L_0=0},\end{aligned} \tag{44}$$

where the square brackets now denote both SU(3) and spin and isospin coupling, and where $a^{+}_{(q_i0)l_im_im_sm_t}$ $(a_{(0q_i)l_im_im_sm_t})$ are single particle creation (annihilation) operators for a particle in the q_ith oscillator shell.

---

[1] J. P. Draayer and Y. Akiyama, J. Math. Phys. **14**, 1904 (1973); Comput. Phys. Comm. **5**, 405 (1973).

[2] D. J. Millener, J. Math. Phys. **19**, 1513 (1978).

[3] J. D. Vergados, Nucl. Phys. A **111**, 681 (1968).

[4] J. P. Elliott, Proc. Roy. Soc. A **245**, 128, 562 (1958).

[5] R. T. Sharp, H. C. von Baeyer, and S. C. Pieper, Nucl. Phys. A **127**, 513 (1969).

[6] H. C. von Baeyer and R. T. Sharp, Nucl. Phys. A **140**, 118 (1970).

[7] G. Rosensteel and D. J. Rowe, Phys. Rev. Lett. **38**, 10 (1977); Ann. Phys. (N. Y.) **126**, 343 (1980).

[8] G. Rosensteel, J. Math. Phys. **21**, 924 (1980).

[9] E. J. Reske and K. T. Hecht, (unpublished).

[10] K. T. Hecht and W. Zahn, Nucl. Phys. A **318**, 1 (1979).

[11] K. T. Hecht, E. J. Reske, T. H. Seligman, and W. Zahn, Nucl. Phys. A **356**, 146 (1981).

[12] Y. Suzuki, E. J. Reske, and K. T. Hecht, Nucl. Phys. A **381**, 77 (1982).

[13] V. Bargmann, Comm. Pure Appl. Math. **14**, 187 (1961).

[14] V. Bargmann, in *Analytic Methods in Mathematical Physics*, edited by R. P. Gilbert and R. G. Newton (Gordon and Breach, New York, 1968), p. 27.

[15] P. Kramer, G. John, and D. Schenzle, *Group Theory and the Interaction of Composite Nucleon Systems* (Vieweg, Braunschweig, 1981), Chap. 4.

[16] L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967).

[17] Sharp *et al.* (Ref. 5) refer to our double-barred coefficient as an SU(3)$\supset$R(3) reduced Clebsch–Gordan coefficient and reserve the term reduced Wigner coefficient for the corresponding 3-$(\lambda\mu)$ symbol. We adhere to the notation of Ref. 1.

[18] L. J. Slater, *Generalized Hypergeometric Functions* (Cambridge U. P., Cambridge, 1966).

[19] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, N. J., 1957).

Reciprocal transformation for one-dimensional conservation equations

Gerald Rosen
Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104

(Received 30 September 1982; accepted for publication 26 November 1982)

One-dimensional conservation equations (OCE) of the form $\partial n/\partial t + \partial f/\partial x = 0$ with $n = n(x,t) > 0$ and $f = f(n, \partial n/\partial x, \partial^2 n/\partial x^2, \cdots)$ admit a symmetric *reciprocal transformation* $x \to x^*(x,t)$, $n \to n^*(x^*,t) \equiv n^{-1}$, $f \to f^* \equiv -n^{-1} f$, which produces an equivalent OCE for n^* in x^* space. Certain OCE of contemporary interest are *reciprocal invariant* in the sense that $f^* = f(n^*, \partial n^*/\partial x^*, \partial^2 n^*/\partial x^{*2}, \cdots)$. There also exists a class of essentially nonlinear OCE for which the reciprocal transformation produces a linear OCE, and thus equations in this class are solvable analytically.

PACS numbers: 02.30.Jr

Consider one-dimensional conservation equations (OCE) of the form

$$\frac{\partial n}{\partial t} + \frac{\partial f}{\partial x} = 0, \tag{1}$$

where the dimensionless density function $n = n(x,t)$ is positive and the flux function $f = f(n, \partial n/\partial x, \partial^2 n/\partial x^2, \cdots)$ is structured algebraically in terms of n and its spatial derivatives. An OCE [Eq. (1)] guarantees existence of a function $x^* = x^*(x,t)$ such that

$$n = \frac{\partial x^*}{\partial x}, \quad f = \frac{-\partial x^*}{\partial t} \tag{2}$$

or, equivalently,

$$dx^* = n\,dx - f\,dt. \tag{3}$$

The latter differential form is endowed with $(x,n,f) \leftrightarrow (x^*,n^*,f^*)$ transposition symmetry if one defines

$$n^*(x^*,t) \equiv [n(x,t)]^{-1}, \quad f^* \equiv -n^{-1} f \tag{4}$$

for then Eqs. (3) and (4) imply

$$dx = n^* dx^* - f^* dt. \tag{5}$$

Since x^* increases monotonically with increasing x according to the first equation in Eqs. (2), the $x \leftrightarrow x^*$ correspondence is one-to-one for fixed t, and thus x can be viewed as a single-valued function of x^* and t. Hence it follows from Eq. (5) that n^* satisfies the OCE

$$\frac{\partial n^*}{\partial t} + \frac{\partial f^*}{\partial x^*} = 0. \tag{6}$$

In Eq. (6), f^* is understood to be expressed in terms of n^* and its derivatives with respect to x^*, i.e.,

$$f^* = -n^* f\Big(n^{*-1}, -n^{*-3}\frac{\partial n^*}{\partial x^*}, 3n^{*-5}\Big(\frac{\partial n^*}{\partial x^*}\Big)^2 - n^{*-4}\Big(\frac{\partial^2 n^*}{\partial x^{*2}}\Big), \cdots\Big), \tag{7}$$

by virtue of the definitions in Eq. (4) and chain-rule differentiation with t held fixed. Thus if $n(>0)$ satisfies the OCE (1) in x space, $n^{-1} \equiv n^*$ (>0) satisfies the equivalent OCE in (6) in x^* space. It is evident by Eqs. (3), (4), and (5) that the correspondence $(x,n,f) \leftrightarrow (x^*,n^*,f^*)$ is symmetric. Observe that the nonsingular one-to-one character of this transformation depends directly on the assumption $n > 0$ through the first equation in Eqs. (2) and the transformation formulas in Eqs. (4). Although the *reciprocal transformation* $(x,n,f) \leftrightarrow (x^*,n^*,f^*)$ has not been discussed in generality heretofore, specialized applications of this transformation have been employed recently by several authors.[1-3]

Observe that Eq. (1) is *reciprocal invariant* if the right side of Eq.(7) reduces to $f(n^*, \partial n^*/\partial x^*, \partial^2 n^*/\partial x^{*2}, \cdots)$, for then the dynamical evolution of n^* in x^* space prescribed by Eq. (6) is identical to the dynamical evolution of n in x space prescribed by Eq. (1). To delineate the class of reciprocal-invariant OCE, one must determine the flux functions which satisfy the condition

$$f^* = f\Big(n^*, \frac{\partial n^*}{\partial x^*}, \frac{\partial^2 n^*}{\partial x^{*2}}, \cdots\Big) \tag{8}$$

with the left side given by Eq. (7). Let us introduce the quantities

$$\xi_k \equiv \Big(n^{-1/2}\frac{\partial}{\partial x}\Big)^k (\ln n), \tag{9}$$

defined for nonnegative integer $k = 0,1,2,3,\cdots$:

$$\xi_0 \equiv \ln n, \quad \xi_1 \equiv n^{-3/2}\frac{\partial n}{\partial x},$$
$$\xi_2 \equiv n^{-2}\frac{\partial^2 n}{\partial x^2} - \frac{3}{2}n^{-3}\Big(\frac{\partial n}{\partial x}\Big)^2, \cdots. \tag{10}$$

If x is replaced by x^* and n is replaced by n^*, the quantities given by Eq. (9) simply change sign,

$$\xi_k \to \xi_k^* \equiv \Big(n^{*-1/2}\frac{\partial}{\partial x^*}\Big)^k \ln n^*$$
$$= \Big(n^{-1/2}\frac{\partial}{\partial x}\Big)^k \ln n^{-1} = -\xi_k, \tag{11}$$

in view of the first members of Eqs. (2) and (4). Moreover, since n and its spatial derivatives are expressible algebraically in terms of the ξ_k's,

$$n = \exp \xi_0, \quad \frac{\partial n}{\partial x} = \xi_1 \exp \tfrac{3}{2}\xi_0,$$
$$\frac{\partial^2 n}{\partial x^2} = (\xi_2 + \tfrac{3}{2}\xi_1^2)\exp 2\xi_0, \cdots, \tag{12}$$

there is no loss of generality in writing

$$f = n^{1/2} g(\boldsymbol{\xi}), \tag{13}$$

where g is a scalar function of $\boldsymbol{\xi} \equiv (\xi_0, \ldots, \xi_N)$ for spatial derivatives up to order N appearing in f.[4] From Eq. (13) and the second definition in Eqs. (4), it follows that

$$f^* = -n^{-1/2}g(\xi). \tag{14}$$

On the other hand, the right side of Eq. (8) is obtained by replacing n by n^* and x by x^* in Eq. (13); according to the first definition in Eqs. (4) and (11), the replacement $(x,n)\rightarrow(x^*,n^*)$ yields

$$f\left(n^*, \frac{\partial n^*}{\partial x^*}, \frac{\partial^2 n^*}{\partial x^{*2}},\cdots\right) = n^{-1/2}g(-\xi). \tag{15}$$

Hence, by substituting Eqs. (14) and (15) into Eq. (8), one finds the necessary and sufficient condition for reciprocal invariance,

$$-g(\xi) = g(-\xi). \tag{16}$$

Therefore we have the following result: Equation (1) is reciprocal invariant if and only if the flux function is expressible as $n^{1/2}$ times an odd function of ξ.

With g in Eq. (13) only required to satisfy the oddness condition [Eq. (16)], the class of reciprocal-invariant OCE is quite broad. The following illustrative reciprocal-invariant OCE are obtained by using Eqs. (10) and (13) to fix f in Eq. (1):

$$g = 2v \sinh \tfrac{1}{2}\xi_0 \Rightarrow \quad \frac{\partial n}{\partial t} + v\frac{\partial n}{\partial x} = 0, \tag{17}$$

$$g = -D\xi_1 \Rightarrow \quad \frac{\partial n}{\partial t} = D\frac{\partial}{\partial x}\left(n^{-1}\frac{\partial n}{\partial x}\right), \tag{18}$$

$$g = -\tfrac{1}{2}D(\cosh \xi_0)^{-1}\xi_1 \Rightarrow$$

$$\frac{\partial n}{\partial t} = D\frac{\partial}{\partial x}\left[(1+n^2)^{-1}\frac{\partial n}{\partial x}\right], \tag{19}$$

$$g = v^{-1}D^2\xi_2 \Rightarrow \frac{\partial n}{\partial t} + v^{-1}D^2\frac{\partial^2}{\partial x^2}\left(n^{-3/2}\frac{\partial n}{\partial x}\right) = 0 \quad (v,D \equiv \text{const}). \tag{20}$$

Equation (17) is the classical linear equation for nondispersive steady-wave propagation, Eqs. (18) and (19) are interesting nonlinear diffusion equations, and Eq. (20) is a nonlinear dispersive-wave-propagation equation. In particular, Eq. (18) has been the subject of recent work[5]; the reciprocal invariance of Eq. (18) implies that the results pertinent to n in x space[5] also hold good for $n^* = n^{-1}$ in x^* space if one makes the appropriate translation of supplementary (boundary and/or initial) conditions.

The most significant practical applications of the reciprocal transformation arise for flux functions of the form

$$f = n\sum_{k=1}^{N} c_k\left(n^{-1}\frac{\partial}{\partial x}\right)^k n^{-1}, \tag{21}$$

where the c_k's are constants. Whereas Eq. (21) makes Eq. (1) essentially nonlinear, the reciprocal transformation yields

$$f^* = -\sum_{k=1}^{N} c_k\frac{\partial^k n^*}{\partial x^{*k}} \tag{22}$$

according to Eqs. (2) and (4); thus Eq. (6) is linear, and the associated initial-value problem is amenable to solution by series-expansion or integral-transform methods. Having solved Eq. (6) for $n^*(x^*,t) = n^{-1}$, one obtains n and x parametrically in terms of x^* and t by integrating Eq. (5). Finally, the algebraic elimination of x^* between the expressions for n and x produces $n = n(x,t)$. As a remarkable illustrative example, the solution to Eqs. (1) and (21), subject to the initial value

$$n(x,0) = (1+ax^2)^{-1/2} \quad (a \equiv \text{positive const}), \tag{23}$$

is obtained as

$$n(x,t) = (e^{bt}+ax^2)^{-1/2}, \tag{24}$$

where

$$b \equiv 2\sum_{j=1}^{N'} c_{2j-1}a^j, \quad N' \equiv \begin{cases} \tfrac{1}{2}N & \text{for } N \text{ even,} \\ \tfrac{1}{2}(N+1) & \text{for } N \text{ odd.}\end{cases}$$

The latter solution is valid without any restriction on the c_k's in Eq. (21) or the value of N.

---

[1] G. Rosen, Phys. Rev. B **19**, 2398 (1979); **23**, 3093 (1981).

[2] J. G. Berryman, J. Math. Phys. **21**, 1326 (1980).

[3] G. Bluman and S. Kumei, J. Math. Phys. **21**, 1019 (1980).

[4] It is interesting to note that Eq. (1) always admits a *self-reciprocal solution* $n = A(t)x^{-2} \Rightarrow n^* = A(t)(x^*)^{-2}$ if g in Eq. (13) is independent of ξ_0; for this self-reciprocal solution $\xi_1 = -2A^{-1/2}$, $\xi_k \equiv 0$ for $k \geqslant 2$ and g is independent of x.

[5] J. G. Berryman and C. J. Holland, J. Math. Phys. **23**, 983 (1982).

Lie transformations, nonlinear evolution equations, and Painlevé forms

M. Lakshmanan and P. Kaliappan
Department of Physics, University of Madras, Autonomous Postgraduate Centre, Tiruchirapalli 620 020, India

(Received 31 August 1981; accepted for publication 29 January 1982)

We present the results of a systematic investigation of invariance properties of a large class of nonlinear evolution equations under a one-parameter continuous (Lie) group of transformations. It is shown that, in general, the corresponding invariant variables (the subclass of which is the usual similarity variables) lead to ordinary differential equations of Painlevé type in the case of inverse scattering transform solvable equations, as conjectured by Ablowitz, Ramani, and Segur. This is found to be also true for certain higher spatial dimensional versions such as the Kadomtsev–Petviashivilli, two dimensional sine–Gordon, and Ernst equations. For the nonsolvable equations considered here this invariance study leads to ordinary differential equations with movable critical points.

PACS numbers: 02.30.Jr, 02.20.Sv

I. INTRODUCTION

The group theoretic analysis of differential equations was advocated by Sophus Lie[1] during the nineteenth century. The basic idea in this analysis is the consideration of the invariance of tangent structural equations under one- or several-parameter transformation groups in conjunction with a given system of differential equations. The invariance conditions enable one to find the infinitesimal transformations and from them the finite transformations, invariant variables, and the Lie algebra associated with a given differential equation. In terms of the invariant variables, the order of an ordinary differential equation can be reduced by one, and in the case of partial differential equations the number of independent variables can be reduced by one. In this way one could obtain a class of interesting similarity solutions, whose importance has been discussed by many authors[2–14] recently in different contexts.

At present there is a revival of interest in the group theoretic analysis of nonlinear partial differential equations. This is due to several reasons. It is known that a class of nonlinear evolution equations solvable by the inverse scattering transform (IST) technique possess a number of common properties such as "solitons," infinite sequence of conservation laws, and Bäcklund transformations with associated geometric and group theoretical properties (see Ref. 15 and the references therein). It is an exciting problem to extend these studies to other systems such as dissipative equations and equations in higher spatial dimensions for which no IST-like technique exists so far. Among the development of a few techniques in this direction, the group theoretical approach is one.[4] The close connection[16–25] between the class of soliton-possessing evolution equations solvable by the IST method and the Painlevé transcendental equations characterized by no movable critical points also prompts one to search for the underlying invariance properties. In fact this connection appears to be even stronger, as some of the recent studies[1,26–30] show that the Bäcklund transformations, Lax criteria for complete integrability, and higher-order conservation laws are closely related to the existence of higher-order Lie–Bäcklund transformations. Such studies also lead to solvability of nonlinear diffusive equations in certain cases.[30] Also certain solvable equations seem to possess a "seeding operator" mechanism in terms of the similarity variables.[31,32]

Motivated by the foregoing considerations we plan to analyze certain nonlinear partial differential equations, which are being intensively studied at present in theoretical physics and applied mathematics, through the use of Lie's method of continuous transformation groups.[1–4] We obtain the appropriate point transformation groups and generators which leave these equations invariant and the corresponding similarity variables and the finite transformations. In terms of the similarity variables, the systems with two independent variables reduce to ordinary differential equations, while for the systems with three independent variables a reduction is made to a partial differential equation with two independent variables. For the latter, again another set of invariant variables is found in terms of which they reduce to ordinary differential equations. Classifying the types of ordinary differential equations which result through the above process, we find that for the systems which are known to be solvable by the IST procedure, and some of their higher dimensional analogs, the ordinary differential equations (under appropriate reductions) fall within one of the 50 canonical forms including the Painlevé transcendental equations with no movable critical points.[2] For the IST nonsolvable equations, the resultant ordinary differential equations in general are found not to fall within the Painlevé class, and so the solutions of such ordinary differential equations are characterized by movable critical points.

The plan of the paper is as follows. To be self-contained, we give a brief outline of Lie's transformation groups as applied to differential equations in Sec. II. In Sec. III, we illustrate the procedure by explicitly working out the invariance properties for two specific equations, namely, generalized KdV equation and two dimensional sine–Gordon equation. In Sec. IV, we present our results in tabular forms for the equations that we have analyzed. In Sec. V, we give a brief

---

[a] On leave from N. G. M. College, Pollachi 642 001, India.

discussion of our results and their implications. In the appendices, we present an analysis of the singular points for the specific cases where movable critical points occur.

II. LIE TRANSFORMATIONS AND SIMILARITY FORMS: THEORY[4]

In this section, we give a brief outline of the theory of Lie's one-parameter group of transformations for invariance of a partial differential equation with two independent variables. Generalization to more variables is straightforward.[4]

A. Invariance and infinitesimal characterization

Consider a partial differential equation with one dependent variable u and two independent variables x and t:

$$H(x,t,u,u_x,u_t,u_{xx},u_{xt},u_{tt},\cdots)=0\,. \tag{2.1}$$

Here subscripts denote partial differentiations. Let a one-parameter (ϵ) group of transformations of the variables x,t,u be taken as

$$x'=f(x,t,u;\epsilon),\quad t'=g(x,t,u;\epsilon),\quad u'=h(x,t,u;\epsilon)\,. \tag{2.2}$$

Let $u=\theta(x,t)$ be a solution of (2.1). If we replace the variables u,x,t in Eq. (2.1) by v, $x'=f(x,t,\theta;\epsilon)$, $t'=g(x,t,\theta;\epsilon)$, Eq. (2.1) becomes

$$H(x',t',v_{x'},v_{t'},v_{x'x'},v_{x't'},v_{t't'},\cdots)=0\,. \tag{2.3}$$

Then $v=\theta(x',t')$ is a solution of (2.3).

We say that the transformations (2.2) leave the Eq. (2.1) invariant iff $v=h(x,t,\theta;\epsilon)$ is a solution to (2.3) whenever $u=\theta(x,t)$ is a solution to (2.1). This condition implies that if Eqs. (2.1) and (2.3) have a unique solution, then

$$\theta(x',t')=h(x,t,\theta(x,t);\epsilon)\,. \tag{2.4}$$

Hence $\theta(x,t)$ satisfies the one-parameter functional equation

$$\theta(f(x,t,\theta;\epsilon),g(x,t,\theta;\epsilon))=h(x,t,\theta;\epsilon)\,. \tag{2.5}$$

Now, expanding (2.2) about the identity $\epsilon=0$, we can generate the following infinitesimal transformations:

$$\begin{aligned} x'&=x+\epsilon\xi(x,t,u)+O(\epsilon^2)\,,\\ t'&=t+\epsilon\tau(x,t,u)+O(\epsilon^2)\,,\\ u'&=u+\epsilon\eta(x,t,u)+O(\epsilon^2)\,. \end{aligned} \tag{2.6}$$

The functions ξ,τ, and η are the infinitesimals of the transformations for the variables x, t, and u, respectively. Then the group (2.6) is extended to the derivative terms. The transformations (2.6) together with the transformations for the first, second,$\cdots$ derivatives are called first, second,$\cdots$ extensions. We shall denote the infinitesimals for $u_x,u_t,u_{xx},u_{xt},\cdots$ by $[\eta_x],[\eta_t],[\eta_{xx}],[\eta_{xt}],\cdots$. Then we have

$$[\eta_x]=\eta_x+(\eta_u-\xi_x)\theta_x-\tau_x\theta_t-\xi_u\theta_u^2-\tau_u\theta_x\theta_t\,. \tag{2.7}$$

Similarly explicit expressions for different higher extensions may be given. Using these various extensions, the infinitesimal criteria for the invariance of (2.1) under the group (2.2) is given by

$$XH|_{H=0}=0\,, \tag{2.8}$$

where the tangent vector field X is given by

$$X=\frac{\partial}{\partial x}+\tau\frac{\partial}{\partial t}+\eta\frac{\partial}{\partial u}+[\eta_x]\frac{\partial}{\partial u_x}+[\eta_{xx}]\frac{\partial}{\partial u_{xx}}+[\eta_t]\frac{\partial}{\partial u_t}+\cdots\,. \tag{2.9}$$

The condition (2.8) provides an algorithm to find ξ, τ, and η. For any solution $u=\theta(x,t)$ of (2.1), Eq. (2.8) may be treated as a form in the derivatives of θ whose coefficients depend on (θ,x,t) and the unknowns (η,ξ,τ). Collecting together the coefficients of like-derivative terms in θ and setting all of them to zero, we get a system of linear partial differential equations from which we can find ξ, τ, and η in practice.

B. Similarity variables, similarity forms and reduction of independent variables

Expanding (2.5) about $\epsilon=0$ with the aid of (2.6), we get

$$\theta(x+\epsilon\xi+O(\epsilon^2),t+\epsilon\tau+O(\epsilon^2))=\theta(x,t)+\epsilon\eta+O(\epsilon^2)\,. \tag{2.10}$$

For known functions of η, ξ, and τ the term of $O(\epsilon)$ leads to the first-order partial differential equation satisfied by $\theta(x,t)$:

$$\xi\theta_x+\tau\theta_t=\eta\,. \tag{2.11}$$

Equation (2.11) is called the *invariant surface condition*. The solutions of (2.11) are obtained by solving the following characteristic equation:

$$\frac{dx}{\xi}=\frac{dt}{\tau}=\frac{d\theta}{\eta}\,. \tag{2.12}$$

The general solution of this equation will involve two arbitrary constants of which one constant takes the role of *similarity variable*, say ζ, and the other, say $f(\zeta)$, which plays the role of a dependent variable. Thus we finally obtain the similarity form of the solution as

$$v=F(x,t,f(\zeta))\,. \tag{2.13}$$

By substituting this relation in Eq. (2.1), we can obtain an ordinary differential equation for f. The results mentioned above can be extended to any number of dependent and independent variables. In this way we can reduce the number of independent variables.

C. Infinitesimal generators and finite transformations

The operator Q given by

$$Q=\xi\frac{\partial}{\partial x}+\tau\frac{\partial}{\partial t}+\eta\frac{\partial}{\partial u} \tag{2.14}$$

is called the infinitesimal operator of the one-parameter group. The solution (2.13) is the corresponding similarity solution for invariance under the action of Q. In other words, the similarity solution (2.13) satisfies the invariant surface condition. The finite transformations are obtained by exponentiation of the infinitesimals through the relations:

$$\begin{aligned} u'&=\exp(\epsilon Q)\cdot u=u+\sum_{n=1}^{\infty}\frac{\epsilon^n}{n!}Q^nu\,,\\ x'&=\exp(\epsilon Q)\cdot x=x+\sum_{n=1}^{\infty}\frac{\epsilon^n}{n!}Q^nx\,,\\ t'&=\exp(\epsilon Q)\cdot t=t+\sum_{n=1}^{\infty}\frac{\epsilon^n}{n!}Q^nt\,. \end{aligned} \tag{2.15}$$

III. EXPLICIT ANALYSIS OF TYPICAL CASES

In this section, we will apply Lie's theory discussed in Sec. II to a couple of specific nonlinear evolution equations to illustrate the method, before presenting the results for the full set (of equations we wish to explore) in Sec. IV.

A. Generalized KdV equation

We consider the invariance of the generalized KdV equation

$$u_t + u^n u_x + u_{xxx} = 0 \quad (n \geqslant 1). \tag{3.1}$$

Following the discussion of Sec. II, the determining equations (2.8) become, after some analysis,

$$\eta_t = \eta_x = \eta_{uu} = 0, \quad -\xi_t + nu^{n-1}\eta - \xi_x u^n + \tau_t u^n = 0,$$

$$\tau_t = 3\xi_x, \quad \xi_{xx} = \xi_u = \tau_x = \tau_u = 0. \tag{3.2}$$

Here subscripts denote partial differentiation.

By solving Eqs. (3.2) we get, for $n = 1$,[6]

$$\xi = \alpha x/3 + \beta t + \gamma, \quad \tau = \alpha t + \delta, \quad \eta = -\tfrac{2}{3}\alpha u + \beta \tag{3.3}$$

and, for $n \geqslant 2$,

$$\xi = \alpha x/3 + \beta, \quad \tau = \alpha t + \delta, \quad \eta = -2\alpha u/3n, \tag{3.4}$$

where the parameters α, β, γ, and δ are constants.

By solving the Lagrange characteristic equation (2.18) for the values of ξ, τ, and η found above, we obtain the following similarity variables ζ and $f(\zeta)$:

For $n = 1$,

$$\begin{aligned}\zeta &= \tfrac{1}{2}\alpha^{5/3}(\alpha t + \delta)^{-1/3}[2\alpha^2 x - 3\beta(\alpha t + \delta) \\ &\quad + 6(\alpha\gamma - \beta\delta)], \\ u &= \tfrac{3}{2}\alpha\beta - \tfrac{9}{4}\beta^2(\alpha t + \beta)^{-2/3} f(\zeta).\end{aligned} \tag{3.5}$$

For $n \geqslant 2$,

$$\begin{aligned}\zeta &= (x + 3\beta/\alpha)(t + \delta/\alpha)^{1/3}, \\ u &= (2/3n)(t + \delta/\alpha)^{-2/3n} f(\zeta).\end{aligned} \tag{3.6}$$

The reduced invariant equation is gotten by substituting (3.5) in (3.1) for $n = 1$ and (3.6) in (3.1) for $n \geqslant 2$. For both the cases the invariant equation can be given as

$$f''' + f^n f' - 2f/n - \zeta f' = 0 \quad \left(' = \frac{d}{d\zeta}\right). \tag{3.7}$$

We have analyzed this equation (see the appendices) for movable critical points by a method given by Ablowitz, Ramani, and Segur.[18] (Hereafter we shall refer to them by ARS.) We find that Eq. (3.7) is of Painlevé type (having no movable critical points) for the cases $n = 1$ and $n = 2$. But it is not of Painlevé type for $n > 2$.

Further for the cases $n = 1, 2$ we can reduce Eq. (3.7) to the second Painlevé type as follows. For $n = 1$, Eq. (3.7) reads

$$f''' + ff' - 2f - \zeta f' = 0. \tag{3.8}$$

We make a transformation[23,24]

$$f = F' - \tfrac{1}{6}F^2. \tag{3.9}$$

Then F satisfies the equation

$$(F'' - \zeta F - \tfrac{1}{18}F^3)'' - \tfrac{1}{3}F(F'' - \zeta F - \tfrac{1}{18}F^3)' = 0. \tag{3.10}$$

This equation can be integrated once to

$$(F'' - \zeta F - \tfrac{1}{18}F^3)' = l \exp\left[+ \int \tfrac{1}{3}F(y)\,dy\right], \tag{3.11}$$

where l is an arbitrary constant. Looking for solutions which are bound as $\zeta \rightarrow \infty$ we have $l = 0$. Another integration then gives

$$F'' - \zeta F - \tfrac{1}{18}F^3 = 0. \tag{3.12}$$

This equation is of the second Painlevé type. For $n = 2$, Eq. (3.7) reads

$$f''' + f^2 f' - f - \zeta f' = 0. \tag{3.13}$$

Integrating once, we have

$$f'' + f^3/3 - \zeta f = \text{const.} \tag{3.14}$$

This equation is also of the second Painlevé type.

The finite transformation Eqs. (2.15) are as follows:

For $n = 1$,

$$\begin{aligned}x' &= A^{1/3}x + (3\beta/2\alpha)(A - A^{1/3})t + [3(\gamma/\alpha)(A^{1/3} - 1) \\ &\qquad + (3\beta\delta/2\alpha^2)(A - 1) - (9\beta\delta/2\alpha^2)(A^{1/3} - 1)], \\ t' &= At + (\delta/\alpha)(A - 1), \\ u' &= A^{-2/3}u - (3\beta/2\alpha)(A^{-2/3} - 1).\end{aligned} \tag{3.15}$$

For $n \geqslant 2$,

$$\begin{aligned}x' &= A^{1/3}x + (3\beta/\alpha)(A^{1/3} - 1), \\ t' &= At + (\delta/\alpha)(A - 1), \\ u' &= A^{-2/3n}u.\end{aligned} \tag{3.16}$$

B.(2 + 1)-dimensional sine–Gordon (sG) equation

The sG equation in (2 + 1) dimensions reads

$$u_{tt} - u_{xx} - u_{yy} + m^2 \sin u = 0. \tag{3.17}$$

We shall apply the following infinitesimal transformations:

$$\begin{aligned}x' &= x + \epsilon\xi_1(x,y,t,u) + O(\epsilon^2), \\ y' &= y + \epsilon\xi_2(x,y,t,u) + O(\epsilon^2), \\ t' &= t + \epsilon\xi_3(x,y,t,u) + O(\epsilon^2), \\ u' &= u + \epsilon\xi_4(x,y,t,u) + O(\epsilon^2).\end{aligned} \tag{3.18}$$

Then the determining equations are of the form

$$\begin{aligned}&\xi_{3u} = \xi_{1u} = \xi_{4uu} = 0, \\ &\xi_{4tt} - \xi_{4uu} - \xi_{4yy} - (\xi_{4y} - 2\xi_{3t})\,m^2 \sin u \\ &\quad + \xi_4 m^2 \cos u = 0, \\ &\xi_{1xx} + \xi_{1yy} - \xi_{1tt} - 2\xi_{4ux} = 0, \\ &\xi_{3xx} + \xi_{3yy} - \xi_{3tt} + 2\xi_{4ut} = 0, \\ &\xi_{2xx} + \xi_{2yy} - \xi_{2tt} - 2\xi_{4uy} = 0. \\ &\xi_{1x} = \xi_{3t}, \quad \xi_{2y} = \xi_{3t}, \quad \xi_{1y} = -\xi_{2x}, \\ &\xi_{2t} = \xi_{3y}, \quad \xi_{1t} = \xi_{3x}.\end{aligned} \tag{3.19}$$

Solving (3.19), the infinitesimals are obtained as

$$\begin{aligned}\xi_1 &= At + Cy + \alpha,\\ \xi_2 &= Bt - Cx + \gamma,\\ \xi_3 &= Ax + By + \beta,\\ \xi_4 &= 0.\end{aligned} \tag{3.20}$$

Then the similarity variables (with $\alpha = \beta = \gamma = 0$) become

$$\begin{aligned}\zeta_1 &= t^2 - x^2 - y^2,\\ \zeta_2 &= Ct - Bx + Ay,\end{aligned} \tag{3.21}$$

and the similarity form for u is

$$u = F(\zeta_1,\zeta_2). \tag{3.22}$$

Then the reduced invariant equation is

$$4\zeta_1 F + 4\zeta_2 F_{\zeta_1\zeta_2} + 6F_{\zeta_1} + KF_{\zeta_2\zeta_2} + m^2 \sin F = 0, \tag{3.23}$$

where $K = C^2 - A^2 - B^2$. The finite transformation group corresponding to the Lorentz invariance is given by

$$(x',y',t',u')^T = \Lambda\,(x,y,t,u)^T, \tag{3.24}$$

where

$$\Lambda = \frac{\delta}{K}\begin{pmatrix} A^2 - C^2 & AB + C\lambda & BC + A\lambda & 0\\ B^2 - C^2 & AB - C\lambda & -BC + B\lambda & 0\\ A^2 - B^2 & AC + B\lambda & BC + A\lambda & 0\\ 0 & 0 & 0 & 1\end{pmatrix} \tag{3.25}$$

with $\lambda = [K(\delta^2 + 2)]^2$, $\cosh[(K\epsilon)^{1/2}] = 1 + \delta$.

Now taking Eq. (3.25) as a new partial differential equation in the variables ζ_1,ζ_2, and F with the following infinitesimal transformations,

$$\zeta_1' = \zeta_1 + \epsilon_1 \xi_5, \quad \zeta_2' = \zeta_2 + \epsilon_1 \xi_6, \quad F' = F + \epsilon_1 \xi_7, \tag{3.26}$$

the determining equations are

$$\begin{aligned}&\xi_7 = 0,\\ &4\zeta_2(\xi_6)_{\zeta_1} + 2K(\xi_6)_{\zeta_2} = 0,\\ &2(\xi_5)_{\zeta_1} - \xi_5/\zeta_1 + (\xi_2/\zeta_1)(\xi_5)_{\zeta_2} = 0,\\ &4\zeta_1(\xi_6)_{\zeta_1\zeta_1} + 4\zeta_2(\xi_6)_{\zeta_1\zeta_2} + 6(\xi_6)_{\zeta_1} + K(\xi_6)_{\zeta_2\zeta_2} = 0,\\ &4\zeta_1(\xi_5)_{\zeta_1\zeta_1} + 6(\xi_5)_{\zeta_1} + K(\xi_5)_{\zeta_2\zeta_2} + 4\zeta_2(\xi_5)_{\zeta_2} = 0.\end{aligned} \tag{3.27}$$

The infinitesimals are

$$\xi_5 = \alpha\zeta_2, \quad \xi_6 = k\alpha/2, \quad \xi_7 = 0, \tag{3.28}$$

where α and k are constants. Then the similarity variables are

$$\zeta = \zeta_1 - \zeta_2^2/k, \quad \varphi = F(\zeta) \tag{3.29}$$

so that the reduced equation becomes

$$4\zeta\varphi_{\zeta\zeta} + 4\varphi_\zeta + m^2 \sin\varphi = 0. \tag{3.30}$$

By making the substitution

$$\omega = e^{i\varphi},$$

Eq. (3.30) reduces to the form

$$\frac{d^2\omega}{d\zeta^2} = \frac{1}{\omega}\left(\frac{d\omega}{d\zeta}\right)^2 - \frac{1}{\zeta}\frac{d\omega}{d\zeta} + \frac{1}{\zeta}(\alpha\omega^2 + \beta), \tag{3.31}$$

where $\alpha = -\beta = -m^2/8$, which is a special case of third Painlevé equation.[2] The finite group corresponding to Eq. (3.28) is

$$\begin{aligned}\zeta_1' &= \zeta_1 + \epsilon\alpha\zeta_2 + \epsilon^2 k\alpha^2/4,\\ \zeta_2' &= \zeta_2 + \epsilon k\alpha/2,\\ F' &= F.\end{aligned} \tag{3.32}$$

IV. EVOLUTION EQUATIONS AND INVARIANT FORMS

The systems that we consider fall into the following three categories.

A. Soliton-possessing IST-solvable evolution equations

(i) Korteweg–de Vries (KdV) equation:

$$u_t + uu_x + u_{xxx} = 0. \tag{4.1}$$

(ii) Modified KdV (MKdV) equation:

$$u_t + u^2 u_x + u_{xxx} = 0. \tag{4.2}$$

(iii) One dimensional sine–Gordon (sG) equation:

$$u_{tt} - u_{xx} + \sin u = 0. \tag{4.3}$$

(iv) Cylindrical KdV (CKdV) equation:

$$u_t + u/2t + 6uu_x + u_{xxx} = 0. \tag{4.4}$$

(v) Boussinesq (B) equation:

$$u_{tt} - u_{xx} - 6(u^2)_{xx} - u_{xxxx} = 0. \tag{4.5}$$

(vi) Nonlinear Schrödinger (NLS) equation:

$$iu_t + u_{xx} + u^2 u^* = 0. \tag{4.6}$$

(vii) Derivative NLS (DNLS)[19]:

$$iu_t = u_{xx} - 4iu^2 u_x^* + 8|u|^4 u. \tag{4.7}$$

(viii) Lund–Regge (LR) system:

$$\begin{aligned}&u_{xt} + \sin u - \left(\frac{\tan^2 u/2}{\sin u}\right) v_x\, v_t = 0,\\ &v_{xt} + (\sin u)^{-1}(u_x v_t + u_t v_x) = 0.\end{aligned} \tag{4.8}$$

(ix) One-dimensional Heisenberg's ferromagnetic system under sterographic projection (equivalent to NLS equation):

$$iu_t + u_{xx} - \frac{2u^*}{1 + uu^*}u_x^2 = 0 \tag{4.9}$$

B. Higher-spatial-dimensional versions

(i) Kadomtsev–Petviashville (KP) equation:

$$(u_t + 6uu_x + u_{xxx})_x + 3\alpha^2 u_{yy} = 0\,. \tag{4.10}$$

(ii) Two dimensional sine–Gordon (2sG) equation:

$$u_{tt} - u_{xx} - u_{yy} + m^2 \sin u = 0\,. \tag{4.11}$$

(iii)(a) Ernst equation for axially symmetric gravitational field (Ernst):

$$(uu^* - 1)[u_{\rho\rho} + (1/\rho)\,u_\rho + u_{zz}] = 2u^*(u_\rho^2 + u_z^2)\,. \tag{4.12}$$

(b) Stationary axially symmetric Einstein–Maxwell (E–M) equation:

$$(uu^* + vv^* - 1)\nabla^2 u = 2\nabla u\cdot(u^*\nabla u + v^*\nabla v),$$
$$(uu^* + vv^* - 1)\nabla^2 v = 2\nabla v\cdot(u^*\nabla u + v^*\nabla v), \qquad (4.13)$$
$$\left(\nabla^2 = \frac{\partial^2}{\partial\rho^2} + \frac{1}{\rho}\frac{\partial}{\partial\rho} + \frac{\partial^2}{\partial z^2}\right).$$

(iv) Stationary axially symmetric Heisenberg's ferromagnetic continuum spin chain (AHS):

$$(uu^* + 1)[u_{\rho\rho} + (1/\rho)\,u_\rho + u_{zz}] = 2u^*(u_\rho^2 + u_z^2). \qquad (4.14)$$

C. Other equations

(i) Generalized KdV(GKdV) equation:

$$u_t + u^n u_x + u_{xxx} = 0, \quad n > 2. \qquad (4.15)$$

(ii) KdV Burger's (KdVB) equation:

$$u_t - \mu u u_x + \nu u_{xxx} = \gamma u_{xx}. \qquad (4.16)$$

(iii) Benjamin–Bona–Mahony (BBM) equation:

$$u_t + u_x + uu_x - u_{xxt} = 0. \qquad (4.17)$$

(iv) Fisher's equation:

$$u_{xx} - u_t + u - u^n = 0. \qquad (4.18)$$

(v) Phi-four (φ^4) equation:

$$u_{tt} - u_{xx} + u - u^3 = 0. \qquad (4.19)$$

For the above equations we have found the infinitesimals of one-parameter point transformation groups. From these infinitesimals we have obtained the corresponding invariant variables and the invariant forms of the solutions. In terms of these similarity variables the invariant ordinary differential equations (ODE's) are obtained. We have further reduced the ODE's with proper substitution to see whether they lead to one of the 50 canonical Painlevé types of equations as enumerated by Ince.[2] We have also adopted the algorithm proposed by ARS[18] in the case of the invariant equations of (4.15), (4.17), and (4.18) for the critical point analysis, and this analysis is given in Appendix A.

In the following, we present our results in tabular forms. Tables I, II, and III denote the same classification of equations as above. Tables IA, IIA, and IIIA contain the infinitesimals of the transformations and the corresponding finite transformations (denoted by primed variables) of the variables involved in the equations. Tables IB, IIB, and IIIB contain the invariant variables, invariant forms of the solutions, and their reduced forms after suitable transformations and further these tables contain the results whether the reduced equations belong to Painlevé type or not.

In Tables IA and IIIA, the symbols ξ, τ, and η denote the infinitesimals of the variables x, t, and u, respectively, and in Tables IB and IIIB the symbols ζ and $f(\zeta)$ denote the invariant variables associated with the values of ξ, τ, and η. Since the Lund–Regge system (4.8) contains one more dependent variable v, we have used the symbols σ and $\varphi(\zeta)$, respectively, to denote the corresponding infinitesimal and the invariant variable. The expressions for σ and $\varphi(\zeta)$ are written along with the results for ζ and $f(\zeta)$.

In Table IIA, ξ_1, ξ_2, ξ_3, and ξ_4 denote the first set of infinitesimals for the variables x, y, t, and u for the KP and 2sG equations and ζ_1, ζ_2, and $F(\zeta_1,\zeta_2)$ denote the invariant variables associated with the above infinitesimals. The symbols ξ_5, ξ_6, and ξ_7 denote the second set of infinitesimals for the new partial differential equation in terms of the new "independent" variables ζ_1 and ζ_2 for the dependent variable $F(\zeta_1,\zeta_2)$, and ζ and $\varphi(\zeta)$ denote the invariant variables for the infinitesimals ξ_5, ξ_6, and ξ_7. In the same table, ξ, τ, and η stand for the infinitesimals of the variables ρ, z, u in the cases of Ernst and HS equations (4.12) and (4.14) and ζ and

TABLE IA. Infinitesimals (ξ, η, and τ) and finite transformations for Eqs. (4.1)–(4.9).

Eqs.	Infinitesimals			Finite transformations		
	ξ	τ	η	x'	t'	u'
KdV	$\alpha x/3 + \beta t + \gamma$	$\alpha t + \delta$	$-2\alpha u/3 + \beta$	$A^{1/3}x + (3\beta/2\alpha)(A - A^{1/3})t$ $+ [(3\gamma/\alpha)(A^{1/3} - 1) + (3\beta\delta/2\alpha^2)(A - 1)$ $- (9\beta\delta/2\alpha^2)(A^{1/3} - 1)]$	$At + (\delta/\alpha)(A - 1)$	$A^{-2/3}u$ $- (3\beta/2\alpha)(A^{-2/3} - 1)$
MKdV	$\alpha x/3 + \beta$	$\alpha t + \delta$	$-\alpha u/3$	$A^{1/3}x + (3\beta/\alpha)(A^{1/3} - 1)$	$At + (\delta/\alpha)(A - 1)$	$A^{-1/3}u$
sG	$\alpha t + \beta$	$\alpha x + \delta$	0	$(x + \delta/\alpha)\cosh A$ $+ (t + \beta/\alpha)\sinh A$	$(t + \beta/\alpha)\cosh A$ $+ (x + \delta/\alpha)\sinh A$	u
CKdV[a]	$\alpha x + \beta x t^{1/2}$ $+ \gamma t^{1/2} + \delta$	$3\alpha t + 2\beta t^{3/2}$	$\beta x/12t^{1/2} - 2\alpha u$ $- 2\beta u t^{1/2} + \gamma/12t^{1/2}$	$Ax + (\delta/\alpha)(A - 1)$	$A^3 t$	$A^{-2}u$
B	α	β	0	$x + \alpha$	$t + \beta$	u
NLS	$\alpha x + \beta t + \gamma$	$2\alpha t + \delta$	$-\alpha u + \frac{1}{2}i\beta u x$ $+ i\lambda u$	$Ax + (\beta/\alpha)(A^2 - A)t$ $+ [(\gamma/\alpha)(A - 1)$ $+ (\beta\delta/2\alpha^2)(A - 1)^2]$	$A^2 t + (\delta/2\alpha)(A - 1)$	$e^{-(\alpha + i\lambda)\epsilon}u$
DNLS	$2\alpha x + \beta$	$4\alpha t + \delta$	$-\alpha u$	$A^2 x + (\beta/2\alpha)(A^2 - 1)$	$A^4 t + (\delta/4\alpha)(A^4 - 1)$	$A^{-1}u$
LR	$\alpha x + \gamma$	$-\alpha t + \delta$	$\eta = 0,\ \sigma = \beta$	$Ax + (\gamma/\alpha)(A - 1)$	$A^{-1}t - (\delta/\alpha)(A^{-1} - 1)$	$u' = u$, $v' = v + \beta$
HS	$\alpha x + \beta$	$2\alpha t + \delta$	$\gamma u^2 + i\lambda u + \hat{\gamma}$	$Ax + (\beta/\alpha)(A - 1)$	$A^2 t + (\delta/2\alpha)(A^2 - 1)$	$(au + b)/(-b^*u + a^*)$, $\lvert a\rvert^2 + \lvert b\rvert^2 = 1$

$\alpha,\beta,\gamma,\delta,a,b,\lambda$, and $\hat{\gamma}$ are arbitrary constants; a, b, $\hat{\gamma}$ are complex. $A = e^{\epsilon\alpha}$.

[a] Finite transformations are given for the case $\beta = 0$ in the infinitesimals.

TABLE IB. Invariant variables (ζ), the invariant form of solutions, and reduction of invariant equations to Painlevé types of equations for the evolution equations (4.1)–(4.9).

Eqs.	Invariant variables (ζ)	Invariant form of solutions (u)	Invariant equations	Reduced forms	P type
KdV	${\tfrac{3}{2}}^{-1/3}\alpha^{-5/3}(\alpha t+\delta)^{-1/3}$ $\times[2\alpha^2x-3\beta(\alpha t+\delta)$ $+6(\alpha\gamma-\beta\delta)]$	$3\beta/2\alpha-3^{-2/3}\alpha^{2/3}$ $\times(\alpha t+\delta)^{-2/3}f(\zeta)$	$f'''=2f+\zeta f'-ff'$	$F''-\zeta F$ $-F^3/18=0$ $(f=F'-F^2/6)$	II
MKdV	$3^{-1/3}(x+3\beta/\alpha)(t+\delta/\alpha)^{-1/3}$	$3^{-1/3}f(\zeta)(t+\delta/\alpha)^{-1/3}$	$f'''=f+\zeta f'-f^2f'$	$f''+f^3/3-\zeta f+C=0$	II
sG	$-(1/2\alpha)[\tfrac{1}{2}\alpha(x^2-t^2)$ $-\beta x+\delta t-(\delta^2-\beta^2)/2\alpha]$	$f(\zeta)$	$\zeta f''+f'+\sin f=0$	$w''=(1/w)w'^2$ $-(1/\zeta)w'-(1/2\zeta)(w^2-1),$ $(w=e^{if})$	III
CKdV[a]	$(x+\delta/\alpha)t^{-1/3}$	$f(\zeta)t^{-2/3}$	$6f'''+36ff'-2\zeta f'-f=0$	$d^2w/dz^2+w^3-zw=0$ $[f=12^{-2/3}w^2,$ $z=12^{-1/3}\zeta]$	II
B	$x-\alpha t/\beta$	$f(\zeta)$	$-f''''+\left(\dfrac{\alpha^2}{\beta^2}-1\right)f''$ $-12ff''-12f'^2=0$	$w''=6w^2+C_1\zeta+C_2$ $[w=f+\tfrac{1}{12}(1-\alpha^2/\beta^2)]$	I
NLS (i) $(\beta=\lambda=0)$	$(x+b/2\alpha)(t+\delta/2\alpha)^{-1/2}$	$-f(\zeta)(t+\delta/2\alpha)^{-1/2}$	$f''+f^2f^*-i\tfrac{1}{2}(\zeta f)'=0$	same as the invariant equation	no movable critical points
(ii) $(\alpha=\lambda=0)$	$(-\beta/\delta)^{1/3}[x$ $-\beta t^2/2\delta-bt/\delta$ $+(b^2/2\delta^2)(\delta/\beta)^{2/3}]$	$(-\beta/\delta)^{1/3}\exp[(i\beta/2\delta)(xt$ $-\beta t^3/3\delta-bt^2/2\delta)]\,p(\zeta)$ $\times\exp[i\textstyle\int q(\zeta)\,d\zeta]$	$2p'q+pq'+(b/\delta)(\beta/\delta)^{-1/3}p'=0,$ $p''-pq^2-(b/\delta)(\beta/\delta)^{-1/3}pq$ $+\zeta p/2-(b^2/4\delta^2)(\beta/\delta)^{-2/3}p+p^3=0$	$w''=2w^3+2w+2ci-\tfrac{1}{2},$ $p^2=-(w'+w^2+\tfrac{1}{2}\zeta)$	II
DNLS	$\dfrac{(x+\beta/2\alpha)}{(2t+\delta/2\alpha)^{1/2}}$	$\dfrac{p(\zeta)\exp[i\int q(\zeta)d\zeta]}{(2t+\delta/2\alpha)^{1/4}}$	$q=p^2-\tfrac{1}{2}\zeta+c/p^2,$ $p''-pq^2-(4p^3+\zeta p)q$ $+8p^5=0$	$w''-w'/2w+6w^3$ $+4\zeta w^2+(-12c+\zeta^2/2)w$ $-2c^2/w=0\ (w=p^2)$	IV
LR	$(x+\gamma/\alpha)(t-\delta/\alpha)$	$u=f(\zeta)$ $v=\dfrac{\beta}{2\alpha}\log\left(\dfrac{x+\gamma/\alpha}{t-\delta/\alpha}\right)$ $+\varphi(\zeta)$	$\varphi'=(c/\zeta)\cot^2f/2,$ $\zeta f''+f'+\sin f$ $+(1/\zeta\sin f)[(\beta^2/4\alpha^2)\tan^2f/2$ $+c^2\cot^2f/2]=0$	$w''=\left[\dfrac{1}{2w}+\dfrac{1}{(w-1)}w'^2\right.$ $\left.\times\dfrac{1}{\zeta}\dfrac{w'}{}+\dfrac{(w-1)^2}{\zeta^2}\left(\dfrac{a}{w}+bw\right)\right]$ $(f=2i\coth^{-1}w^{1/2},$ $a=i\beta^2/8\alpha^2,\ b=ic^2/2)$	V
HS $(\gamma=\lambda=0)$	$(x+\beta/\alpha)(t+\delta/2\alpha)^{1/2}$	$p(\zeta)\exp[i\textstyle\int q(\zeta)\,d\zeta]$	$q=\dfrac{(1+p^2)}{2p^2}+\dfrac{c(1+p^2)^2}{p^2},$ $p''-pq-pq^2-[2p/(1+p^2)]$ $X[p'^2-p^2q^2]=0$	$\dfrac{d^2w}{dz^2}=\left(\dfrac{1}{2w}+\dfrac{1}{w-1}\right)\left(\dfrac{dw}{dz}\right)^2$ $-\dfrac{1}{z}\dfrac{dw}{dz}+\dfrac{(w-1)^2}{z^2}\left(aw+\dfrac{b}{w}\right)$ $[w=p^2,z=e^{\zeta},a=-2c^2,$ $b=2(c^2+c+\tfrac{1}{4})]$	V

α,β,γ, and δ are arbitrary constants. c, C_1 and C_2 are integration constants.

[a] Finite transformations are given for the case $\beta=0$ in the infinitesimals.

$f(\zeta) = p(\zeta)\exp[i \int q(\zeta)\, d\zeta]$ denote the corresponding invariant variables. For Eq. (4.13), in addition to the above, σ stands for the infinitesimals corresponding to v and $\varphi(\zeta) = K(\zeta)\exp[i\int T(\zeta)\, d\zeta]$ denotes another invariant variable.

In finding the finite transformations, and invariant variables in some cases, certain parameters in the infinitesimals are taken to be zero either to avoid lengthy expressions or due to the difficulties encountered in solving the associated Lagrange characteristic equation (2.12). This is indicated at the proper places in the tables.

It might be noted that partial results on some of these equations in class I have appeared already in the literature, and we give references to these results wherever necessary.

Finally, we might add that, although we have found in the present paper all the Lie-point symmetries of the evolutions equations under consideration, we have presented only some of the reduced equations that can be obtained using these symmetries. A further group theoretical analysis is necessary in order to find all the similarity variables. To appreciate this, we may consider the case of the KdV equation discussed here. We find that the use of the full four-parameter group (3.3) leads to the second Painlevé equation (3.12). On the other hand, Fokas and Ablowitz[33] have shown recently that a proper combination of the Galilean and time-invariance groups, cf. (3.3), of the KdV equation leads to the first Painlevé equation. We hope to carry out a systematic analysis of such combinations for the equations considered in this paper in the near future. We thank the referee of this paper for insisting on the necessity of this paragraph and pointing out Ref. 33 to us.

V. DISCUSSION

From our results, in general, it appears that IST-solvable one-dimensional systems and their higher-dimensional versions on reduction to ordinary differential equations can be identified with one of the Painlevé types of equations with no movable critical points. The results seem to corroborate the conjecture by ARS[16–19] on the interconnection between soliton and Painlevé equations. Further, ARS[19] extend the scope of inverse scattering methods to the Painlevé type of ordinary differential equations. They do this by postulating that the linear integral equation

$$K(x,y) = F((x+y)/2) + \int_x^\infty K(x,z)\, N(x,z;y)\, dz, \quad y \geqslant x, \quad (5.1)$$

where $N(x,z;y)$ assumes different forms involving F's for dif-

TABLE IIA. Infinitesimals and finite transformations for the higher-dimensional equations (4.10)–(4.14).[a]

Eqs.	Infinitesimals	Finite transformations
KP	First set of transformations $\xi_1 = ax/3 + \mu t + \beta + by + dy\, t$ $\xi_2 = 2a/3 + c - 6\alpha^2 bt - 3\alpha^2\, dt^2$ $\xi_3 = at + \delta$ $\xi_4 = -2au/3 + \mu/6 + d/6$	for $d = 0$ $x' = Ax + (3b/a)A(A-1)y + A(A-1)[(3\mu/2a)(A+1)$ $-(27\alpha^2 b^2/a^2)(A-1)]t + (3\beta/a)(A-1) + (9bc/2a^2)(A-1)^2$ $+(3\mu\delta/2a^2)(A^3 - 3A + 2) - (27\alpha^2 b^2\delta/a^3)(A-1)^3$ $y' = A^2 y - (18\alpha^2 b/a)A^2(A-1)t + (3c/2a)(A^2 - 1)$ $-(9\alpha^2 b\delta/a^2)(2A^3 - 3A^2 + 1)$ $t' = A^3 t + (\delta/a)(A^3 - 1),\ u' = uA^{-2} - (\mu/4a)(A^{-2} - 1)$ $(A = e^{\epsilon a/3})$
	Second set of transformations $\xi_5 = B\zeta_2 + C$ $\xi_6 = -9\alpha^2 B$ $\xi_7 = \frac{1}{18}(B\zeta_2 - C)$	$\zeta_1' = \zeta_1 + \epsilon(B\zeta_2 + C) - \frac{9}{2}\alpha^2 B^2\epsilon^2$ $\zeta_2' = \zeta_2 - 9\alpha^2 B\epsilon$ $F' = F + (\epsilon/18)(B\zeta_2 - C) - \frac{1}{4}\alpha^2 B^2\epsilon^2$
2sG	First set of transformations $\xi_1 = At + Cy,\ \xi_2 = Bt - Cx,$ $\xi_3 = Ax + By,\ \xi_4 = 0$	$[x',y',t',u']^T = \Lambda\, [x,y,t,u]^T,$ where Λ is as in Eq. (3.25)
	Second set of transformations $\xi_5 = \alpha\zeta_2,\ \xi_6 = K\alpha/2,\ \xi_7 = 0$	$\zeta_1' = \zeta_1 + \epsilon\alpha\zeta_2 + \frac{1}{4}\epsilon^2 K\alpha^2,$ $\zeta_2' = \zeta_2 + \epsilon K\alpha/2,\ F' = F$
Ernst	$\xi = \alpha\rho,\ \tau = \alpha z + \beta,\ \eta = \gamma u^2 + i\hat{\beta}u - \gamma^*$	$\rho' = A\rho,\ z' = Az + \dfrac{\beta}{\alpha}(A-1),$ $u' = (au + b)/b^* u + a^*),\ (aa^* - bb^* = 1)$
AHS	$\xi = \alpha\rho,\ \tau = \alpha z + \beta,\ \eta = \gamma u^2 + i\hat{\beta}u + \gamma^*$	$\rho' = A\rho,\ z' = Az + \dfrac{\beta}{\alpha}(A-1),$ $u' = (au + b)/(-b^* u + a^*),\ (aa^* + bb^* = 1)$
E–M	$\xi = \alpha\rho,\ \tau = \alpha z + \beta,$ $\eta = \alpha u^2 + ibu + iav + \gamma uv - \alpha^*,$ $x = \gamma u^2 + i\beta u + iav + buv - \gamma^*$	$\rho' = A\rho,\ z' = Az + (\beta/\alpha)(A-1),$ $u' = \dfrac{a_{11}u + a_{12}v + a_{13}}{a_{31}u + a_{32}v + a_{33}},$ $v' = \dfrac{a_{21}u + a_{22}v + a_{23}}{a_{31}u + a_{32}v + a_{33}}$ where $\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$ is an arbitrary SU(2,1) matrix

[a] $A = e^{\epsilon\alpha}$.

TABLE IIB. Invariant variables and invariant equations for higher-dimensional systems (4.10)–(4.14) and their reduction to Painlevé type.

Eqs.	Invariant variables	Invariant form of solutions	Invariant equations	Reduced equations	P type
KP	First set of transformations				
	$\zeta_1 = (a^{-5/3}/2(at+\delta)^{1/3})$ $\times[2a^2x - 6aby - 3(\mu + 18\alpha^2 b/a)$ $\times(at+\delta) - 18b(6\alpha^2 b\delta + ac)/a$ $+ 6(a\beta - \mu\delta)]$ $\zeta_2 = (a^{-4/3}/2(at+\delta)^{3/2})$ $\times[2a^2y + 36\alpha^2 b(at+\delta)$ $+ 3(6\alpha^2 b\delta + ac)]$	$F(\zeta_1,\zeta_2)$ $= \frac{3(at+\delta)^{2/3}}{2a^{5/3}}(\mu/6 - 2au/3)$	$\frac{3\partial F}{\partial\zeta_1} + (\zeta_1 + 18F)\frac{\partial^2 F}{\partial\zeta_1^2}$ $+ 18\left(\frac{\partial F}{\partial\zeta_1}\right)^2 + 2\zeta_2\frac{\partial^2 F}{\partial\zeta_1\partial\zeta_2}$ $-\frac{3\partial^4 F}{\partial\zeta_1^4} - \frac{9\alpha^2\partial^2 F}{\partial\zeta_2^2} = 0$	same as the invariant equation	
	Second set of transformations $\zeta = -9\alpha^2 B\zeta_1 - \frac{1}{2}B\zeta_2^2 - C\zeta_2$	$\varphi(\zeta) = (-162\alpha^2 B)\mathrm{F}$ $-\frac{1}{2}B\zeta_2^2 - C\zeta_2$	$2187(\alpha^2 B)^3\varphi'''$ $+ (\zeta + \varphi)\varphi'' + (\varphi')^2$ $+ 2\varphi' + 1 = 0$	$\frac{d^2\chi}{dz^2} = 6\chi^2 + \lambda z$ $[\chi = -(\varphi + \zeta), (6K)^{1/2}z$ $= (\zeta + q/p), K = 4374(\alpha^2 B)^3$ $\lambda = 6{\cdot}6p^{1/2}$, where p,q are arbitrary const]	I
	Special case: For $a = \mu = \beta = c = \delta = 0$ in the infinitesimals (cf. Table IIA) and $\alpha^2 = 1/3$ in KP [Eq. (4.10)] $\zeta_1 = t$, $\zeta_2 = x + \frac{1}{4} - y^2/t$	$F(\zeta_1,\zeta_2) = u$	$\left(\frac{\partial F}{\partial\zeta_1} + \frac{F}{2\zeta_1} + 6F\frac{\partial F}{\partial\zeta_2}\right.$ $\left. + \frac{\partial^3 F}{\partial\zeta^3}\right)_{\zeta_2} = 0$ [same as CKdV Eq. (4.4)]	cf. Table I B	II
2sG	First set of transformations $\zeta_1 = t^2 - x^2 - y^2,$ $\zeta_2 = Ct - Bx + Ay$	$F(\zeta_1,\zeta_2) = u$	$4\zeta_1\frac{\partial^2 F}{\partial\zeta_1^2} + 4\zeta_2\frac{\partial^2 F}{\partial\zeta_1\partial\zeta_2} + 6\,\frac{\partial F}{\partial\zeta_1} + (C^2 - A^2 - B^2)\frac{\partial^2 F}{\partial\zeta_2^2}$ $+ m^2\sin F = 0$	$w = e^{i\varphi},\ \alpha = -\beta = -m^2/8,$	
2sG	Second set of transformations $\zeta = (\zeta_1 - \zeta_2^2)/(C^2 - A^2 - B^2)$	$F(\zeta_1,\zeta_2) = \varphi(\zeta)$	$4\zeta\varphi'' + 4\varphi' + m^2\sin\varphi = 0$	$\frac{d^2w}{d\zeta^2} = \frac{1}{w}\left(\frac{dw}{d\zeta}\right)^2 - \frac{1dw}{\zeta d\zeta}$ $+ \frac{1}{\zeta}(\alpha w^2 + \beta)$	III
Ernst	$\zeta = [(z + \beta/\alpha)^2 + \rho^2]^{1/2}/\rho$ $-(z + \beta/\alpha)/\rho$	For $\gamma = \beta = 0,$ $u = f(\zeta) = p(\zeta)$ $\times\exp[i\int q(\zeta)\,d\zeta]$	$q = c(p^2 - 1)^2/p^2\zeta,$ $p'' - [2p/(2p^2 - 1)](p')^2$ $+ \frac{p'}{\zeta} + c^2(p^2 - 1)^3(p^2 + 1)/p^3\zeta^2 = 0$	$w'' = \left(\frac{1}{2w} + \frac{1}{w-1}\right)w'^2$ $-\frac{w'}{\zeta} + \frac{(w-1)^2}{\zeta^2}$ $\times\left(\frac{\alpha w + \beta}{w}\right)$ $(\alpha = -\beta = -2c^2, w = p^2)$	V

TABLE II B. (*Continued.*)

E–M	$\zeta = \log\{[(z+\beta/\alpha)^2+\rho^2]^{1/2}/\rho$ $-(z+\beta/\alpha)\rho\}$	$v = \hat{K}(\zeta)e^{i\int T(\zeta)d\zeta}$ $u = p(\zeta)e^{i\int q(\zeta)d\zeta}$	$(\hat{K}+p^2-1)(\hat{K}''-KT^2)$ $= 2\hat{K}(\hat{K}'^2-\hat{K}^2T^2)$ $+ 2p(\hat{K}'p'-\hat{K}Tpq),$ $(\hat{K}^2+p^2-1)(2\hat{K}T+\hat{K}T')$ $= 4\hat{K}^2\hat{K}'T+2p(\hat{K}'pq+p'\hat{K}T)$ Another set of similar Eqs. by replacing $\hat{K},T,$ p,q by $p,q,\hat{K},T$, resp.	$\frac{(\hat{K}')^2}{2(\hat{K}^2+p^2-1)^2}$ $=(C-D)^2\frac{1}{8\hat{K}^2}-\frac{C^2\hat{K}^2}{2}+\hat{E},$ $\frac{p'^2}{2(\hat{K}^2+p^2-1)^2}$ $=-(C+D)^2\frac{1}{8p^2}-C^2p^2/2+F$ (C,D,E,F are integration constants)	solutions for $\hat{K},p$ have no movable critical points
HS	$\zeta = [(z+\beta/\alpha)^2+\rho^2]^{1/2}/\rho$ $-(z+\beta/\alpha)/\rho$	for $\gamma=\hat{\beta}=0$ $u=f(\zeta)$ $=p(\zeta)e^{i\int q(\zeta)d\zeta}$	$q=c(p^2+1)/p^2\zeta,$ $p''-\frac{2p}{p^2+1}p'^2+p'/\zeta$ $+\frac{c^2(p^2+1)^3(p^2-1)}{p^3\zeta^2}=0$	$w''=[1/2w+1/(w-1)]w'^2$ $-\frac{w'}{\zeta}+\frac{(w-1)^2}{\zeta^2}\left(\frac{\alpha w+\beta}{w}\right)$ $(\alpha=-\beta=-2c^2, w=-p^2)$	V

ferent evolution equations, is reducible to the corresponding PDE for assumed linear PDE's for the F's. The ordinary differential equations are then obtained by a suitable ansatz for the function $F(x,y)$[19] in terms of self-similar solutions, and these ordinary differential equations are transformed into a Painlevé type.

In our approach we consider group-theoretic or geometric aspects to reduce the IST-solvable equations to ordinary differential equations of Painlevé type. Further, we note that the IST-nonsolvable equations that we have analyzed are reduced, in terms of invariant variables, to ordinary differential equations with movable critical points in general. It is interesting to note that the Fisher equation (4.18) can be reduced to first and second Painlevé types for the cases $n = 2,3$, with specific values of parameters in the invariant variables. This leads to biologically interesting wave solutions of Fisher equation with specific wave speeds $5/6^{1/2}$ in the case[34] of $n = 2$ and $3/2^{1/2}$ in the case[35] of $n = 3$.

Also our analysis leads to physically interesting classes of solutions in some important problems. The Ernst equations with electromagnetic fields corresponding to the coupled Einstein–Maxwell system are explicitly solved for special solutions in terms of elementary functions with a minimal set of integration constants.[36] These solutions are of solitary type and are bounded everywhere. We have also obtained new solutions for the Heisenberg ferromagnetic spin system in its continuous limit with circular, spherical, planar and axial symmetries.[37,38] These solutions exhibit point singularities of defect type in radial variables and so are of physical interest. A model biochemical reaction–diffusion system proposed by Prigogine in one-dimensional unbounded media is reduced to one of Painlevé type for a special value of a parameter in the similarity variable.[39] This study enables us to present a class of one-parameter solutions of simple nature satisfying conditions of biological interest.

ACKNOWLEDGMENTS

P. K. would like to thank the Principal and the management of N. G. M. College, Pollachi, for granting study leave, and the University Grants Commission, New Delhi, for financial assistance under the Faculty Improvement Programme. The work of M. L. forms part of a research scheme supported by the Uuniversity Grants Commission, India.

APPENDIX A

In this appendix, we describe briefly an algorithm proposed by ARS[18] to find certain necessary conditions for an ordinary differential equation to be of Painlevé type and then apply this algorithm to test for the invariant equations of the GKdV and BBM equations (cf. Table IIIB) to be of Painlevé type, or not.

Let us consider an nth order ODE:

$$\frac{d^n w}{dz^n} = F\left(z,w,\frac{dw}{dz},\ldots,\frac{d^{n-1}w}{dz^{n-1}}\right). \tag{A1}$$

For Eq. (A1) to be of Painlevé type, it is *necessary* that it should have no movable branch points, either algebraic or logarithmic. Such necessary conditions may be obtained as follows.

TABLE IIIA. Infinitesimals and finite transformations for Eqs. (4.14)–(4.19).[a]

Eq.	Infinitesimals			Finite transformations		
	ξ	τ	η	x'	t'	u'
GKdV $(n>2)$	$\frac{1}{3}\alpha x+\beta$	$\alpha t+\delta$	$-2\alpha u/3n$	$A^{1/3}x+(3\beta/\alpha)(A^{1/3}-1)$	$At+(\delta/\alpha)(A-1)$	$A^{-2/3n}u$
KdVB	$-\mu\beta t+\delta$	α	β	$x+\epsilon\mu\beta t+\epsilon\delta-\frac{\epsilon^2}{2}\alpha\mu\beta$	$t+\epsilon\alpha$	$u+\epsilon\beta$
BBM	δ	$\alpha t+\beta$	$-\alpha u-\alpha$	$x+\epsilon\delta$	$At+(\beta/\alpha)(A-1)$	$A^{-1}u$ $+(A^{-1}-1)$
Fisher	α	β	0	$x+\alpha\epsilon$	$t+\beta\epsilon$	u
φ^4	$\alpha t+\beta$	$\alpha x+\delta$	0	$(x+\delta/\alpha)\cosh A$ $+(t+\beta/\alpha)\sinh A$	$(t+\beta/\alpha)\cosh A$ $+(x+\delta/\alpha)\sinh A$	u

[a] $A=e^{\alpha\epsilon}$.

TABLE IIIB. Invariant forms and equations with movable critical points for the nonsolvable systems (4.14)–(4.19).

Eqs.	Invariant variables (ζ)	Invariant form of solutions (u)	Invariant equations	Reduced forms	P type
GKdV $(n>2)$	$(x+3\beta/\alpha)(\alpha+\beta)^{-1/3}$	$(\alpha t+\delta)^{-2/3n}f(\zeta)$	$f'''+f^nf'-(2f/n+f')=0$	same as the invariant equation	no [App. A]
KdBV	$x+\mu\beta t^2/2x-\delta t/\alpha$	$\beta t/\alpha+f(\zeta)$	$\alpha\nu f'''-\alpha\gamma f''$ $-(\alpha\mu f+\delta)f'+\beta=0$	$\frac{d^2W}{dZ^2}=6W^2$ $+\left\{K_1-K_2\ln\left[\left(\frac{12\gamma^2}{25\mu}\right)^{1/2}Z\right]\right\}$ $\times\frac{1}{Z^4}$, $(K_1,K_2=\text{const.})$	not in general [App. B]
BBM	$\alpha x-\delta\ln(\alpha t+\beta)$	$f(\zeta)/(\alpha t+\beta)-1$	$f'''+f''+ff'$ $-\lambda(f+f')=0,\ \lambda=\delta^2/\alpha^2$	same as the invariant equation	not in general [App. A]
Fisher	$\beta x-\alpha t$	$f(\zeta)$	$\beta^2f''+\alpha f'+f-f^n=0$	same as the invariant equation	not in general [App. C]
φ^4	$-(1/2\alpha)[\frac{1}{2}\alpha(x^2-t^2)$ $+\beta x-\delta t-(\delta^2-\beta^2)/2\alpha]$	$f(\zeta)$	$\zeta f''+f'+f-f^3=0$	same as the invariant equation	not in general [App. D]

We shall assume that the dominant behavior of the function w in a sufficiently small neighborhood of a movable singularity z_0, is algebraic. In other words,

$$w\sim\alpha(z-z_0)^p \quad \text{as } z\to z_0, \tag{A2}$$

where $\mathrm{Re}(p)$ is negative. We can find the possible sets of values of α and p by requiring that two or more terms in (A1) may balance each other. The terms that balance each other are called leading terms. We may have several choices of p. We require the following two cases to be considered in our discussion.

If all the p's are not integers, then (A2) represents the dominant behavior of w near the algebraic branch point z_0 and so the function w will have a movable branch point. In such a situation, the expansion (A2) should further be proved to be asymptotic near z_0.

If all the possible p's are negative integers, then (A2) may represent the first term of the Laurent series of the solution, for each p, valid in a deleted neighborhood of z_0. In this case a solution for w is

$$w=(z-z_0)^p\sum_{j=0}^{\infty}a_j(z-z_0)^j,\quad 0<|z-z_0|<R. \tag{A3}$$

Apart from the arbitrary constant z_0, if there are $n-1$ aribitrary coefficients a_j, then we have the n constants of integration for Eq. (A1). The powers at which the arbitrary constants enter are called *resonances*. To find the resonances, substitute

$$w=\alpha(z-z_0)^p+\beta(z-z_0)^{p+r} \tag{A4}$$

into Eq. (A1), retaining only the leading terms. The reduced equation (to leading order in β) will be

$$Q(r)(z-z_0)^q\cdot\beta=0,\quad q\geqslant p+r-n, \tag{A5}$$

where $Q(r)$ is a polynomial in r. The roots of $Q(r)=0$ will determine the resonances. Let $r_1\leqslant r_2\leqslant\cdots\leqslant r_s$ denote the positive integral roots of $Q(r)=0$ $(r_s\leqslant n-1)$. Substitute

$$w=\alpha(z-z_0)^p+\sum_{j=1}^{r_s}a_j(z-z_0)^{p+j} \tag{A6}$$

in the given equation. Requiring that the coefficient of $(z-z_0)^{p+j-n}$ (where n is the order of the equation) should vanish, we have the condition

$$Q(j)a_j-R_j(z_0,\alpha,a_1,\ldots,a_{j-1})=0. \tag{A7}$$

For $j<r$, Eq. (A7) determines a_j. At the resonance

$r = r_1$, $Q(r_1) = 0$. If $R_{r_1} = 0$, a_{r_s} is an arbitrary constant of integration, and we can proceed to find the next coefficient. If $R_{r_1} \neq 0$, then there is no solution of the form (A6). In this case we have to introduce logarithmic terms as follows:

$$w = \alpha(z - z_0)^p + \sum_{j=1}^{(r_1-1)} a_j(z - z_0)^{p+j} + [a_{r_1} + b_{r_1} \log(z - z_0)](z - z_0)^{p+r} + \cdots . \qquad \text{(A8)}$$

The coefficient b_{r_1} can be determined by the condition that the coefficient of $(z - z_0)^{p + r_1 - n}$ should vanish, and it can be seen that a_{r_1} is arbitrary. Continuing the expansion (A8) to higher orders, we introduce more and more logarithmic terms. But the introduction of logarithmic terms in (A6) means that w has movable logarithmic branch points. Thus the condition that $R_{r_1} = 0$ is a necessary condition for Eq. (A1) to be of Painlevé type.

Now we shall apply the above procedure to the invariant equations of GKdV and BBM (cf. Table IIIB) to see whether these equations meet the necessary conditions to be of Painlevé type.

Invariant equation of GKdV

The invariant equation (cf. Table IIIB) of GKdV is

$$f''' = -f^n f' + (2/n)f + \zeta f' \quad (n > 2). \qquad \text{(A9)}$$

Substitute

$$f \sim \alpha(\zeta - \zeta_0)^p \qquad \text{(A10)}$$

in (A9). We find that there is only one possibility

$$p = -2/n, \quad \alpha = [-(2/n + 1)(2/n + 2)]^{1/n} \qquad \text{(A11)}$$

with leading terms f''' and $-f^n f'$. Since $n > 2$, Eq. (A9) will have a movable branch point of order $-2/n$ provided (A10) is asymptotic near ζ_0. To see this asymptotic nature, we define

$$f = v^{-(2/n)}.$$

The equation for v is

$$-\frac{2}{n}v^2v''' + \frac{6}{n}\left(\frac{2}{n} + 1\right)vv'v'' - \frac{2}{n} \times \left(\frac{2}{n} + 1\right)\left(\frac{2}{n} + 2\right)v'^3 - \frac{2}{n}v' - \frac{2}{n}v^3 + \frac{2\zeta}{n}v^2v' = 0. \qquad \text{(A12)}$$

There is a regular solution of (A12), that is, regular at ζ_0, if $v(\zeta_0) = 0$, $v'(\zeta_0) + (2/n + 1)(2/n + 2)[v'(\zeta_0)]^3 = 0$, $v''(\zeta_0)$ is a finite quantity, and $v'''(\zeta_0)$ is finite. Then $v(\zeta)$ is analytic at ζ_0 and so (A10) is asymptotic near ζ_0. Thus Eq. (A9) is not of Painlevé type.

Also it may be noted that the KdV (or MKdV) satisfies the necessary conditions for Painlevé type as discussed above. One can easily verify that $p = -2$ (or $p = -1$) and the resonances are $r = 4,6$ (or $r = 3,4$) for KdV (or MKdV) and no logarithmic terms be added at the resonances.

Invariant equation of BBM

Invariant equation (cf. Table IIIB) of BBM is

$$f''' = -f'' - ff' + \lambda(f + f'). \qquad \text{(A13)}$$

Substituting

$$f \sim \alpha(\zeta - \zeta_0)^p \qquad \text{(A14)}$$

in (A13), we find $\alpha = -12$ and $p = -2$ with leading terms f''' and ff'. By putting $f = \alpha(\zeta - \zeta_0)^p + \beta(\zeta - \zeta_0)^{p+r}$ in $f''' = -ff'$, we find that $Q(r) = r^3 - 9r^2 + 14r + 24$. By solving $Q(r) = 0$, we find that the resonances are $r = -1,4,6$. So we have $r_1 = 4$ and $r_2 = 6$. Suppose that

$$f \sim -12(\zeta - \zeta_0)^{-2} + \sum_{j=1}^{6} a_j(\zeta - \zeta_0)^{p+j}; \qquad \text{(A15)}$$

using condition (A7), we find that

$$a_1 = 12/5, \quad a_2 = \lambda + 1/25, \quad a_3 = \lambda + 1/125. \qquad \text{(A16)}$$

But when $j = r_1$ $(= 4)$, we have $R_{r_1}(\zeta_0, a_1, a_2, a_3) = 0$, which means that we have to introduce logarithmic terms in (A15). Thus we find that, at the resonance $r_1 = 4$,

$$f \sim -12(\zeta - \zeta_0)^{-2} + (12/5)(\zeta - \zeta_0)^{-1} + (\lambda + 1/25) + (\lambda + 1/125)(\zeta - \zeta_0) + [a_4 - (6\lambda/25)\log(\zeta - \zeta_0)](\zeta - \zeta_0)^2 + \cdots . \qquad \text{(A17)}$$

Since f contains logarithmic term, Eq. (A13) is not of Painlevé type, in general, if the expansion (A14) is asymptotic near ζ_0, which one can prove as in the previous case.

APPENDIX B

The invariant equation for the KdVB equation (4.16) is

$$\alpha\nu f''' - \alpha\gamma f'' - (\alpha\mu f + \delta)f' + \beta = 0. \qquad \text{(B1)}$$

Integrating once, we have

$$f'' - (\gamma/\nu)f' - (\mu/2\nu)f^2 - \delta f/\alpha\nu + (\beta/\alpha\nu)\zeta + C_1 = 0, \qquad \text{(B2)}$$

where C_1 is the integration constant. Now making the transformations

$$Z = (25\mu\nu/12\gamma^2)^{1/2}\exp(\gamma\zeta/5\nu), \qquad \text{(B3)}$$

and

$$f = W(Z)\exp(2\gamma\zeta/5\nu) - (1/\mu)(6\gamma^2/25\nu + \delta/\alpha), \qquad \text{(B4)}$$

Eq. (B2) can be rewritten as

$$\frac{d^2W}{dZ^2} = 6W^2 + \{K_1 - K_2\ln[(12\gamma^2/25\mu\nu)^{1/2}Z]\}\frac{1}{Z^4} \qquad \text{(B5)}$$

where the constants

$$K_1 = (625/12\gamma^4)\nu^2(6\gamma^2/25\nu + \delta/\alpha) \times (3\gamma^2/25\nu - \delta/2\alpha) + C \qquad \text{(B6)}$$

$$K_2 = (3125/12)\beta\mu\nu^3/\alpha\gamma^5,$$

and

$$C = -625\mu\rho^3C_1/12\gamma^4.$$

It is known that an equation of the form (Ref. 2, p. 324)

$$\frac{d^2W}{dZ^2} = 6W^2 + S(Z) \qquad \text{(B7)}$$

is free from movable critical points only when

$$S(Z) = pZ + q \quad (p,q\text{const}), \qquad \text{(B8)}$$

corresponding to the first Painlevé transcedents. Hence Eq. (B1) has movable critical points, in general [see below Eq. (C11)].

APPENDIX C

The invariant equation for the Fisher's equation (4.18) is

$$\beta^2 f'' + \alpha f' + f - f^n = 0, \tag{C1}$$

where α and β are arbitrary parameters. It is necessary that n should be less than or equal to 3 for the equation (C1) to be of Painlevé type (Ref. 2, p. 326). For $n = 2$, Eq. (C1) becomes

$$\beta^2 f'' + \alpha f' + f - f^2 = 0. \tag{C2}$$

By making the transformations

$$\begin{aligned} W &= \lambda(\zeta) f + \mu(\zeta), \\ Z &= \varphi(\zeta), \end{aligned} \tag{C3}$$

where

$$\lambda = -6\varphi'^2$$

$$\varphi' = \exp(-\alpha\zeta/5\beta^2),$$

$$\mu = +(3\alpha^2/25\beta^2) + \tfrac{1}{2},$$

equation (C3) reduces to

$$\frac{d^2W}{dZ^2} = 6W^2 + \frac{1}{\beta^2}\left[\left(\frac{3\alpha^2}{25\beta^2}\right)^2 - \frac{1}{4}\right]\frac{1}{\lambda\varphi'^2}. \tag{C4}$$

This equation is not in general of Painlevé type. But for the choice $\alpha/\beta = 5/6^{1/2}$, Eq. (C4) reduces to

$$\frac{d^2W}{dZ^2} = 6W^2. \tag{C5}$$

This is of Painlevé type. For $n = 3$, Eq. (C1) becomes

$$\beta^2 f'' + \alpha f' + f - f^3 = 0 \tag{C6}$$

or

$$f'' = -\alpha f'/\beta^2 - f/\beta^2 + f^3/\beta^2. \tag{C7}$$

Making a scale transformation $f \to 2\beta^2 f$, we get

$$f'' = -\alpha f'/\beta^2 - f/\beta^2 + 2f^3. \tag{C8}$$

This equation is not of Painlevé type in general. But for $\alpha/\beta = 3/2^{1/2}$, the above equation can be written as

$$f'' = -3q(\zeta)f' - [q'(\zeta) + 2q^2(\zeta)]f + 2f^3, \tag{C9}$$

where $q = 1/2^{1/2}\beta$. Now Eq. (C9) can be transformed into

$$\frac{d^2W}{dZ^2} = 2W^2, \tag{C10}$$

with the following transformations (see Ref. 2, p. 334):

$$\begin{aligned} W &= f\exp(1/2^{1/2}\beta\zeta), \\ Z &= -2^{1/2}\beta\exp[-(1/2^{1/2}\beta)\zeta]. \end{aligned} \tag{C11}$$

Equation (C10) is of Painlevé type.

We might also point out that in the case of KdVB equation, even though the general invariant Eq. (B1) has movable critical points, the special case $\beta = 0$ and $C_1 = 0$, Eq. (B2) is similar to (C2) and therefore the results of $n = 2$, Fisher's equation, hold here also.

APPENDIX D

The invariant equation for the φ^4 equation (4.19) is

$$\zeta f'' + f' + f - f^3 = 0. \tag{D1}$$

By making the transformations $z = 2\zeta^{1/2}$ and $w = 2^{1/2}f$ in (D1), it reduces to

$$\frac{d^2f}{dz^2} = -\frac{1}{2z}\frac{df}{dz} - w + 2w^3. \tag{D2}$$

For Eq. (D2) to be of P type, it should be expressible in the form (see Ref. 2, p. 328)

$$\frac{d^2f}{dz^2} = -3q(z)\frac{dw}{dz} + 2w^3 - [q'(z) + 2q^2(z)]w. \tag{D3}$$

Since this is not possible, Eq. (D1) is not of P type. However, when $\alpha = 0$, in the infinitesimals of the φ^4 equation (cf. Table IIIA), the solution is of the elliptic function type.

[1] For a list of original references, see, e.g., R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).

[2] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. IV.

[3] W. F. Ames, *Nonlinear Partial Differential Equations* (Academic, New York, 1965), Vol. I, Chap. 4, Vol. II (1972), Chap. 2.

[4] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer-Verlag, Berlin, 1974).

[5] B. K. Harrison and F. B. Estabrook, J. Math. Phys. **12**, 653 (1971).

[6] H. Shen and W. F. Ames, Phys. Lett. A **49**, 313 (1974).

[7] W. Chester, J. Inst. Math. Appl. **19**, 343 (1977).

[8] W. H. Steeb, Z. Naturforsch. **33a**, 742 (1978).

[9] B. Leroy, Lett. Nuovo Cimento **22**, 17 (1978).

[10] M. Lakshmanan and P. Kaliappan, Phys. Lett. A **71**, 166 (1979).

[11] P. Kaliappan and M. Lakshmanan, J. Phys. A **12**, 249 (1979).

[12] S. F. Johnson, K. E. Lonngren, and D. R. Nicholson, Phys. Lett. A **74**, 393 (1979).

[13] M. Humi, J. Math. Phys. **18**, 1705 (1977).

[14] M. Boiti and F. Pempinelli, Nuovo Cimento **51**, 70 (1979).

[15] R. K. Bullough and P. J. Caudrey, *Solitons* (Springer, New York, 1980), Chap. 1.

[16] M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 103 (1977).

[17] M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978).

[18] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).

[19] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 1014 (1980).

[20] B. M. McCoy, C. A. Tracy, and T. T. Wu, J. Math. Phys. **18**, 1058 (1977); Phys. Lett. A **61**, 283 (1977).

[21] R. S. Johnson, Phys. Lett. A **72**, 197 (1979).

[22] D. V. Chudnovsky, "Riemann Monomodromy Problem, Isomonodromy Equations and Completely Integrable Systems," Cargese lectures (June 1979), to be published.

[23] R. R. Rosales, Proc. R. Soc. London Ser. A **361**, 265 (1978).

[24] J. W. Miles, Proc. R. Soc. London Ser. A **361**, 277 (1978).

[25] H. Flaschka and A. C. Newell, Comm. Math. Phys. **76**, 65 (1980).

[26] S. Kumei, J. Math. Phys. **16**, 2461 (1975); **18**, 256 (1977).

[27] P. J. Olver, J. Math. Phys. **18**, 1212 (1977).

[28] N. H. Ibragimov and A. B. Shabat, Dokl. Akad. Nauk SSSR **244**, 57 (1979) [Sov. Phys. Dokl. **24**, 15 (1979)].

[29] A. S. Fokas, J. Math. Phys. **21**, 1318 (1980).

[30] G. Bluman and S. Kumei, J. Math. Phys. **21**, 1019 (1980).

[31] H. C. Morris and R. K. Dodd, Phys. Lett. A **75**, 20 (1979).

[32] H. Flaschka, J. Math. Phys. **21**, 1016 (1980).

[33] A. S. Fokas and M. J. Ablowitz, "On a unified approach to transformations and elementary solutions of Painlevé equations," J. Math. Phys. (1981), in press.

[34] M. J. Ablowitz and A. Zeppetella, Bull. Math. Biol. **41**, 835 (1979).

[35] P. Kaliappan and M. Lakshmanan, unpublished results.

[36] P. Kaliappan and M. Lakshmanan, J. Math. Phys. **22**, 2447 (1981).

[37] M. Lakshmanan, P. Kaliappan and M. Daniel, Phys. Lett. A **75**, 97 (1979).

[38] M. Lakshmanan, M. Daniel, and P. Kaliappan, J. Phys. C **13**, 4743 (1980).

[39] P. Kaliappan, M. Lakshmanan, and P. K. Ponnuswamy, J. Phys. A **13**, L227 (1980).

Kaluza–Klein theories on bundles with homogeneous fibers. I

R. Percacci
Istituto di Fisica Teorica dell'Università di Trieste, Italy,
Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Italy, and Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy
S. Randjbar-Daemi
International Centre for Theoretical Physics, Trieste, Italy

(Received 26 February 1982; accepted for publication 29 October 1982)

We analyze some geometric aspects of Kaluza–Klein theories under the assumption that the $(4 + d)$-dimensional space is a bundle over space–time M with fiber G/H. We formulate the most general metric in the bundle which leads, upon dimensional reduction of the Ricci scalar, to a G-gauge invariant Lagrangian. We find that the treatment of Brans–Dicke-like scalars given by some authors is inconsistent with the bundle-theoretic interpretation.

PACS numbers: 02.40. – k

I. INTRODUCTION

An important progress in Kaluza–Klein (KK) theories has been the discovery of Luciani,[1] later emphasized by Witten,[2] that the same four-dimensional G-Yang–Mills Lagrangian could be obtained if instead of assuming that the extra dimensions belong to the group manifold G, they are realized as a coset space G/H. In this way, considerable economy can be achieved in the number of dimensions needed to realize a given G gauge symmetry, and, furthermore, the number of possible geometries is greatly enlarged. An extensive review on the subject has appeared recently[3]; our purpose here is to fill two gaps: the bundle-theoretic description of the coset space case and the generalization of this theory to include Brans–Dicke-like scalars.

Most of the current literature on the subject of "more dimensions" tends to emphasize that $(4 + d)$-dimensional space–time E is "close" to a product (in metrical sense) of Minkowski space M and some compact d-dimensional space B, and hence all quantities have to be expanded in eigenfunctions of differential operators on B to yield the full spectrum of the four-dimensional theory. In our opinion, there are some objections to this viewpoint: to start with, even the first nonzero eigenvalue of these operators will be of the order of Planck's mass and so will all masses in the theory. (We disregard for the moment the ingenious constructions of Ref. 4). It is well known that a particle with Planck's mass is a very problematic object, and we feel that it is safer to restrict ourselves purely to the zero mass sector, at least as long as we do not have a meaningful theory of quantum gravity. Another related shortcoming of this viewpoint is that it makes sense only in a perturbative approach; for instance, if the Wheeler–Hawking[5] space–time foam picture has some physical reality, then it would be inappropriate to approximate space–time M by Minkowski space because the extra dimensions are supposed to have exactly the same scale as the fluctuations in the metric.[6] For these reasons, it seems that, at present, a consistent treatment of KK theories can only be done classically. In this framework, there is no *a priori* reason to prefer the product structure $E = M \times B$: Letting M be any pseudo-Riemannian manifold and E a nontrivial bundle over M with fiber B would do equally well.[7] Studying these more general possibilities is also a preliminary step to any future nonperturbative quantum treatment of the subject.

The use of bundles is not only relevant in the discussion of global problems. Even locally, say on an open set $U \subset M$, there is a distinction between saying that E is a locally trivial fiber bundle or a Cartesian product $U \times B$. The point is that a bundle space over U is isomorphic to $U \times B$ but the isomorphism (local trivialization) is not given *a priori*, while in the case of $U \times B$ the isomorphism is canonically given. Thus one says that $U \times B$ is a trivial bundle over U, while E need only be trivializable over U.[8] Since trivializations correspond in physical language to gauges, requiring E to be a bundle embodies the concept of gauge symmetry into the geometry. This has been discussed several times in the literature, and we need not insist on it.[9] The important point is the following: Given M and B, the requirement that E be a bundle and that all fields have the canonical dependence on the coordinates of B which is prescribed by bundle theory will yield the gauge invariant sector of the KK theory.

Sections II–VII contain an account of the geometry underlying KK theories on bundles with homogeneous fibers. These sections generalize previous works on the case of principal bundles.[10] The reader who is not so interested in geometrical details should start from the metric (7.1) or (7.2) and will find in Sec. VIII the dimensional reduction of the Ricci scalar of E. We have done this by taking into account a possible dependence of the fiber metric on the base point, so that Brans–Dicke scalars appear, their number varying between 1 and $\frac{1}{2}d(d + 1)$, according to the choice of G and H. These scalars turn out to be singlets under G, and their Lagrangian is similar to that of a nonlinear sigma model. All our results can be specialized to the case of a principal bundle by putting $H = \{e\}$ and making the appropriate changes in the indices. The scalar Lagrangian has been obtained previously in this special case (Refs. 4 and 11) but our results are different, as we discuss in Sec. IX.

We give here a summary of our notation. As a rule, tildes are used for quantities referring to a principal bundle and bars for quantities referring to an associated bundle. The abstract Lie algebra $\mathfrak{G}$ has a basis T_i, $i = 1,...,n = \dim G$; The first $m = \dim H$ generators span $\mathfrak{H}$ and the remaining

0022-2488/83/040807-08$02.50

$d = n - m$ span a complementary space $\mathfrak{P}$. We use indices $\hat{i}, \hat{j},\cdots$ running over $1,\ldots,m$ to label the generators of H and $\alpha, \beta,\cdots$ running over $1,\ldots,d$ to label the generators of G lying in $\mathfrak{P}$. The coordinates on M are x^μ with $\mu = 0,1,2,3$ and those on G/H are y^α with $\alpha = 1,\ldots,d$. We do not need to make distinctions between holonomic and anholonomic indices.

II. HOMOGENEOUS SPACES

For us, a homogeneous space X is a differentiable manifold on which a Lie group G acts transitively and effectively on the left[12]; the group action is a differentiable map $\bar{L}$: $G \times X \rightarrow X$ and fixing $a \in G$ we obtain maps $\bar{L}_a : X \rightarrow X$ by $\bar{L}_a(x) =: \bar{L}(a, x)$. If O is a point of X and $H =: \{a \in G \,|\, \bar{L}_a(O) = O\}$ is the isotropy group of O, then X is diffeomorphic to G/H, the space of left cosets gH. In particular, if $H = \{e\}$, then X is diffeomorphic to G itself; in this case the left action us denoted L: $G \times G \rightarrow G$ and consists of left multiplication. On G there is also a right action R: $G \times G \rightarrow G$ which commutes with L and consists of right multiplication. If we denote by μ: $G \rightarrow G/H$ the canonical projection which maps $g \in G$ to the coset gH, then

$$\mu \circ R_h = \mu \quad \forall \; h \in H \tag{2.1}$$

and

$$\mu \circ L_g = \bar{L}_g \circ \mu \quad \forall \; g \in G. \tag{2.2}$$

Let $\mathfrak{G}$ be the (abstract) Lie algebra of G and $\mathfrak{H}$ be the Lie subalgebra of H. We will assume that G/H is a reductive homogeneous space,[13] which means that there exists a linear subspace $\mathfrak{P}$ of $\mathfrak{G}$ such that

$$\mathfrak{G} = \mathfrak{H} \oplus \mathfrak{P} \tag{2.3}$$

and, denoting by $\mathrm{Ad}_{\mathfrak{G}}(H)$ the restriction of $\mathrm{Ad}(G)$ to H,

$$\mathrm{Ad}_{\mathfrak{G}}(H)\,\mathfrak{P} \subseteq \mathfrak{P}. \tag{2.4}$$

This condition is very weak (for instance, it is always satisfied if H is compact); it is useful because it allows us to identify the tangent space $T_0(G/H)$ with $\mathfrak{P}$ and hence to transport properties of the algebra to the coset space. Choosing the basis T_i of $\mathfrak{G}$ as explained at the end of the previous section, the only nonvanishing structure constants are $C_{\hat{i}\hat{j}}{}^{\hat{k}}$, $C_{\hat{i}\alpha}{}^{\beta}$, $C_{\alpha\beta}{}^{k}$.

III. VECTOR FIELDS ON HOMOGENEOUS SPACES

The Lie algebra $\mathfrak{G}$ of G is by definition the algebra of left-invariant vector fields over G; since every left-invariant vector field is determined by its value at e, $\mathfrak{G}$ is isomorphic as a vector space to T_eG. In a similar manner, T_eG is isomorphic to the space of right-invariant vector fields. Let X_i be a basis for T_eG; then we define the vector fields e_i^{R} and e_i^{L} by

$$e_i^{\mathrm{R}}(g) =: L_{g*}X_i, \tag{3.1}$$

$$e_i^{\mathrm{L}}(g) =: R_{g*}X_i. \tag{3.2}$$

The following relations hold ($\forall a \in G$):

$$L_{a*}e_i^{\mathrm{R}} = e_i^{\mathrm{R}} \quad \text{(left-invariance)}, \tag{3.3}$$

$$R_{a*}e_i^{\mathrm{L}} = e_i^{\mathrm{L}} \quad \text{(right-invariance)}, \tag{3.4}$$

$$R_{a*}e_i^{\mathrm{R}} = \mathrm{Ad}(a^{-1})^k{}_i e_k^{\mathrm{R}}, \tag{3.5}$$

$$L_{a*}e_i^{\mathrm{L}} = \mathrm{Ad}(a)^k{}_i e_k^{\mathrm{L}}. \tag{3.6}$$

Furthermore, at each $g \in G$

$$e_i^{\mathrm{L}}(g) = \mathrm{Ad}(g^{-1})^k{}_i e_k^{\mathrm{R}}(g). \tag{3.7}$$

In infinitesimal form, (3.5) and (3.6) read

$$[e_i^{\mathrm{R}}, e_j^{\mathrm{R}}] = C_{ij}{}^k e_k^{\mathrm{R}}, \tag{3.8}$$

$$[e_i^{\mathrm{L}}, e_j^{\mathrm{L}}] = -C_{ij}{}^k e_k^{\mathrm{L}}, \tag{3.9}$$

and (3.3) and (3.4) read

$$[e_i^{\mathrm{R}}, e_i^{\mathrm{L}}] = 0. \tag{3.10}$$

Equation (3.9) may seem strange at first sight; it can be derived by direct computation from (3.8) and (3.7), but we give here an alternative simple proof: Let I:$G \rightarrow G$ by $I(g) = g^{-1}$; if v is a left-invariant vector field, $I_* v$ is right-invariant and furthermore, $I_*(v(e)) = -v(e)$. Thus $I_* e_i^{\mathrm{R}} = -e_i^{\mathrm{L}}$ and $[e_i^{\mathrm{L}}, e_j^{\mathrm{L}}] = [I_* e_i^{\mathrm{R}}, I_* e_j^{\mathrm{R}}] = C_{ij}{}^k I_* e_k^{\mathrm{R}} = -C_{ij}{}^k e_{\mathrm{R}}^{\mathrm{L}}$.

The vector fields e_i^{L} generate the left translations and the e_i^{R} generate the right translations[14]

$$e_i^{\mathrm{L}}(g) = \frac{d}{dt} L_{\exp(tT_i)}(g)\bigg|_{t=0}, \tag{3.11}$$

$$e_i^{\mathrm{R}}(g) = \frac{d}{dt} R_{\exp(tT_i)}(g)\bigg|_{t=0}. \tag{3.12}$$

On the coset space G/H there are vector fields K_i generating the left action $\bar{L}$:

$$K_i(y) = \frac{d}{dt} \bar{L}_{\exp(tT_i)}(y)\bigg|_{t=0}. \tag{3.13}$$

From (2.2) it follows that

$$K_i = \mu_* e_i^{\mathrm{L}}. \tag{3.14}$$

Also, from (2.1) it follows that

$$\mu_* e_i^{\mathrm{R}} = 0 \quad \text{for} \;\; T_{\hat{i}} \in \mathfrak{H}, \tag{3.15}$$

while $\mu_* e_\alpha^{\mathrm{R}}$ is not defined. The K_i transform under $\bar{L}$ as follows:

$$\bar{L}_{a*}K_i = \mathrm{Ad}(a)^j{}_i K_j; \tag{3.16}$$

hence from (3.9) and (3.14)

$$[K_i, K_j] = -C_{ij}{}^k K_k. \tag{3.17}$$

There are no G-invariant vector fields on G/H, but the linear combinations of vector fields K_i with $T_i \in \mathfrak{H}$ are invariant under $\bar{L}_h$ with $h \in H$. These vector fields vanish at O; on the other hand, the vector fields K_α (with $T_\alpha \in \mathfrak{P}$) are all nonzero in a neighborhood V of O and in fact are linearly independent; thus they may be taken as a basis for vectors on V. In particular, at O the K_α can be identified with the generators T_α in the canonical isomorphism $\mathfrak{P} \rightarrow T_0(G/H)$. From (3.17), we obtain the Lie brackets of this basis

$$[K_\alpha, K_\beta] = -C_{\alpha\beta}{}^{i} K_i{}^{\gamma} K_\gamma. \tag{3.18}$$

Here $K_i{}^\gamma$ denote the components of the generator K_i on the basis $\{K_\alpha\}$; in particular, $K_\alpha{}^\gamma = \delta_\alpha{}^\gamma$ and $K_{\hat{i}}{}^\gamma$ are certain functions of y vanishing at O.

We now derive a formula which will be needed later. Let σ:$G/H \supseteq V \rightarrow G$ be a local section of μ, i.e., a choice of a representative group element in each coset. Then from (3.16) we have $K_i(y) = L_{\sigma(y)*}(\mathrm{Ad}(\sigma(y)^{-1})^j{}_i K_j(O))$ $= \mathrm{Ad}(\sigma(y)^{-1})^\beta{}_i L_{\sigma(y)*} K_\beta(O)$; if we put $i = \alpha$, we have a relation between the bases at y and O, which is invertible. If we denote by $[\mathrm{Ad}(\sigma(y)^{-1})]^{-1\beta}{}_{\alpha}$ the inverse of the $d \times d$ matrix $\mathrm{Ad}(\sigma(y)^{-1})^{\alpha}{}_{\beta}$, then $L_{\sigma(y)*}(K_{\beta}(O))$ $= [\mathrm{Ad}(\sigma(y)^{-1})]^{-1\gamma}{}_{\beta}K_{\gamma}(y)$. Introducing this in the previous relation, we have

$$K_i{}^{\gamma}(y) = \mathrm{Ad}(\sigma(y)^{-1})^{\beta}{}_i\,[\mathrm{Ad}(\sigma(y)^{-1})]^{-1\gamma}{}_{\beta}, \tag{3.19}$$

which implies

$$[\mathrm{Ad}(\sigma(y)^{-1})]^{-1\beta}{}_{\gamma} = \mathrm{Ad}(\sigma(y))^{i}{}_{\gamma}K_i{}^{\beta}. \tag{3.20}$$

IV. RIEMANNIAN STRUCTURES ON HOMOGENEOUS SPACES

In the following we will need some results on the existence of group invariant metrics on homogeneous spaces. We merely quote them and make some comments; the interested reader may find more details in Kobayashi and Nomizu.[13] It is convenient to start from the case of the group manifold.

Proposition 4.1: There is a one–one correspondence between left-invariant Riemannian structures on the group G and inner products in the Lie algebra ($\mathfrak{G}$); there is a one–one correspondence between bi-invariant Riemannian structures on G and $\mathrm{Ad}(G)$ invariant inner products in $\mathfrak{G}$.

Let us denote by γ the inner product in $\mathfrak{G}$; γ defines a left-invariant metric h in the following way: If X, $Y \in T_g G$,

$$h_g(X,Y) = \gamma(L_{g*}^{-1}X, L_{g*}^{-1}Y), \tag{4.1}$$

where we have used the canonical isomorphism $\mathfrak{G} \to T_e G$. Notice that the left-invaraint metric has constant components in the basis of left-invariant vector fields

$$h^{\mathrm{R}}_{ij}(g) = h_g(e_i^{\mathrm{R}}, e_j^{\mathrm{R}}) = \gamma_{ij}, \tag{4.2}$$

but not in the basis of right-invariant vector fields (generators of left translations)

$$h^{\mathrm{L}}_{ij}(g) = h_g(e_i^{\mathrm{L}}, e_j^{\mathrm{L}}) = \mathrm{Ad}(g^{-1})^k{}_i \mathrm{Ad}(g^{-1})^l{}_j \gamma_{kl}. \tag{4.3}$$

It may sometimes be desirable to have a bi-invariant metric; in this case Proposition 4.1 says that γ has to satisfy

$$\gamma_{ij} = \mathrm{Ad}(a)^k{}_i \mathrm{Ad}^l(a)_j \gamma_{kl} \quad \forall \ a \in G. \tag{4.4}$$

The Cartan–Killing form in $\mathfrak{G}$ has this property, but it is not necessarily the unique choice.

In the case of a nontrivial isotropy group we have

Proposition 4.2: There is a one–one correspondence between left-invariant Riemannian structures on G/H and $\mathrm{Ad}_{\mathfrak{P}}(H)$ invariant inner products in the subspace $\mathfrak{P}$ of $\mathfrak{G}$.

Because of (2.4), the matrix $\mathrm{Ad}(a)^i{}_j$ for $a \in H$ is block-diagonal. The condition of the proposition then means

$$\gamma_{\alpha\beta} = \mathrm{Ad}(a)^{\gamma}{}_{\alpha} \mathrm{Ad}(a)^{\delta}{}_{\beta} \gamma_{\gamma\delta} \quad \forall \ a \in H, \tag{4.5}$$

$\gamma_{\alpha\beta}$ being the components of $\gamma_{\mathfrak{P}}$. The reason for this condition is easily understood. If $\sigma: G/H \to G$ is a local section of μ and $X, Y \in T_y\ G/H$, we define the left-invariant metric $\bar{h}$ on G/H in analogy to (4.2)

$$\bar{h}_y(X,Y) = \gamma_{\mathfrak{P}}(\bar{L}_{\sigma(y)*}^{-1}X, \bar{L}_{\sigma(y)*}^{-1}Y). \tag{4.6}$$

The element $\sigma(y)$ is not unique; one could choose another local section $\sigma'(y) = \sigma(y)h(y)$ with $h: G/H \to H$; then

$$\begin{aligned}\gamma_{\mathfrak{P}}&(\bar{L}_{\sigma'(y)*}^{-1}X, \bar{L}_{\sigma'(y)*}^{-1}Y)\\ &= \gamma_{\mathfrak{P}}(\bar{L}_{h(y)*}^{-1}\bar{L}_{\sigma(y)*}^{-1}X, \bar{L}_{h(y)*}^{-1}\bar{L}_{\sigma(y)*}^{-1}Y)\\ &= \gamma_{\mathfrak{P}}(\mathrm{Ad}(h(y)^{-1})\bar{L}_{\sigma(y)*}^{-1}X, \mathrm{Ad}(h(y)^{-1})\bar{L}_{\sigma(y)*}^{-1}Y),\end{aligned}$$

where we have used (3.16). Thus definition (4.6) is independent of the section σ iff $\gamma_{\mathfrak{P}}$ is $\mathrm{Ad}(H)$ invariant. In the case when $H = \{e\}$ this condition becomes void and we are led back to Proposition 4.1.

Finally, we want to point out the existence of averaging procedures for inner products in the Lie algebra. If γ is an inner product in $\mathfrak{G}$ and h is given by (4.1), then denoting by $\eta^L(g)$ the left-invariant Haar measure $\sqrt{h}\, d^n g$, we have from (4.3) $[V_G = \int_G \eta^L(g)]$:

$$\gamma_{ij} = \frac{1}{V_G}\int_G \eta^L(g) h_g(e_i^{\mathrm{R}}, e_j^{\mathrm{R}}) = \frac{1}{V_G}\int_G \eta^L(g)\gamma_{ij}. \tag{4.7}$$

On the other hand, from (4.4)

$$\begin{aligned}\gamma'_{ij} &=: \frac{1}{V_G}\int_G \eta^L(g) h_g(e_i^{\mathrm{L}}, e_j^{\mathrm{L}})\\ &= \frac{1}{V_G}\int_G \eta^L(g)\ \mathrm{Ad}(g^{-1})^k{}_i\ \mathrm{Ad}(g^{-1})^l{}_j \gamma_{kl}\end{aligned} \tag{4.8}$$

is different from γ_{ij}, unless γ is $\mathrm{Ad}(G)$-invariant. However, even if γ is not, γ' *is* $\mathrm{Ad}(G)$-invariant:

$$\begin{aligned}\mathrm{Ad}(a)^i{}_k\ &\mathrm{Ad}(a)^j{}_l \gamma'_{ij}\\ &= \frac{1}{V_G}\int_G \eta^L(g) h_g\ (\mathrm{Ad}(a)^i{}_k e_i^{\mathrm{L}}(g), \mathrm{Ad}(a)^j{}_l e_j^{\mathrm{L}}(g))\\ &= \frac{1}{V_G}\int_G \eta^L(g) h_g(L_{a*}(e_k^{\mathrm{L}}(a^{-1}g)), L_{a*}(e_l^{\mathrm{L}}(a^{-1}g))\\ &= \frac{1}{V_G}\int_G \eta^L(g) h_{a^{-1}g}(e_k^{\mathrm{L}}(a^{-1}g), e_l^{\mathrm{L}}(a^{-1}g))\\ &= \gamma'_{kl} \quad \forall \ a \in G.\end{aligned} \tag{4.9}$$

We will say that γ' is the G average of γ. From (4.8) it follows that γ' is really an inner product: $\gamma'(X,Y) = 0\ \forall\ Y \Leftrightarrow X = 0$ and $\gamma'(X,X) \geqslant 0$, $\forall\ X \in \mathfrak{G}$.

V. BUNDLES WITH HOMOGENEOUS FIBERS

Let $(P,\pi,M;G)$ be a principal bundle over M with structure group G. P is locally trivializable, i.e., $\forall x \in M$ there exists an open set $U \subseteq M$, with $x \in U$, and a diffeomorphism $\psi: \mathrm{U} \times \mathrm{G} \to \pi^{-1}(\mathrm{U})$. Since we are mostly interested in the local structure, in the following we will always refer to this particular open set U. If $R_a: G \to G$ is the right multiplication in G by $a \in G$, there is a right action $\tilde{R}_a: P \to P$, which in any trivialization is given by

$$\tilde{R}_a(\psi(x,g)) =: \psi(x, R_a(g)). \tag{5.1}$$

To every trivialization ψ there is related a preferred local section $s: U \to P$ which is defined by $s(x) = \psi(x,e)$; conversely, every local section defines a trivialization (a gauge). Let s and s' be two sections related by

$$s'(x) = \tilde{R}_{a(x)}(s(x)), \tag{5.2}$$

where $a: U \to G$ is the "gauge function." If $p = \psi(x,g)$, then, in the trivialization ψ' defined by s', $p = \psi'(x,a^{-1}g)$. We will interpret gauge transformations passively as changes of local trivialization. We define the diffeomorphisms $\phi_x : G \rightarrow \pi^{-1}(x)$ by $\phi_x(g) = \psi(x,g)$. Then we have

$$\phi'_x = \phi_x \circ L_{a(x)}. \tag{5.3}$$

It is well known that $(G,\mu,G/H;H)$ is a principal bundle. Let $(E,\eta,M;G/H)$ be the fiber bundle with fiber G/H associated with $(P,\pi,M;G)$; by a well-known theorem[13] E is the quotient of P by the right action of H (as a subgroup of G); we denote by $\tau: P \rightarrow E$ the canonical projection. Then $(P,\tau,E;H)$ is a principal bundle and, $\forall\, x \in M$, $(\pi^{-1}(x),\tau,\eta^{-1}(x);H)$ is a principal bundle isomorphic to $(G,\mu,G/H;H)$.[15] If ψ is a local trivialization of P over U as above, then there is a local trivialization of E over $U \bar{\psi}: U \times G/H \rightarrow \eta^{-1}(U)$, which is defined as follows: If $w = \tau(p)$ and $p = \psi(x,g)$, then $w = \bar{\psi}(x,\mu(g))$. If we construct the diffeomorphisms $\bar{\phi}_x : G/H \rightarrow \eta^{-1}(x)$ out of $\bar{\psi}$ in the same way ϕ_x was constructed out of ψ, then we have a commutative diagram

$$\begin{tikzcd} P \supseteq \pi^{-1}(x) \arrow[d, "\tau"'] & G \arrow[l, "\phi_x"'] \arrow[d, "\mu"] \\ E \supseteq \eta^{-1}(x) & G/H \arrow[l, "\bar{\phi}_x"'] \end{tikzcd} \tag{5.4}$$

If $w = \bar{\psi}(x,y)$ for $y \in G/H$, $\bar{\psi}'$ is the trivialization of E derived from the trivialization ψ' of P given by (5.2), then $w = \bar{\psi}'(x,\bar{L}_{a^{-1}}(y))$. The map $\bar{s} = \tau \circ s: U \rightarrow E$ is a local section of E and $\bar{s}(x) = \bar{\psi}(x,0)$, where $0 = \mu(e) = eH$ is the "origin" of G/H. It should be noted, however, that while G acts freely on itself, the action on G/H is assumed to be only effective and hence the assignment of the section $\bar{s}$ is not sufficient to specify completely the local trivialization of E: The section $\bar{s}$ is unchanged by the gauge transformations $a: U \rightarrow H$. Hence, to have a complete description of the gauge structure, we will always need to refer to the principal bundle.

Let Γ be a connection (a distribution of horizontal spaces) in P; then a connection $\bar{\Gamma}$ is defined in E in the following way: A vector in E is horizontal if it is the image under τ_* of a horizontal vector in P (this is the definition given in Ref. 13, specialized to the case when the left action of G on the typical fiber of the associated bundle is transitive).

VI. BASES

In the principal bundle P it is customary[10] to assume as a basis for the vertical vectors the so-called fundamental vector fields, which are the generators of the group action $\tilde{R}$ on the space P:

$$\tilde{e}_i^{\mathrm{R}}(p) = \frac{d}{dt} \tilde{R}_{\exp(tT_i)}(p) \bigg|_{t=0}. \tag{6.1}$$

On the associated bundles there is no group action, and hence there are no fundamental vector fields; therefore, the basis has to be chosen in some other way.

To this end let us first remark that the vector fields e_i^{R} can be defined in a given trivialization by

$$\tilde{e}_i^{\mathrm{R}}(p) = \phi_{x*} e_i^{\mathrm{R}}(g) \tag{6.2}$$

if $p = \phi_x(g)$; this follows easily from (5.1) and (6.1). That this definition does not depend on the trivialization follows from the left invariance of the vector fields e_i^{R}: From (5.3)

$$\phi'_{x*} e_i^{\mathrm{R}} = \phi_{x*} \circ L_{a(x)*} e_i^{\mathrm{R}} = \phi_{x*} e_i^{\mathrm{R}}.$$

The $\tilde{e}_i^{\mathrm{R}}$ are not the only possible choice, however: Any basis, for instance a coordinate basis on G, can be transported to $\pi^{-1}(U)$ by means of the diffeomorphisms ϕ_x. The other choice which will be used in this paper are the right-invariant vector fields

$$\tilde{e}_i^{\mathrm{L}}(p) = \phi_{x*} e_i^{\mathrm{L}}(g). \tag{6.3}$$

Since the ϕ_x are diffeomorphisms, all of the relations (3.3)–(3.10) will be satisfied by the fields $\tilde{e}_i^{\mathrm{L}}$ and $\tilde{e}_i^{\mathrm{R}}$ in each fiber; in particular their Lie brackets are given by

$$[\tilde{e}_i^{\mathrm{R}},\tilde{e}_j^{\mathrm{R}}] = C_{ij}{}^k \tilde{e}_k^{\mathrm{R}}, \tag{6.4}$$

$$[\tilde{e}_i^{\mathrm{L}},\tilde{e}_j^{\mathrm{L}}] = -C_{ij}{}^k \tilde{e}_k^{\mathrm{L}}, \tag{6.5}$$

$$[\tilde{e}_i^{\mathrm{L}},\tilde{e}_j^{\mathrm{R}}] = 0. \tag{6.6}$$

The same type of construction can be applied to the case of the associated bundles.

Given a basis on G/H, one can induce a basis for vertical vectors on $\eta^{-1}(x)$ using the diffeomorphism $\bar{\phi}_x$ as in (6.2) and (6.3). However, due to the nonexistence of G-invariant vector fields on G/H, this basis will always depend upon the choice of the trivialization. We will always use what we will call the Killing basis for vertical vectors: If $w = \bar{\psi}(x,y)$, we define

$$\bar{e}_i(w) = \bar{\phi}_{x*} K_i(y). \tag{6.7}$$

As in Sec. III, the subset of vector fields $\{\bar{e}_\alpha\}$ is then the desired basis; its Lie brackets are obtained by acting with $\bar{\phi}_{x*}$ on Eq. (3.18):

$$[\bar{e}_\alpha,\bar{e}_\beta] = -C_{\alpha\beta}{}^i K_i{}^\gamma(y) \bar{e}_\gamma. \tag{6.8}$$

In a different trivialization

$$\bar{e}'_\alpha(w) = \bar{\phi}'_{x*} K_\alpha(y) = \bar{\phi}_{x*} \circ \bar{L}_{a(x)*} K_\alpha(y) = \mathrm{Ad}(a)_\alpha^j \bar{e}_j(w). \tag{6.9}$$

In view of (3.14) and (5.4), we have

$$\bar{e}_i = \tau_* \tilde{e}_i^{\mathrm{L}}. \tag{6.10}$$

In the complementary directions, there are two convenient choices. If the open set U is a coordinate patch for M and $\{\partial_\mu\}$ is a natural basis for vectors on U, then we may construct on $\eta^{-1}(U)$ either the vector fields tangent to the sections $\bar{s}_y : U \rightarrow \eta^{-1}(U)$ defined by $\bar{s}_y(x) = \bar{\psi}(x,y)$, or their horizontal parts. At a given point $w = \bar{\psi}(x,y) = \bar{s}_y(x)$, the vectors tangent to the submanifold $\bar{s}_y(U)$ and projecting upon the vectors ∂_μ on U will be denoted by $\bar{\partial}_\mu$; their horizontal parts with respect to the distribution $\bar{\Gamma}$ will be denoted $\bar{e}_\mu$. Similarly, in the principal bundle we denote by $\tilde{\partial}_\mu$ the vectors tangent to the constant section $s_g : U \rightarrow \pi^{-1}(U)$ defined by $s_g(x) = \psi(x,g)$ and by $\tilde{e}_\mu$ their horizontal parts with regard to Γ (the covariant derivatives); then we have

$$\bar{\partial}_\mu = \tau_* \tilde{\partial}_\mu, \quad \bar{e}_\mu = \tau_* \tilde{e}_\mu. \tag{6.11}$$

Let ω be the connection form on P; in a given trivialization ψ we may decompose

$$\omega = \omega_\mu^i \tilde{d}x^\mu \otimes T_i, \tag{6.12}$$

where $\tilde{d}x^\mu$ are the duals of the $\tilde{\partial}_\mu$. The components depend on the coordinates by

$$\omega^i_\mu(x,g) = \mathrm{Ad}(g^{-1})^i{}_k\,\omega^k_\mu(x,e). \tag{6.13}$$

The gauge potential on M is (in the gauge defined by the trivialization ψ)

$$A = s^*\omega, \tag{6.14}$$

and hence we may identify

$$A^{\,i}_{\,\mu}(x) = \omega^i_\mu(x,e). \tag{6.15}$$

Similarly, let $\Omega = \mathrm{hor}\; d\omega$ be the curvature form on P; we may decompose

$$\Omega = \tfrac{1}{2}\,\Omega_{\mu\nu}{}^i\tilde{d}x^\mu \wedge \tilde{d}x^\nu \otimes T_i. \tag{6.16}$$

The components depend on the coordinates by

$$\Omega_{\mu\nu}{}^i(x,g) = \mathrm{Ad}(g^{-1})^i{}_j\,\Omega_{\mu\nu}{}^j(x,e). \tag{6.17}$$

The field strength on M is (in the gauge defined by the trivialization ψ)

$$F = s^*\Omega, \tag{6.18}$$

and hence we may identify

$$F_{\mu\nu}{}^i(x) = \Omega_{\mu\nu}{}^i(x,e). \tag{6.19}$$

The horizontal basis vectors are given by

$$\tilde{e}_\mu = \tilde{\partial}_\mu - \omega^i_\mu\,\tilde{e}^{\mathrm{R}}_i, \tag{6.20}$$

or, using the vertical vectors (6.3) and applying relation (3.7),

$$\tilde{e}_\mu = \tilde{\partial}_\mu - A^{\,i}_{\,\mu}\tilde{e}^{\mathrm{L}}_i. \tag{6.21}$$

From (6.10) and (6.11) we then obtain the form of the horizontal vectors in E:

$$\begin{aligned}\bar{e}_\mu &= \bar{\partial}_\mu - A^{\,i}_{\,\mu}\,\bar{e}_i\\ &= \bar{\partial}_\mu - \bar{A}^{\,\alpha}_{\,\mu}\,\bar{e}_\alpha,\end{aligned} \tag{6.22}$$

where in the second line we have used the linearly independent vectors only and have defined

$$\bar{A}_\mu{}^\alpha = A^{\,i}_{\,\mu}\,K^\alpha_i = A^{\,\alpha}_{\,\mu} + A^{\,\hat{i}}_{\,\mu}\,K^\alpha_{\hat{i}}\,. \tag{6.23}$$

We need the Lie brackets between the basis vectors. In P

$$[\tilde{\partial}_\mu,\tilde{e}^{\mathrm{R}}_i] = 0, \tag{6.24}$$

$$[\tilde{e}_\mu,\tilde{e}^{\mathrm{R}}_i] = 0, \tag{6.25}$$

and

$$[\tilde{\partial}_\mu,\tilde{e}^{\mathrm{L}}_i] = 0, \tag{6.26}$$

$$[\tilde{e}_\mu,\tilde{e}^{\mathrm{L}}_i] = A_\mu{}^k C_{ki}{}^j\tilde{e}^{\mathrm{L}}_j\,. \tag{6.27}$$

Furthermore, using (6.17) and (3.7),

$$\begin{aligned}[\tilde{e}^{\mathrm{L}}_\mu,\tilde{e}^{\mathrm{L}}_\nu] &= -\,\Omega_{\mu\nu}{}^i\tilde{e}^{\mathrm{R}}_i\\ &= -\,F_{\mu\nu}{}^i\tilde{e}^{\mathrm{L}}_i.\end{aligned} \tag{6.28}$$

In E

$$[\bar{\partial}_\mu,\bar{e}_\alpha] = 0, \tag{6.29}$$

$$[\bar{e}_\mu,\bar{e}_\alpha] = d_{\mu\alpha}{}^\beta\bar{e}_\beta, \tag{6.30}$$

where

$$d_{\mu\alpha}{}^\beta = A^{\,k}_{\,\mu}c_{k\alpha}{}^jK^\beta_j = A^{\,k}_{\,\mu}(c_{k\alpha}{}^\beta + c_{k\alpha}{}^{\hat{j}}K^\beta_{\hat{j}}), \tag{6.31}$$

and

$$[\bar{e}_\mu,\bar{e}_\nu] = d_{\mu\nu}{}^\alpha\bar{e}_\alpha, \tag{6.32}$$

where

$$d_{\mu\nu}{}^\alpha = F_{\mu\nu}{}^iK^\alpha_i = F_{\mu\nu}{}^\alpha + F_{\mu\nu}{}^{\hat{j}}K^\alpha_{\hat{j}}. \tag{6.33}$$

VII. GAUGE-INVARIANT RIEMANNIAN STRUCTURES

According to our general scheme explained in the Introduction, we need to find the most general metric $\bar{g}$ on the bundle space E which is compatible with the gauge structure, in the following sense:

(i) $\bar{g}$ is defined in a trivialization independent way;

(ii) $\bar{g}$ makes the horizontal and vertical spaces orthogonal to each other.

The first condition guarantees that the theory will be gauge invariant, while the second constitutes the link between the gauge structure and the Riemannian geometry of E.

The most general metric which satisfies these requirements can be fixed at a point w by assigning the inner products in the vertical and horizontal subspaces of T_wE separately.

Let g be a (pseudo-) Riemannian structure in M; we require $\bar{g}$ to be such that η_* maps the horizontal space at each point w isometrically onto $T_{\eta(w)}M$. Concerning the vertical spaces, one usually requires that the diffeomorphisms $\bar{\phi}_x$ be isometries when G/H is given some metric $\bar{h}$. This defines a trivialization independent metric in each fiber $\eta^{-1}(x)$ iff $\bar{h}$ is left-invariant, because then $\forall\, X,Y \in \mathrm{ver}\; T_wE$ we have under (5.3):

$$\begin{aligned}&\bar{h}_{\bar{\phi}'^{-1}_x(w)}(\bar{\phi}'^{-1}_{x*}X,\bar{\phi}'^{-1}_{x*}Y)\\ &\qquad= \bar{h}_{\bar{L}^{-1}_a\circ\bar{\phi}^{-1}_x(w)}(\bar{L}^{-1}_{a*}\circ\bar{\phi}^{-1}_{x*}X,\bar{L}^{-1}_{a*}\circ\bar{\phi}^{-1}_{x*}Y)\\ &\qquad= (\bar{L}^{-1*}_a h)_{\bar{\phi}^{-1}_x(w)}(\bar{\phi}^{-1}_{x*}X,\bar{\phi}^{-1}_{x*}Y)\\ &\qquad= \bar{h}_{\bar{\phi}^{-1}_x(w)}(\bar{\phi}^{-1}_{x*}X,\bar{\phi}^{-1}_{x*}Y).\end{aligned}$$

We will be more general than that because conditions (i) and (ii) allow to vary the metric $\bar{h}$ from fiber to fiber within the class of left-invariant metrics. This can be properly formalized in the following way. Let $\{T_\alpha\}$ be an oriented basis for $\mathfrak{P}$ as usual; all other bases with the same orientation are related to this by the action of an element of the group $\mathrm{GL}^+(d,\mathbb{R})$; the space of all inner products in $\mathfrak{P}$ is then $\mathrm{GL}^+(d,\mathbb{R})/\mathrm{SO}(d)$. This space is diffeomorphic to $\mathbb{R}^{d(d+1)/2}$ and taking as a chart this diffeomorphism, the coordinates of a point γ in $\mathrm{GL}^+(d,\mathbb{R})/\mathrm{SO}(d)$ are nothing but the components $\gamma_{\alpha\beta} = \gamma(T_\alpha,T_\beta)$. By Proposition 4.2 the space $\mathfrak{T}$ of left-invariant metrics in G/H is the algebraic submanifold of $\mathrm{GL}^+(d,\mathbb{R})/\mathrm{SO}(d)$ defined by Eq. (4.5).

Now let $\gamma{:}M\to\mathfrak{T}$ be a smooth map and let $\bar{h}(x)$ denote the left-invariant metric on G/H corresponding to $\gamma(x)$ through Proposition 4.2. We require $\bar{g}$ to be such that the diffeomorphism $\bar{\phi}_x{:}G/H\to\eta^{-1}(x)$ be an isometry when G/H is given the metric $\bar{h}(x)$. The components $\gamma_{\alpha\beta}$ will be seen in the next section to behave like scalar fields on M; the number of independent fields, however, equals the dimension of $\mathfrak{T}$ and has to be found case by case by computing the rank of the algebraic system (4.5). Furthermore, it is clear from this geometric definition that the fields $\gamma_{\alpha\beta}$ are completely unaffected by changes of trivialization; in other words, they must behave like singlets under gauge transformations.

We now look for the components of the metric $\bar g$ at a point $w \in E$. If $\bar\psi$ is a local trivialization such that $w = \bar\psi(x,y)$ and $\{\bar e_\alpha, \bar e_\mu\}$ is the "horizontal lift basis" given by Eqs. (6.7) and (6.11), then

$$\bar g_{\mu\nu}(w) = \bar g_w(\bar e_\mu, \bar e_\nu) = : g(\eta_* \bar e_\mu, \eta_* \bar e_\nu)|_{\eta(w)}$$

$$= g(\partial_\mu, \partial_\nu)|_x = g_{\mu\nu}(x),$$

$$\bar g_{\mu\alpha}(w) = \bar g_w(\bar e_\mu, \bar e_\alpha) = 0,$$

$$\bar g_{\alpha\beta}(w) = \bar g_w(\bar e_\alpha, \bar e_\beta) = : \bar h\,(\eta(w))(\bar\phi_{x*}^{-1}\bar e_\alpha, \bar\phi_{x*}^{-1}\bar e_\beta)|_{\bar\phi_x^{-1}(w)}$$

$$= \bar h\,(x)(K_\alpha, K_\beta)|_y = : \bar h_{\alpha\beta}(x,y).$$

Thus the metric $\bar g$ has components

$$\begin{pmatrix} g_{\mu\nu}(x) & 0 \\ 0 & \bar h_{\alpha\beta}(x,y) \end{pmatrix}. \tag{7.1}$$

Transforming to the basis $\{\bar e_\alpha, \bar\partial_\mu\}$, we obtain

$$\begin{pmatrix} g_{\mu\nu}(x) + \bar h_{\alpha\beta}(x,y) K_i^{\ \alpha}(y) K_j^{\ \beta}(y)\, A^{\ i}_\mu(x)\, A^{\ j}_\nu(x) & A^{\ i}_\mu(x) K_i^{\ \alpha}(y) \bar h_{\alpha\beta}(x,y) \\ \bar h_{\alpha\beta}(x,y)\, A^{\ i}_\nu(x) K_i^{\ \beta}(y) & \bar h_{\alpha\beta}(x,y) \end{pmatrix}, \tag{7.2}$$

when we have explicitly indicated the coordinate dependence of all quantities in order to emphasize that this is precisely the ansatz of Ref. 1, with the generalization of the x dependence of $\bar h$. It is easy to specialize to the case when $H = \{e\}$ (principal bundle); in this case $\mathfrak{P} = \mathfrak{G}$, the indices $\alpha, \beta, \cdots$ are replaced everywhere by the indices $i, j, \cdots$ and the quantities $K_i^{\ \alpha}$ reduce to $\delta_i^{\ j}$. The metric (7.2) then reduces to the usual "ansatz" for theories with spontaneous compactification. It should be observed that the use of $A^{\ i}_\mu(x)$ in (7.2) is compatible with the bundle structure only if the basis $\tilde e_i^{\rm L}$ is used; had we used the fundamental vector fields $\tilde e_i^{\rm R}$ as basis for the vertical directions, then the quantities $\omega_\mu^i(x,g)$ given by (6.13) and (6.15) would appear in (7.2) (as in Ref. 10).

VIII. DIMENSIONAL REDUCTION OF THE RICCI SCALAR

We start from the metric (7.1) in the horizontal lift basis $\{\bar e_\alpha, \bar e_\mu\}$, whose structure functions are given by (6.8), (6.30), and (6.31), and, using well-known formulae, we compute the Ricci scalar of its Riemannian connection. We give the intermediate steps of our computation. The connection coefficients are [the quantities d were introduced in Eqs. (6.30) and (6.32)]:

$$\begin{aligned}
&\bar\Gamma_{\beta\gamma}{}^{\alpha} = \Gamma_{\beta\gamma}{}^{\alpha}(x,y) = \tfrac12 \bar h^{\,\alpha\delta}(\bar h_{\gamma\epsilon} c_{\delta\beta}{}^{i} + \bar h_{\beta\epsilon} c_{\delta\gamma}{}^{i}) K_i^{\ \epsilon} - \tfrac12 c_{\beta\gamma}{}^{i} K_i^{\ \alpha}\,,\\
&\bar\Gamma_{\beta\mu}{}^{\alpha} = \tfrac12 \bar h^{\,\alpha\delta} \bar e_\mu \bar h_{\beta\delta} + \tfrac12 (\bar h^{\,\alpha\delta} \bar h_{\beta\gamma} d_{\delta\mu}{}^{\gamma} + d_{\beta\mu}{}^{\alpha})\,,\\
&\bar\Gamma_{\mu\beta}{}^{\alpha} = \tfrac12 \bar h^{\,\alpha\delta} \bar e_\mu \bar h_{\beta\delta} + \tfrac12 (\bar h^{\,\alpha\delta} \bar h_{\beta\gamma} d_{\delta\mu}{}^{\gamma} - d_{\beta\mu}{}^{\alpha})\,,\\
&\bar\Gamma_{\mu\nu}{}^{\alpha} = \tfrac12 d_{\mu\nu}{}^{\alpha}\,,\\
&\bar\Gamma_{\alpha\mu}{}^{\lambda} = \Gamma_{\mu\alpha}{}^{\lambda} = \tfrac12 g^{\lambda\rho} \bar h_{\alpha\beta} d_{\rho\mu}{}^{\beta}\,,\\
&\bar\Gamma_{\beta\gamma}{}^{\lambda} = -\tfrac12 g^{\lambda\rho} \bar e_\rho \bar h_{\beta\gamma} + \tfrac12 g^{\lambda\rho} (\bar h_{\beta\delta} d_{\rho\gamma}{}^{\delta} + \bar h_{\gamma\delta} d_{\rho\beta}{}^{\delta})\,,\\
&\bar\Gamma_{\mu\nu}{}^{\lambda} = \Gamma_{\mu\nu}{}^{\lambda}(x).
\end{aligned} \tag{8.1}$$

The partially contracted Ricci tensor is (here $\nabla_\mu V^\lambda = \bar e_\mu V^\lambda + \Gamma_{\mu\nu}{}^{\lambda} V^\nu$):

$$\begin{aligned}
\bar g^{\beta\delta} R_{\beta\delta} = {}& (R_{G/H} + \tfrac14 g^{\mu\nu} g^{\rho\sigma} \bar h_{\alpha\beta} d_{\mu\rho}{}^{\alpha} d_{\nu\sigma}{}^{\beta} - \tfrac12 \bar h^{\,\beta\delta} \nabla_\rho \nabla^\rho \bar h_{\beta\delta} + h^{\,\beta\delta} \nabla_\mu (g^{\mu\rho} \bar h_{\delta\epsilon} d_{\rho\beta}{}^{\epsilon})\\
& - \tfrac14 g^{\mu\nu} \bar h^{\,\alpha\beta} \bar h^{\,\gamma\delta} (\bar e_\mu \bar h_{\alpha\beta})(\bar e_\nu \bar h_{\gamma\delta}) + \tfrac12 g^{\mu\nu} \bar h^{\,\alpha\gamma} \bar h^{\,\beta\delta} (\bar e_\mu \bar h_{\alpha\beta})(\bar e_\nu \bar h_{\gamma\delta}) - g^{\mu\nu} \bar h^{\,\alpha\gamma} (\bar e_\mu \bar h_{\alpha\gamma}) d_{\beta\nu}{}^{\beta}\\
& + g^{\mu\nu} \bar h^{\,\beta\delta} (\bar e_\mu \bar h_{\delta\gamma}) d_{\beta\nu}{}^{\gamma} - g^{\mu\nu} d_{\alpha\mu}{}^{\alpha} d_{\beta\nu}{}^{\beta}\,,
\end{aligned} \tag{8.2}$$

$$\begin{aligned}
\bar g^{\mu\nu} R_{\mu\nu} = {}& R_M - \tfrac12 g^{\mu\nu} g^{\rho\sigma} \bar h_{\alpha\beta} d_{\mu\rho}{}^{\alpha} d_{\nu\sigma}{}^{\beta} - \nabla_\mu g^{\mu\nu} (\tfrac12 \bar h^{\,\alpha\beta} \bar e_\nu \bar h_{\alpha\beta} + d_{\alpha\nu}{}^{\alpha}) - \tfrac14 g^{\mu\nu} \bar h^{\,\alpha\gamma} \bar h^{\,\beta\delta} (\bar e_\mu \bar h_{\alpha\beta})(\bar e_\nu h_{\gamma\delta})\\
& - g^{\mu\nu} \bar h^{\,\beta\delta} (\bar e_\mu \bar h_{\delta\gamma}) d_{\beta\nu}{}^{\gamma} - \tfrac12 g^{\mu\nu} d_{\beta\mu}{}^{\alpha} d_{\alpha\nu}{}^{\beta} - \tfrac12 g^{\mu\nu} \bar h^{\,\alpha\delta} \bar h_{\beta\gamma} d_{\alpha\mu}{}^{\beta} d_{\delta\nu}{}^{\gamma}\,.
\end{aligned} \tag{8.3}$$

After some manipulations we finally obtain

$$\begin{aligned}
\bar R = {}& \bar g^{\beta\delta} \bar R_{\beta\gamma} + \bar g^{\mu\nu} \bar R_{\mu\nu}\\
= {}& R_M + R_{G/H} - \tfrac14 g^{\mu\nu} g^{\rho\sigma} \bar h_{\alpha\beta} d_{\mu\rho}{}^{\alpha} d_{\nu\sigma}{}^{\beta} - \bar e_\mu g^{\mu\nu} (\bar h^{\,\alpha\beta} \bar e_\nu \bar h_{\alpha\beta} + 2 d_{\alpha\nu}{}^{\alpha}) - \Gamma_{\mu\rho}{}^{\mu} g^{\mu\nu} (\bar h^{\,\alpha\beta} \bar e_\nu \bar h_{\alpha\beta} + 2 d_{\alpha\nu}{}^{\alpha})\\
& - \tfrac14 g^{\mu\nu} (\bar h^{\,\alpha\beta} \bar h^{\,\gamma\delta} + \bar h^{\,\alpha\gamma} \bar h^{\,\beta\delta}) (\bar e_\mu \bar h_{\alpha\beta})(\bar e_\nu \bar h_{\gamma\delta}) - g^{\mu\nu} \bar h^{\,\alpha\gamma} (\bar e_\mu \bar h_{\alpha\gamma}) d_{\beta\nu}{}^{\beta} - g^{\mu\nu} \bar h^{\,\beta\delta} (\bar e_\mu \bar h_{\delta\gamma}) d_{\beta\nu}{}^{\gamma}\\
& - g^{\mu\nu} d_{\alpha\mu}{}^{\alpha} d_{\beta\nu}{}^{\beta} - \tfrac12 g^{\mu\nu} (d_{\beta\mu}{}^{\alpha} d_{\alpha\nu}{}^{\beta} + \bar h^{\,\alpha\delta} \bar h_{\beta\gamma} d_{\alpha\mu}{}^{\beta} d_{\delta\nu}{}^{\gamma})\,.
\end{aligned} \tag{8.4}$$

So far we have not used the condition that the metric $\bar h$ be left invariant; this enters in (8.4) through formula (6.22):

$$\bar e_\mu \bar h_{\alpha\beta} = \bar\partial_\mu \bar h_{\alpha\beta} - A^{\ i}_\mu \bar e_i \bar h_{\alpha\beta}\,.$$

For a left-invariant metric [$\mathscr{L}$ = Lie derivative; use Eq. (6.8)]:

$$\bar e_i \bar h_{\alpha\beta} = \mathscr{L}_{e_i} (\bar h\,(\bar e_\alpha, \bar e_\beta)) = (\mathscr{L}_{e_i} \bar h\,)(\bar e_\alpha, \bar e_\beta) + \bar h\,(\mathscr{L}_{e_i} \bar e_\alpha, \bar e_\beta) + \bar h\,(e_\alpha, \mathscr{L}_{e_i} \bar e_\beta) = -C_{i\alpha}{}^{j} K_j^{\ \gamma} h_{\gamma\beta} - C_{i\beta}{}^{j} K_j^{\ \gamma} h_{\alpha\gamma}\,. \tag{8.5}$$

Introducing in (8.4), we find

$$\bar{R} = R_M + R_{G/H} - \tfrac{1}{4} g^{\mu\nu} g^{\rho\sigma} \bar{h}_{\alpha\beta} K^{\alpha}_{i} K^{\beta}_{j} F_{\mu\rho}{}^{i} F_{\nu\sigma}{}^{j}$$
$$- \bar{\nabla}_{\mu} (g^{\mu\nu} \bar{h}^{\alpha\beta} \bar{\partial}_{\nu} \bar{h}_{\alpha\beta})$$
$$- \tfrac{1}{4} g^{\mu\nu} (\bar{h}^{\alpha\beta} \bar{h}^{\gamma\delta} + \bar{h}^{\alpha\gamma} \bar{h}^{\beta\delta})(\bar{\partial}_{\mu} \bar{h}_{\alpha\beta})(\bar{\partial}_{\nu} \bar{h}_{\gamma\delta}), \qquad (8.6)$$

where $\bar{\nabla}_{\mu} v^{\lambda} = \bar{\partial}_{\mu} v^{\lambda} + \Gamma_{\mu\nu}{}^{\lambda} v^{\nu}$. The last step in the dimensional reduction is the integration over the fibers. To this end, we first need the y dependence of $\bar{h}_{\alpha\beta}$. This follows from (4.6) and (3.16)

$$\bar{h}_{\alpha\beta}(x,y) = \mathrm{Ad}(\sigma(y)^{-1})^{\gamma}_{\alpha}\ \mathrm{Ad}(\sigma(y)^{-1})^{\delta}_{\beta} \bar{h}_{\gamma\delta}(x,0) \qquad (8.7)$$

and $\bar{h}_{\alpha\beta}(x,0) = \gamma_{\alpha\beta}(x)$. The y dependence of $\bar{h}^{\alpha\beta}$ is by the matrix inverse to $\mathrm{Ad}(\sigma(y)^{-1})^{\gamma}_{\alpha}$, which in our basis is given by Eq. (3.20):

$$\bar{h}^{\alpha\beta}(x,y) = K_i{}^{\alpha} K_j{}^{\beta}\ \mathrm{Ad}(\sigma(y))^{i}_{\gamma}\ \mathrm{Ad}(\sigma(y))^{j}_{\delta} \bar{h}^{\gamma\delta}(x,0) \qquad (8.8)$$

and $\bar{h}^{\alpha\beta}(0) = \gamma^{\alpha\beta}(x)$. Since the y dependence is factored, the terms containing $\bar{\partial}_{\nu} \bar{h}_{\alpha\beta}$ have no dependence on y; thus we may replace $\bar{h}$ by γ there. The Ricci scalar $R_{G/H}$ is y independent because $\mathscr{L}_{e_i} R_{G/H} = 0\ \forall\ i$; since it is a function of γ, it will act as a kind of potential for the scalar fields. The term containing F^2 is y dependent through $\bar{h}_{\alpha\beta}$.

We now make some definitions. First, it is convenient to write $\gamma{:}M \rightarrow \mathrm{GL}^{+}(d,\mathbb{R})/\mathrm{SO}(d)$ as $\gamma = \delta \cdot \varphi$, where $\delta{:}M \rightarrow \mathbb{R}^{+}$ and $\varphi{:}M \rightarrow \mathrm{SL}(d,\mathbb{R})/\mathrm{SO}(d)$. In components $\gamma_{\alpha\beta}(x) = \delta(x) \cdot \varphi_{\alpha\beta}(x)$ with $(\delta(x))^{d} = \det(\gamma_{\alpha\beta}(x))$ and $\det(\varphi_{\alpha\beta}(x)) = 1$. Similarly, we may scale the metric $\bar{h}{:}\bar{h}_{\alpha\beta}(x,y) = \delta(x) \bar{k}_{\alpha\beta}(x,y)$, $\bar{k}_{\alpha\beta}$ being related to $\varphi_{\alpha\beta}$ through (4.6): $\bar{k}_{\alpha\beta}(x,0) = \varphi_{\alpha\beta}(x)$. Then

$$V(x) =: \int_{G/H} d^{d}y\ [\bar{h}(x,y)]^{1/2} = \delta^{d/2}(x) V_{G/H}, \qquad (8.9)$$

$$V_{G/H} =: \int_{G/H} d^{d}y\ [\bar{k}(x,y)]^{1/2}, \qquad (8.10)$$

$$\lambda_{ij}(x) =: \frac{1}{V_{G/H}} \int_{G/H} d^{d}y\ [\bar{k}(x,y)]^{1/2} \bar{k}_{\alpha\beta}(x,y) K^{\alpha}_{i}(y) K^{\beta}_{j}(y), \qquad (8.11)$$

λ_{ij} is the G average of $\bar{k}_{ij}(x,y)$ on G/H, analogous to (4.10). It is a function of $\varphi_{\alpha\beta}$, but not of δ.

A calculation similar to (4.9) shows that λ_{ij} is a (symmetric) Ad(G)-invariant tensor in $\mathfrak{G}$. If G is simple, the adjoint representation is irreducible and a basis in $\mathfrak{G}$ can be chosen so that the Cartan-Killing form is proportional to the Euclidean inner product. Therefore, applying Theorem 1, Appendix 5 of Ref. 13 (Vol. I), we find $\lambda_{ij} = c\delta_{ij}$ for some $c \neq 0$.

After some manipulations and discarding a total divergence, we obtain

$$\int_{E} d^{4}x\, d^{d}y\, \bar{g}^{1/2} \bar{R} = V_{G/H} \int_{M} d^{4}x \sqrt{g}\ \{\delta^{d/2} R_M$$
$$- \tfrac{1}{4} g^{\mu\nu} g^{\rho\sigma} \delta^{d/2+1} \lambda_{ij} F_{\mu\rho}{}^{i} F_{\nu\sigma}{}^{j}$$
$$- \tfrac{1}{4} g^{\mu\nu} \delta^{d/2} (\varphi^{\alpha\beta} \varphi^{\gamma\delta} + \varphi^{\alpha\gamma} \varphi^{\beta\delta})\, \partial_{\mu} \varphi_{\alpha\beta} \partial_{\nu} \varphi_{\gamma\delta}$$
$$- \tfrac{1}{2} \delta^{d/2-1} g^{\mu\nu}\, \partial_{\mu} \delta \varphi^{\alpha\beta} \partial_{\nu} \varphi_{\alpha\beta}$$
$$+ \tfrac{1}{4} d(d-1) g^{\mu\nu} \delta^{d/2-2} \partial_{\mu} \delta \partial_{\nu} \delta + W[\delta,\varphi]\}, \qquad (8.12)$$

where

$$W[\delta,\varphi] = -\delta^{d/2-1} \varphi^{\alpha\beta} [\tfrac{1}{2} C_{\alpha\gamma}{}^{\delta} C_{\beta\delta}{}^{\gamma} + C_{\alpha j}{}^{\gamma} C_{\beta\gamma}{}^{j}$$
$$+ C_{\alpha\gamma}{}^{\gamma} C_{\beta\delta}{}^{\delta} + \tfrac{1}{4} \varphi_{\gamma\delta} \varphi^{\epsilon\varphi} C_{\alpha\epsilon}{}^{\gamma} C_{\beta\varphi}{}^{\delta}]. \qquad (8.13)$$

IX. DISCUSSION OF THE FOUR-DIMENSIONAL THEORY

The Lagrangian (8.12) and (8.13) describes the low-energy behavior of a KK theory with homogeneous fibers. Its gauge invariance is evident from the fact that the scalars δ, $\varphi_{\alpha\beta}$ are singlets under G and from the Ad-invariance of λ. There are some surprising aspects in this Lagrangian. The first is the fact that a full G-Yang–Mills field becomes dynamically active, in spite of the fact that the fibers are not the full group; it might be expected that part of the information contained in the connection Γ in the principal bundle P gets lost when it is projected to the connection $\bar{\Gamma}$ in the associated bundle E by means of τ. This has been discussed, e.g. in Ref. 16. However, this is not so, and it can be proven that $\bar{\Gamma}$ contains exactly as much information as Γ if the group G acts effectively on the fibers.[17] If G does not act effectively on G/H, the kernel of the action (i.e., the subset of elements which leaves G/H fixed) is a normal subgroup G_0 of G and $G' = G/G_0$ acts effectively on G/H. If $H' = H/G_0$, then G/H can also be written G'/H'; in this case the gauge group which becomes dynamically active is G'.

The other surprising aspect is the role of the scalars. They have a complicated nonpolynomial Lagrangian and their interactions with the other fields lie only in the Brans–Dicke-like $\delta^{d/2}$ factor in the gravitational kinetic term and $\delta^{d/2+1}\lambda_{ij}$ in the Yang–Mills term. If the scalar degrees of freedom were frozen, then the Lagrangian would reduce to an Einstein–Yang–Mills one with a cosmological constant coming from W.

We have already mentioned in Sec. VII that the number of independent scalar fields has to be worked out case by case. In the case of a principal bundle, it is the maximal number $\frac{1}{2} n(n+1)$ $(n = \dim G)$ but when H is nontrivial, condition (4.5) has to be satisfied in addition. This can be written as $A_{(\alpha\beta)}{}^{(\gamma\delta)} \gamma_{\gamma\delta} = 0$ with $A_{(\alpha\beta)}{}^{(\gamma\delta)} = A_{(\beta\alpha)}{}^{(\gamma\delta)} = A_{(\alpha\beta)}{}^{(\delta\gamma)}$ and the bi-indices $(\alpha\beta)$ run from 1 to $\frac{1}{2} d(d+1)$; the number of independent scalars is then $\frac{1}{2} d(d+1) - \mathrm{rank}\, A$. In some cases it is possible to obtain more general information in other ways. For instance, in the case of the sphere $S^{n} = \mathrm{SO}(n+1)/\mathrm{SO}(n)$, the linear isotropy representation of $\mathrm{SO}(n)$ on $\mathfrak{P} \approx T_0(S^{n})$ is just the standard action of $\mathrm{SO}(n)$ on $\mathbb{R}^{n}$ and the only inner product which is invariant under this action is the Euclidean one. Hence for sphere bundles there will be only one scalar field, corresponding to scale transformations of the fibers. The same result holds for projective space $\mathbb{F}\,\mathbb{P}^{n}$, where $\mathbb{F}$ is either $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{H}$. The isotropy groups are, respectively, $\mathrm{SO}(n)$, $\mathrm{U}(n)$, $\mathrm{Sp}(n)$ and they act in the standard manner on $\mathfrak{P} \approx \mathbb{R}^{n}$, $\mathbb{C}^{n}$, $\mathbb{H}^{n}$; in each case there is a unique invariant inner product. In each case when there is a unique independent scalar, this can be assumed to be the scale factor δ and (8.12) simplifies: The third and fourth terms in curly brackets vanish and in the second term λ_{ij}

becomes fixed. In these cases the only interaction between the scalars, gravity, and the Yang–Mills fields is in the form of varying coupling constants, as in Brans–Dicke theory. Concerning the term W, it should be stressed that it cannot be simply interpreted as a potential, because the kinetic term for the scalars is nonpolynomial. For $S_d = \mathrm{SO}(d+1)/\mathrm{SO}(d)$, $\varphi_{\alpha\beta} = \delta_{\alpha\beta}$ and $C_{ijk} = \epsilon_{ijk}$, so $W[\delta] = \frac{1}{2}(d^4 - \frac{3}{2}d^3 - \frac{1}{2}d^2 + d)\,\delta^{d/2-1}$.

The most unexpected feature of the scalars is perhaps the fact that they behave like singlets under the group G. This is in contrast to all previous calculations, at least to our knowledge, but it is absolutely necessary if the theory is to be interpreted in terms of fiber bundles. It is a direct consequence of the fact that the Riemannian structure in the bundle space is required to be trivialization-independent. The relaxation of this condition would lead to the results of Refs. 4 and 11.

In the case when $H = \{e\}$, the bundle E reduces to the principal bundle P, and in this special case the existence of the right action $\tilde{R}$ of G on P and of the fundamental vector fields (6.1) and (6.2) allows a direct check of the singlet nature of the scalars. Namely, if we use the basis $\{\tilde{e}_i^{\mathrm{R}}, \tilde{e}_\mu\}$ and if $p = \psi(x,g) \in P$, we get from (4.2)

$$\begin{aligned}\tilde{g}_{ij}(p) = \tilde{g}_p(\tilde{e}_i^{\mathrm{R}}, \tilde{e}_j^{\mathrm{R}}) &= h(x)(\phi_{x*}^{-1}\tilde{e}_i^{\mathrm{R}}, \phi_{x*}^{-1}\tilde{e}_j^{\mathrm{R}})|_g \\ &= h(x)(e_i^{\mathrm{R}}, e_j^{\mathrm{R}})|_g = \gamma_{ij}(x)\end{aligned}$$

independently of g, because h is left-invariant and the vector fields e_i^{R} are left-invariant. Since the vector fields $\tilde{e}_i^{\mathrm{R}}$ are trivialization-independent, the result is itself trivialization-independent, and the γ_{ij} are singlets.

The dimensional reduction discussed in this paper is obtained by substituting the generalized KK ansatz into the action. This should be distinguished from the mechanism of spontaneous compactification,[18] whereby one assumes that the bundle structure discussed in this paper is realized as the ground state of some $(4+d)$-dimensional theory. Indeed, it is known that an Einstein–Yang–Mills theory in $(4+d)$ dimensions with the gauge group H will always admit such a bundle structure as a solution of the equations of motion provided G/H is symmetric.[19]

Finally, we wish to give a geometric basis to a statement made by Witten in Ref. 2. The group $G \times G$ acts on G transitively and effectively from the left by $(a,b)g = agb^{-1}$. The isotropy group of this action at e is ΔG, the diagonal in $G \times G$, i.e., the set of all pairs (a,a). Therefore, G can be thought of as a coset space $G = (G \times G)/\Delta G$. Given a principal G-bundle P, we may form the fiber product $P \underset{G}{\times} P$, which is a principal bundle with fiber $G \times G$, and then use the action above to form an associated bundle $E = (P \underset{G}{\times} P) \underset{(G \times G)}{\times} G$. Although this bundle has fiber G, it is not a principal bundle. We can then apply the main result of the present paper to conclude that a Yang–Mills field with group $G \times G$ will become dynamically active. This is a nice illustration of the fact that the nature of the gauge group does not depend only on the choice of the fiber, but also on the choice of group action on the fiber.

ACKNOWLEDGMENTS

The authors are grateful to J. Strathdee, Y. M. Cho, and E. Sezgin for several constructive conversations. We also wish to thank the referee for his careful reading. S.R.-D. would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

[1] J. F. Luciani, Nucl. Phys. B **135**, 111 (1978).

[2] E. Witten, Nucl. Phys. B **186**, 412 (1981).

[3] Abdus Salam and J. Strathdee, Ann. Phys. **141**, 316 (1982).

[4] J. Scherk and J. H. Schwarz, Nucl. Phys. B **153**, 61 (1979).

[5] J. A. Wheeler in *Relativity, Groups and Topology*, edited by B. deWitt and C. deWitt (Gordon and Breach, New York, 1964); S. Hawking, Nucl. Phys. B **144**, 349 (1978).

[6] If d extra dimensions are compactified to a scale comparable to Planck's length $l_{P,4} \sim (K_4)^{1/2}$ (K_4 is Newton's constant), then from (8.12) the gravitational constant in $(4+d)$ dimension has to be of the order $K_{4+d} \sim K_4(l_{P,4})^d \sim K_4^{(d+2)/2}$. Thus the length at which one expects a large fluctuation of the $(4+d)$-dimensional metric is

$$l_{P,4+d} \simeq (K_{4+d})^{1/(d+2)} \sim l_{P,4}.$$

[7] From the point of view of the dimensionally reduced theory this means that the classical Yang–Mills field A_μ^i in Eq. (7.2) is allowed to have nonvanishing topological charge.

[8] W. Thirring, Acta Phys. Aust. Suppl. **XIX**, 439 (1978).

[9] W. Drechsler and M. E. Mayer, *Fiber Bundle Techniques in Gauge Theories*, Lecture Notes in Physics, Vol. 67 (Springer-Verlag, Berlin, 1977); M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1978).

[10] Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); L. N. Chang, K. I. Macrae, and F. Mansouri, Phys. Rev. D **13**, 235 (1976); A. S. Schwarz, CMP **56**, 79 (1977); P. Forgacs and N. Manton, CMP **72**, 15 (1980); J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980); A. Trautman, Bull. Acad. Polon. Sci. **27**, 7 (1979) and Acta Phys. Austr. Suppl. **XXIII** (1981) and other quoted therein; M. E. Mayer, *Lecture Notes in Physics* (Springer-Verlag, Berlin, 19xx), Vo. 116, p. 291; Hadronic J. **4**, 108 (1980); Acta Phys. Aust. Suppl. **XXIII**, 477–490 (1981); R. Percacci, J. Math. Phys. **22**, 1892 (1981).

[11] Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975).

[12] The condition that G acts effectively will be important in the following because we will interpret gauge transformations as motions (actually isometries) in the fibers of E and G is isomorphic to a group of motions only if it acts effectively. If the action was not effective, two different elements of G could produce the same motion.

[13] S. Kobayashi and F. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963, Vol. 1, and 1969, Vol. II).

[14] Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis Manifolds and Physics* (North-Holland, Amsterdam 1982), rev. ed.

[15] N. Steenrod, *The Topology of Fibre Bundles* (Princeton U.P., Princeton, New Jersey, 1951).

[16] G. Domokos and S. Kovesi-Domokos, Nuovo Cimento A **44**, 318 (1978).

[17] This can be understood intuitively by means of the following argument, for which we are grateful to H. Urbantke. If G is compact, every G/H can be embedded equivariantly into a vector space V which carries a representation of G; G/H can then be imagined as an orbit of G in V. Choose a particular frame in V whose vectors have end points in G/H; all other frames which are obtained from the given one by the action of G will be called "G-frames" on G/H. If the action of G on G/H is effective, there will be a one–one correspondence between elements of G and G frames on G/H. Thus the principal bundle P can be thought of as the bundle of G frames of the bundle E. A connection $\bar{\Gamma}$ prescribes a parallelism in E, i.e., a way to move points from one fiber to another. But G frames are ordered collections of points in G/H and hence $\bar{\Gamma}$ also says how to parallelly propagate G frames from one fiber to another, i.e., how to move points from one fiber to another in P. Thus a connection Γ can be reconstructed in P.

[18] J. Scherk and J. H. Schwarz, Phys. Lett. B **57**, 463 (1975); E. Cremmer and J. Scherk, Nucl. Phys. B **103**, 393 (1976); B **108**, 409 (1976).

[19] S. Randjbar-Daemi and R. Percacci, "Spontaneous compactification of a $(4+d)$-dimensional Kaluza–Klein theory into $M_4 \times G/H$ for arbitrary G and H," Preprint IO/82/50, ICTP, Trieste, 1982.

Error estimates of solutions and mean of solutions of stochastic differential systems

G. S. Ladde[a] and M. Sambandham[b]
Department of Mathematics, University of Texas at Arlington, Arlington, Texas 76019

(Received 4 September 1981; accepted for publication 12 February 1982)

Stochastic differential equations are considered. Estimates in terms of statistical properties are given for the difference between the solutions and solutions of the mean of stochastic differential systems. For this purpose necessary theorems are developed and sufficient conditions are given to obtain error estimates. A few examples are worked out to demonstrate the usefulness of the results.

PACS numbers: 02.50.Ey

1. INTRODUCTION

Mathematical modeling of several real world problems lead us to differential equations. In formulating the mathematical model one can ignore the randomness in the system and obtain a deterministic model. Such a deterministic model of a dynamic system can be described by systems of deterministic differential equations. However, if one incorporates the inherent randomness of a system into the mathematical model, then the dynamic of the system will be described by a stochastic differential equation with random parameters. In general, the laws governing the random phenomena, and the corresponding parameters, are not precisely known. Therefore, one is interested in approximating a stochastic model by means of a deterministic model. Such an approximation will lead us to the study of estimating the error response corresponding to stochastic and deterministic models.

The objective of this paper is to estimate the error between the solution and the solution of the mean of a random differential equation. Very recently[1,2] problems of this nature with regard to roots of random polynomials have been investigated. Furthermore, certain relationships between the random eigenvalue problem and its corresponding mean problem have been discussed in Ref. 3. The present study provides a tool that verifies to what extent the deterministic mathematical model differs from the corresponding stochastic model. The mathematical conditions are given in terms of statistical properties of rate coefficients and the initial data of the system. In addition to this, we have obtained some sufficient conditions to guarantee the boundedness of solution with probability one (w.p. 1). The purpose of obtaining such a result is to develop suitable and more feasible conditions to estimate the difference between the solution of the mean and the solution. Several remarks and examples are given to indicate the usefulness of the result. Some related deterministic results can be found in Refs. 4–6.

We organize our article as follows. In Sec. 2 we prove a few auxiliary results. In Sec. 3 some results for the boundedness of the solutions with probability one are developed. Main theorems are proved in Sec. 4, and to illustrate the theory a few examples are worked out in Sec. 5.

2. GENERALIZED VARIATION OF CONSTANTS FORMULA

For our discussion we consider the initial value problem

$$y'(t,\omega) = F(t,y(t,\omega),\omega), \quad y(t_0,\omega) = y_0(\omega) \tag{2.1}$$

and the corresponding mean system of differential equations

$$m'(t) = f(t,m(t)), \quad m(t_0) = m_0 = E(y_0(\omega)), \tag{2.2}$$

where $f(t,z) = E[F(t,z,\omega)]$. From (2.1) and (2.2) we have

$$y'(t,\omega) = f(t,y(t,\omega)) + R(t,y(t,\omega),\omega), \quad y(t_0,\omega) = y_0(\omega), \tag{2.3}$$

where

$$R(t,y(t,\omega),\omega) = F(t,y(t,\omega),\omega) - E[F(t,y(t,\omega),\omega)].$$

In our presentation we will be using the following initial value problem also:

$$x' = f(t,x), \quad x(t_0,\omega) = y_0(\omega) = x_0(\omega), \tag{2.4}$$

where f is as defined in (2.2).

Hereafter the notations and definitions are adopted from Ref. 7. Without further mention we assume that all the inequalities and relations involving random quantities are valid with probability 1 (w.p. 1). Now we assume the following hypotheses. $(\mathrm{H}_1) R \in M[R_+ \times R^n, R[\Omega, R^n]]$ and R is almost sure sample continuous in x for fixed $t \in R_+$; $f \in C[R_+ \times R^n, R^n]$, f_x exists, and $f_x \in C[R_+ \times R^n, R^{n^2}]$. (H_2) The random function F in (2.1) satisfies suitable regularity conditions so that the initial value problem (2.1) has sample solution process existing for $t \geqslant t_0$.

The above conditions imply that $x(t,\omega) = x(t,t_0,x_0(\omega))$ is a unique solution of (2.4), and further that $x(t,\omega)$ is sample continuously differentiable with respect to (t_0,x_0).

Let $\Phi(t,t_0,x_0(\omega))$ be the fundamental solution of the variational system associated with (2.4). Further assume that (H_3) $V(t,x) \in C[R_+ \times R^n, R^m]$ and V_x exists and is continuous for $(t,x) \in R_+ \times R^n$.

We now state and prove a few lemmas.

Lemma 2.1: Let the hypotheses (H_1)–(H_3) be satisfied and $x(t,\omega) = x(t,t_0,x_0(\omega))$ be the sample solution of (2.4) and $y(t,\omega) = y(t,t_0,y_0(\omega),\omega)$ be the sample solution of (2.3) for $t \geqslant 0$. Then

---

[a] Research partially supported by U. S. Army Research Grants Nos. DAAG29-80-C-0060 and DAAG29-81-G-0008.

[b] Post-Doctoral Research Fellowship support by the Government of India, No. 6-21/79-Ns-5.

$$V(t,y(t,\omega)) = V(t,x(t,\omega)) + \int_{t_0}^{t} V_x(t,x(t,s,y(s,\omega))) \times \Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds. \qquad (2.5)$$

Proof: Let $x(t,s,y(s,\omega))$ be the sample solution processes of (2.4) through $(s,y(s))$, and $y(s,\omega) = y(s,t_0,y_0(\omega),\omega)$ be the sample solution processes of (2.3) through (t_0,y_0). From the hypotheses (H_1)–(H_3) and Ref. 7, Theorem 2.6.4, we obtain

$$\frac{dV}{ds}(t,x(t,s,y(s,\omega))) = V_x(t,x(t,s,y(s,w)))\left[\frac{d}{ds}x(t,s,y(s,\omega)) + \frac{d}{dy}x(t,s,y(s,\omega))\frac{d}{ds}y(s,\omega)\right]$$
$$= V_x(t,x(t,s,y(s,\omega)))\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega) \quad \text{w.p. } 1. \qquad (2.6)$$

Integrating (in the sample sense) both the sides from t_0 to t, and noting that $x(t,t,y(t,t_0,y_0(\omega),\omega)) = y(t,t_0,y_0(\omega),\omega)$, we obtain

$$V(t,y(t,\omega)) = V(t,x(t,\omega)) + \int_{t_0}^{t} V_s(t,s(t,s,y(s,\omega)))\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds.$$

This completes the proof of the lemma.

Some particular cases of Lemma 1.1 will be illustrated in the following remarks.

Remark 2.1: Let $V(t,x(t,\omega)) = x(t,\omega)$. Then in Lemma 2.1, (2.5) reduces to

$$y(t,\omega) = x(t,\omega) + \int_{t_0}^{t} \Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds.$$

This gives the result of Theorem 2.7.3.[7]

Remark 2.2: If $V(t,x(t,\omega)) = \|x(t,\omega)\|^2$ then in Lemma 2.1, (2.5) reduces to

$$\|y(t,\omega)\|^2 = \|x(t,\omega)\|^2 + 2\int_{t_0}^{t} x(t,s,y(s,\omega))\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds.$$

Now we prove another lemma which gives the expressions for the difference between the solutions of the unperturbed system (2.3) and the initial value problem (2.4).

Lemma 2.2: Suppose all the hypotheses of Lemma 2.1 hold. Then

$$V(t,y(t,\omega) - x(t,\omega)) = V(t,0) + \int_{t_0}^{t} V_x[t,x(t,s,y(s,\omega)) - x(t,s,x(s,\omega))]\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds.$$

Proof: By following the proof of Lemma 2.1 we have the relation

$$\frac{dV}{ds}(t,x(t,s,y(s,\omega)) - x(t,s,x(s,\omega))) = V_x[t,x(t,s,y(s,\omega)) - x(t,s,x(s,\omega))]\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega) \quad \text{w.p. } 1. \qquad (2.7)$$

Integrating (2.7) from t_0 to t, and noting that $x(t,t,y(t,t_0,y_0(\omega),\omega)) = y(t,t_0,y_0(\omega),\omega)$, we complete the proof of the lemma.

The next lemma gives the expression for the difference between the solution of the perturbed system (2.3) and the solution of the mean system of equations (2.2).

Lemma 2.3: Suppose that all the hypotheses of Lemma 2.1 are satisfied. Then

$$V(t,y(t,\omega) - m(t)) = V(t,x(t,\omega) - m(t)) + \int_{t_0}^{t} V_x[t,x(t,s,y(s,\omega)) - x(t,s,m(s))] \times \Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds, \qquad (2.8)$$

where $m(t) = m(t,t_0,m_0)$ is a solution of (2.2).

Proof: By noting the fact that $x(t,s,m(s)) = m(t)$, the proof of the lemma follows from Lemma 2.2.

A particular case of Lemma 2.3 is illustrated in the following remark, which will be very much useful in studying the statistical properties of solution processes.

Remark 2.3: If $V(t,x) = \|x\|^2$, then from (2.8) we obtain

$$\|y(t,w) - m(t)\|^2 = \|x(t,\omega) - m(t)\|^2 + \int_{t_0}^{t} [x(t,s,y(s,\omega)) - x(t,s,m(s))]\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)ds.$$

3. BOUNDEDNESS SOLUTIONS

In this section we consider the boundedness of the solution process w.p. 1 of (2.1). The obtained result will justify some of the assumption that will be made in the following sections in order to obtain estimates of the difference between the solution of the mean (2.2) and the solution of (2.1).

Assume that F in (2.1) satisfies assumption (H_2). Then the differential system (2.1) is said to be:

(SB_1) bounded w.p. 1 (or almost surely sample bounded) if for each $\alpha > 0, t_0 \in J$ there exists a positive function $\beta = \beta(t_0,\alpha)$ which is continuous in t_0 for each α such that the inequality

$$\|y_0(\omega)\| < \alpha, \quad \text{w.p. } 1$$

implies

$$\|y(t,\omega)\| < \beta, \quad \text{w.p. } 1, \ t > t_0,$$

(SB_2) uniformly bounded w.p. 1 if β in (SB_1) is independent of t_0.

Consider the following random comparison differential systems:

$$u'(t,\omega) = g(t,u(t,\omega),\omega), \quad u(t_0,\omega) = u_0(\omega), \tag{3.1}$$

where $g \in M[J \times R^m, R[\Omega, R^m]]$ is such that $g(t,u,\omega)$ satisfies the Caratheodory conditions in (t,u) w.p. 1 and $g(t,u,\omega)$ is quasimonotonic, nondecreasing in u for fixed t w.p. 1.

The differential system (3.1) is said to be: (SB_1^*) bounded w.p. 1 if given $\alpha > 0$, $t_0 \in J$ there exists a positive function $\beta = \beta(t_0,\alpha)$ such that

$$\sum_{i=1}^{m} u_{i0}(\omega) < \alpha, \quad \text{w.p. } 1$$

implies

$$\sum_{i=1}^{m} u_i(t,\omega) < \beta, \quad t > t_0, \quad \text{w.p. } 1\,;$$

(SB_2^*) uniformly bounded w.p. 1 if β in (SB_1^*) is independent of t_0.

We now prove the following theorem, which is a probabilistic version of Ref. 8, Theorem 3.13.1.

Theorem 3.1: Assume that

(i) $g \in M[J \times R^m, R[\Omega, R^m]]$ and $g(t,u,\omega)$ is sample continuous and quasimonotone nondecreasing in u for fixed $t \in J$,

(ii) $V \in C[J \times R^n, R[\Omega, R^m]]$ satisfies a local Lipschitz condition in y w.p. 1 and for $(t,y) \in J \times R^n$

$$D^{+}_{(2.1)} V(t,y(t,\omega)) < g(t, V(t,y(t,\omega)))\,,$$

(iii) for $(t,y) \in J \times R^n$, $V(t,0) \equiv 0$ and

$$\sum_{i=1}^{m} V_i(t,y(t,\omega)) > b(\|y\|)\,,$$

where $b \in K$ on the interval $0 < u < \infty$ and $b(u) \to \infty$ as $u \to \infty$. Then

$$(SB_1^*) \Rightarrow SB_1\,.$$

Proof: Let $\alpha > 0$ and $t_0 \in J$ be given and let $\|y_0(\omega)\| < \alpha$. From the hypothesis on $V(t,y(t,\omega))$, there exists a number $\alpha_1 = \alpha_1(t_0,\alpha)$ such that

$$\|y_0(\omega)\| < \alpha \text{ and } V(t_0,y_0(\omega)) < \alpha_1$$

are satisfied, simultaneously. From the equiboundedness of (3.1), we have

$$\sum_{i=1}^{m} r_i(t,t_0,u_0(\omega),\omega) < \beta_1, \quad t > t_0 \tag{3.2}$$

if

$$\sum_{i=1}^{m} u_{i0}(\omega) < \alpha_1\,.$$

Further as $b(u) \to \infty$ as $u \to \infty$ we can choose a $\beta = \beta(t_0,\alpha)$, such that

$$b(\beta) > \beta_1(t_0,\alpha)\,. \tag{3.3}$$

Now let

$$u_0(\omega) = V(t_0,y_0(\omega))\,.$$

Then from Ref. 7, Theorem 2.8.1, we obtain

$$V(t,y(t,\omega)) < r(t,t_0,u_0(\omega))\,, \quad t > t_0 \tag{3.4}$$

where $r(t,t_0,u_0(\omega),\omega)$ is the maximal solution of (3.1).

Suppose that there exists a solution $y(t,t_0,y_0(\omega))$ with $\|y_0(\omega)\| < \alpha, \Omega_1 \subset \Omega, P(\Omega_1) > 0$, and $t \geqslant t_0$ such that

$$\|y(t_1,t_0,y_0(\omega),\omega)\| = \beta, \quad \omega \in \Omega_1$$

and

$$\|y(t,t_0,y_0(\omega),\omega)\| < \beta, \quad t \in (t_0,t_1)\,.$$

Then from (iii), (3.2)–(3.4), we obtain

$$b(\beta) < \sum_{i=1}^{m} V_i(t,y(t_1,t_0,y_0(\omega),\omega))$$
$$< \sum_{i=1}^{m} r_i(t_1,t_0,u_0(\omega),\omega) < \beta_1(t_0,\alpha) < b(\beta)\,.$$

This establishes the fact that SB_1 holds.

4. MAIN RESULTS

To develop theorems on the estimations of solutions we assume the following:

$$(H_4) \| V_x[t,x(t,s,y(s,\omega)) - x(t,s,m(s))]$$
$$\times \Phi(t,s,y(s,\omega)) R(s,y(s,\omega),\omega)\|$$
$$< C(\|y(s,\omega) - m(s)\|) g(s,\omega),$$

where $C \in C[R_+,R_+]$ and nondecreasing on $R_+, g \in M[R_+, R[\Omega, R_+]]$ and is sample Lebesgue integrable.

(H_5) $b(\|x\|) < V(t,x)\| < a(\|x\|)$,

where $a,b \in C[R_+,R_+]$; a is differentiable on R_+; b^{-1} exists and it is nondecreasing and continuous on R_+.

Theorem 4.1: Suppose the hypotheses (H_1)–(H_5) are satisfied. Then

$$b(\|y(t,\omega) - m(t)\|) < H^{-1}\left[\int_{t_0}^{t} g(s,\omega)ds + H(N(t,\omega))\right], \quad \text{w.p.1}$$

where

$$N(t,\omega) = a(\|x(t,\omega) - m(t)\|),$$

$$\frac{dH(s)}{ds} = \frac{1}{h(s)}\,,$$

and

$$h(s) = C(b^{-1}(s)).$$

Moreover, if H^{-1} is a concave function, then

$$E[b(\|y(t,\omega) - m(t)\|)]$$
$$< H^{-1}\left[E\left(\int_{t_0}^{t} g(s,\omega)ds\right) + E(H(N(t,\omega)))\right].$$

Proof: From Lemma 2.3, (H_4), and (H_5) we obtain

$$b(\|y(t,\omega) - m(t)\|) < a(\|x(t,\omega) - m(t)\|)$$
$$+ \int_{t_0}^{t} C(\|y(s,\omega) - m(s)\|) g(s,\omega)ds. \tag{4.1}$$

Let

$$r(t,\omega) = \int_{t_0}^{t} C(\|y(s,\omega) - m(s)\|) g(s,\omega)ds.$$

Therefore

$$r'(t,\omega) = C(\|y(t,\omega) - m(t)\|) g(t,\omega). \tag{4.2}$$

From (4.1) we obtain

$$b(\|y(t,\omega) - m(t)\|) < N(t,\omega) + r(t,\omega). \tag{4.3}$$

From (4.2) and (4.3) we obtain

$$r'(t,\omega)\leqslant h\,(N(t,\omega)+r(t,\omega))g(t,\omega). \tag{4.4}$$

Let

$$N(t,\omega)+r(t,\omega)=u(t,\omega). \tag{4.5}$$

Differentiating (4.5) on both the sides we obtain

$$N'(t,\omega)+r'(t,\omega)=u'(t,\omega),$$

$$u(t_0,\omega)=N(t_0,\omega). \tag{4.6}$$

Using (4.6) in (4.4) we get

$$u'(t,\omega)\leqslant N'(t,\omega)+h\,(u(t,\omega))g(t,\omega). \tag{4.7}$$

Now we consider the following comparison differential equation:

$$v'(t,\omega)=N'(t,\omega)+h\,(v(t,\omega))g(t,\omega), \tag{4.8}$$

with

$$v(t_0,\omega)=u(t_0,\omega).$$

Since $h\,(\cdot)$ is nondecreasing and $g(t,\omega)$ is non-negative, by an application of Ref. 5, Corollary 2.1, we obtain

$$v(t,\omega)\leqslant H^{-1}\left[\int_{t_0}^{t} g(s,\omega)ds+H\,(N\,(t,\omega))\right]. \tag{4.9}$$

From (4.5)–(4.9) and an application of a comparison theorem (Ref. 7, Theorem 2.5.1), (4.7) reduces to

$$r(t,\omega)\leqslant -N(t,\omega)+H^{-1}\left[\int_{t_0}^{t} g(s,\omega)ds+H\,(N\,(t,\omega))\right]. \tag{4.10}$$

From (4.3) and (4.10) we obtain

$$b\,(\|y(t,\omega)-m(t\,)\|)H^{-1}\left[\int_{t_0}^{t} g(s,\omega)ds+H\,(N\,(t,\omega))\right]. \tag{4.11}$$

This proves the first part of the theorem.

If H^{-1} is concave, then taking expectation on both the sides of (4.11), and using Jensen's inequality,[7] we obtain

$$E\,(b\,(\|y(t,\omega)-m(t\,)\|))\leqslant H^{-1}\left[E\int_{t_0}^{t} g(s,\omega)ds+EH\,(N\,(t,\omega))\right].$$

This proves the second part of the theorem.

In the following corollary we prove another inequality which gives the estimates in terms of the initial conditions for (2.2) and (2.4).

Corollary 4.1: Suppose that all the hypotheses of Theorem 4.1 are satisfied and

$$\|\Phi\,(t,s,y(s,\omega))\|\leqslant K,$$

where K is a positive constant. Then

$$b\,(\|y(t,\omega)-m(t\,)\|)\leqslant H^{-1}\left[H\,(\overline{N}\,)(t_0,\omega))+\int_{t_0}^{t} g(s,\omega)ds\right], \tag{4.12}$$

where $\overline{N}\,(t_0,\omega)=a(K\,\|x_0(\omega)-m_0\|)$.

Moreover, if H^{-1} is concave then

$$E\,(b\,(\|y(t,\omega)-m(t\,)\|))$$

$$\leqslant H^{-1}\left[E\,(H\,(\overline{N}\,(t_0,\omega)))+E\left(\int_{t_0}^{t} g(s,\omega)ds\right)\right]. \tag{4.13}$$

Proof: We prove the first part of the corollary only. The second part follows from Theorem 4.1.

Application of Theorem 4.1 gives

$$b\,(\|y(t,\omega)-m(t\,)\|)\leqslant a(\|x(t,\omega)-m(t\,)\|)$$

$$+\int_{t_0}^{t} C\,(\|y(s,\omega)-m(s)\|)\cdot g(s,\omega)ds.$$

From Theorem 2.7.4,[7] we can derive

$$x(t,t_0,x_0(\omega))-x(t,t_0,m(t_0))$$

$$=\int_0^1 \Phi\,(t,t_0,m_0+s(x_0(\omega)-m_0))ds(x_0(\omega)-m_0).$$

Using this inequality and the hypotheses of the corollary, we obtain

$$\|x(t,t_0,x_0(\omega))-x(t,t_0,m_0)\|\leqslant K\,\|x_0(\omega)-m_0\|. \tag{4.14}$$

From (4.14), and an application of Theorem 4.1, the proof of the corollary is complete.

To illustrate the feasibility of assumption (H_4), and the fruitfulness of the above result, we present a particular example.

Illustration 4.1: Let

$$V(t,x)=\|x\|^2 \quad\text{and}\quad \|\Phi\,(t,s,y(s,\omega))\|\leqslant K. \tag{4.15}$$

Note that (H_4) is feasible provided that the solution processes of (2.1) are bounded w.p.1. This can be tested by using the boundedness results in Sec. 3. From (4.15) we obtain

$$\|V_x\,(t,x)\|\leqslant 2\|x\|. \tag{4.16}$$

From the boundedness assumption on the solution processes (2.1), together with (4.15) and (4.16), one can find g such that

$$\|V_x\,(t,x(t,s,y(s,\omega))-x(t,s,m(s)))\Phi\,(t,s,y(s,\omega))R\,(s,y,(s,\omega),\omega)\|$$

$$\leqslant 2\|x(t,s,y(s,\omega))-x(t,s,m(s))\|g(s,\omega), \tag{4.17}$$

where $g(s,\omega)\in M\,[R_+,R\,[\Omega,R\,]]$. Now by an application of Corollary 4.1 , we get

$$b\,(\|y(t,\omega)-m(t\,)\|)\leqslant H^{-1}\left[K\,\|x_0(\omega)-m_0\|+\int_{t_0}^{t} g(s,\omega)ds\right]$$

$$=\left[K\,(\|x_0(\omega)-m_0\|)+\int_{t_0}^{t} g(s,\omega)ds\right]^2, \tag{4.18}$$

since $H\,(s)=s^{1/2}$. By noting the fact that $b\,(u)=u^2=a(u)$, (4.18) reduces to the following form:

$$\|y(t,\omega)-m(t\,)\|^2\leqslant\left[K\,\|x_0(\omega)-m_0\|+\int_{t_0}^{t} g(s,\omega)ds\right]^2. \tag{4.19}$$

Taking expectation on both sides of (4.19) we get

$$E\,\|y(t,\omega)-m(t\,)\|^2\leqslant E\left[K\,(\|x_0(\omega)-m_0\|)+\int_{t_0}^{t} g(s,\omega)ds\right]^2$$

$$=E\left[K^2\|x_0(\omega)-m_0\|^2+2K\,\|x_0(\omega)-m_0\|\int_{t_0}^{t} g(s,\omega)ds\right.$$

$$\left.+\left(\int_{t_0}^{t} g(s,\omega)ds\right)^2\right]. \tag{4.20}$$

Remark 4.1: We remark that by selecting various conditions on the process $g(s,\omega)$ and $\|x_0(\omega)-m_0\|$ one can obtain more attractive estimates. For example:

(i) Taking the square root and then expectation on both the sides of (4.19) we obtain

$$E\,\|y(t,\omega)-m(t)\|\leqslant E\,[K\,\|x_0(\omega)-m_0\|\,]+E\left[\int_{t_0}^{t}g(s,\omega)ds\right]. \tag{4.21}$$

(ii) When $\|x_0(\omega)-m_0\|$ and $g(s,\omega)$ are independent random processes, from (4.20) we get

$$\begin{aligned}E\,\|y(t,\omega)-m(t)\|^2&\leqslant K^2E\,(\|x_0(\omega)-m_0\|^2\\&\quad+2KE\,(\|x_0(\omega)-m_0\|)E\left(\int_{t_0}^{t}g(s,\omega)ds\right)\\&\quad+E\left[\int_{t_0}^{t}\int_{t_0}^{t}(g(s,\omega)g(u,\omega))dsdu\right].\end{aligned} \tag{4.22}$$

Further, if $g(s,\omega)$ is any stationary Gaussian process and $Eg(s,\omega)$ is a constant (which we can take to be zero) then $E\,(g(s,\omega)g(u,\omega))$ depends only on $s-u$. Therefore, if $Eg(s,\omega)=0$ then in this case (4.22) reduces to

$$E\,\|y(t,\omega)-m(t)\|^2\leqslant K^2E\,(\|x_0-m_0\|)^2+\int_{t_0}^{t}\int_{t_0}^{t}c(u-s)duds,$$

where $E\,(g(s,\omega)g(u,\omega))=c(u-s)$.

So far we considered the right-hand side of (H_4) to be a product of two random functions. If the right-hand side of (H_4) is a product of a deterministic function and a random function we get another interesting inequality which we state in the following corollary.

Corollary 4.2: Suppose that all the hypotheses of Theorem 4.1 are satisfied; and instead of (H_4) we assume the following:

$$(\mathrm{H}_6)\|\,V_x(t,x(t,s,y(s,\omega))-x(t,s,m(s)))\Phi\,(t,s,y(s,\omega))R\,(s,y(s,\omega),\omega)\|\leqslant w(s)g(s,\omega),$$

where $w(s)\in C\,[R_+,R_+]$. Then

$$\begin{aligned}E\,(b\,(\|y(t,\omega)-m(t)\|))&\leqslant E\,[a(\|x(t,\omega)-m(t)\|)]\\&\quad+\left[\int_{t_0}^{t}w^2(s)ds\right]^{1/2}E\left[\int_{t_0}^{t}g^2(s,\omega)ds\right]^{1/2}.\end{aligned} \tag{4.23}$$

Moreover, if $\|\Phi\,(t,s,y(s,\omega))\|\leqslant K$, then

$$\begin{aligned}E\,(b\,\|y(t,\omega)-m(t)\|)&\leqslant E\,(a(K\,\|x_0(\omega)-m_0\|))\\&\quad+\left[\int_{t_0}^{t}w^2(s)ds\right]^{1/2}E\left[\int_{t_0}^{t}g^2(s,\omega)ds\right]^{1/2}.\end{aligned}$$

Proof: From (H_6), an application of the proof of Theorem 4.1, Corollary 4.1, and Holder's inequality, the proof of the corollary follows.

Now we present another illustration to exhibit the feasibility of (H_6), as well as the applications of Corollary 4.2.

Illustration 4.2: Let

$$V(t,x)=\|x\|\quad\text{and}\quad\|\Phi\,(t,s,y(s,\omega))\|\leqslant K. \tag{4.24}$$

Then

$$\|\,V_x(t,x)\|\leqslant 1.$$

Again by the boundedness assumption of the solution process of (2.1) it is easy to see that

$$\begin{aligned}&V_x(t,x(t,s,y(s,\omega))-x(t,s,m(s)))\Phi\,(t,s,y(s,\omega))R\,(s,y(s,\omega),\omega)\\&\quad\leqslant g(s,\omega),\end{aligned} \tag{4.25}$$

where $g\in M\,[R_+,R\,[\Omega,R\,]]$. Now take $b\,(u)=a(u)=u$. From (4.27), (4.28), and an application of Corollary 4.1 we have

$$\|y(t,\omega)-m(t)\|\leqslant K\,\|x_0(\omega)-m_0\|+\int_{t_0}^{t}g(s,\omega), \tag{4.26}$$

since $H\,(s)=s$. Taking expectation on both sides we get

$$E\,\|(t,\omega)-m(t)\|\leqslant KE\,(\|x_0(\omega)-m_0\|)+E\left(\int_{t_0}^{t}g(s,\omega)ds\right). \tag{4.27}$$

Remark 4.2: We remark that from (4.26) we can also get the mean square estimates. For example, squaring (4.26) and using the known inequality $(a+b)^2\leqslant 2(a^2+b^2)$ and taking expectation on both the sides we get

$$\begin{aligned}&E\,\|y(t,\omega)-m(t)\|^2\\&\quad\leqslant 2\left[K^2E\,\|x_0(\omega)-m_0\|^2+E\left(\int_{t_0}^{t}g(s,\omega)ds\right)^2\right].\end{aligned} \tag{4.28}$$

From the estimates (i) (4.21), (4.27) and (ii) (4.20), (4.22), (4.28) one can immediately note the role of different types of functions V.

In the next theorem we consider the right-hand side of (H_4) to be a function of $V(\cdot)$. Its immediate application is illustrated in Remark 4.3.

Theorem 4.2: Let the hypotheses (H_1)–(H_3) be satisfied. Further assume that

$$\begin{aligned}(\mathrm{H}_7)\,&V_x\,[t,x(t,s,y(s,\omega))-x(t,s,m(s))]\cdot\Phi\,(t,s,y(s,\omega))\\&\times R\,(s,y(s,\omega),\omega)\|\leqslant C\,[V(\,y(s,\omega)-m(s))]g(s,\omega).\end{aligned}$$

Then

$$V(t,y(t,\omega)-m(t))\leqslant H^{-1}\left[\int_{t_0}^{t}g(s,\omega)ds+H\,(M\,(t,\omega))\right],\quad\text{w.p.1} \tag{4.29}$$

where

$$M\,(t,\omega)=V(t,x(t,\omega)-m(t))\quad\text{and}\quad H\,(s)=\int\frac{ds}{C\,(s)}.$$

Moreover, if H^{-1} is a concave function then

$$\begin{aligned}&E\,[\,V(t,y(t,\omega)-m(t))]\\&\quad\leqslant H^{-1}\left[E\int_{t_0}^{t}g(s,\omega)ds+E\,\{H\,(M\,(t,\omega))\}\right].\end{aligned} \tag{4.30}$$

Proof: From Lemma 2.3 we obtain

$$\begin{aligned}V(t,y(t,\omega)-m(t))&=V(t,x(t,\omega)-m(t))\\&\quad+\int_{t_0}^{t}V_x\,[t,x(t,s,y(s,\omega))-x(t,s,m(s))]\\&\quad\times\Phi\,(t,s,y(s,\omega))R\,(s,y(s,\omega),\omega)ds\\&\leqslant V(t,x(t,\omega)-m(t))\\&\quad+\int_{t_0}^{t}C\,[\,V(t,y(s,\omega)-m(s))]g(s,\omega)ds.\end{aligned} \tag{4.31}$$

Let

$$r(t,\omega)=\int_{t_0}^{t}C\,[\,V(t,y(s,\omega)-m(s))]g(s,\omega)ds. \tag{4.32}$$

Differentiating (4.32) on both sides we get

$$r'(t,\omega)=C\,[\,V(t,y(t,\omega)-m(t))]g(t,\omega).$$

Now the proof of the theorem follows by applying the method used in Theorem 4.1.

Remark 4.3: Suppose $V(t,x) = \|x\|^p$. Assume that solution processes of (2.1) are bounded w.p.1. Then one can compute that

$$\begin{aligned}\|V_x(t,x(t,s,y(s,\omega)) - x(t,s,m(s)))\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)\| \\ \leqslant p\|x(t,s,y(s,\omega)) - x(t,s,m(s))\|^{p-1}g(s,\omega) \\ = C[V(t,x(t,s,y(s,\omega)) - x(t,s,m(s))]\,pg(s,\omega),\end{aligned} \qquad (4.33)$$

where $C(r) = r^{(p-1)/p}$. Application of Theorem 4.2 gives

$$\begin{aligned}\|y(t,\omega) - m(t)\|^p &\leqslant H^{-1}\left[p\int_{t_0}^t g(s,\omega)ds + p\|x(t,\omega) - m(t)\|\right] \\ &\leqslant \left[\int_{t_0}^t g(s,\omega)ds + \|x(t,\omega) - m(t)\|\right]^p.\end{aligned} \qquad (4.34)$$

Taking expectation on both sides of (4.34) we obtain

$$E\,\|y(t,\omega) - m(t)\|^p \leqslant E\left[\int_{t_0}^t g(s,\omega)ds + \|x(t,\omega) - m(t)\|\right]^p. \qquad (4.35)$$

Using the known inequality $(a+b)^p \leqslant 2^p(a^p + b^p)$, (4.35) reduces to

$$\begin{aligned}&E\,\|y(t,\omega) - m(t)\|^p \\ &\leqslant 2^p\left[E\left(\int_{t_0}^t g(s,\omega)ds\right)^p + E\,\|x(t,\omega) - m(t)\|^p\right].\end{aligned} \qquad (4.36)$$

5. EXAMPLES

In this section we present a few simple and illustrative examples.

Example 5.1: Consider the following differential equation:

$$y'(t,\omega) = a(t,\omega)y(t,\omega), \quad y(t_0,\omega) = y_0(\omega), \qquad (5.1)$$

where $a \in M[R,R[\Omega,R]]$ and satisfies enough regularity conditions for existence of a solution process for $t \geqslant t_0$:

$$y'(t,\omega) = E(a(t,\omega))y(t,\omega) + [a(t,\omega) - E(a(t,\omega))]\,y(t,\omega),$$

$$y(t_0,\omega) = y_0(\omega), \qquad (5.2)$$

$$m'(t) = E(a(t,\omega))m(t)), \quad m(t_0) = E(y_0(\omega)), \qquad (5.3)$$

$$x' = E(a(t,\omega))x, \quad x(t_0,\omega) = x_0(\omega), \qquad (5.4)$$

with $y_0(\omega) = x_0(\omega)$.

Solving (5.4) we get

$$x(t,\omega) = x_0(\omega)\exp\left[\int_{t_0}^t E(a(s,\omega))ds\right] \qquad (5.5)$$

and

$$\Phi(t,t_0,x_0(\omega)) = \exp\left[\int_{t_0}^t E(a(s,\omega))ds\right]. \qquad (5.6)$$

(i) Suppose that $V(t,x) = x$ and let $a(z) = b(z) = z$. Then we get

$$\begin{aligned}&|V_x(t,x(t,s,y(s,\omega)) \\ &\quad - x(t,s,m(s)))\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)| \leqslant g(s,\omega),\end{aligned} \qquad (5.7)$$

where

$$g(s,\omega) = |a(s,\omega) - Ea(s,\omega)|\,\exp\left[\int_s^t Ea(u,\omega)du\right].$$

From Theorem 4.1 we get

$$|y(t,\omega) - m(t)| \leqslant H^{-1}\left[\int_{t_0}^t g(s,\omega) + H(N(t,\omega))\right], \qquad (5.8)$$

where

$$H(x) = H^{-1}(z) = z,$$

$$N(t,\omega) = |x(t,\omega) - m(t)|.$$

From (5.5), (5.6) and an application of (5.7) and (5.8), we have

$$\begin{aligned}|y(t,\omega) - m(t)| &< |x_0(\omega) - m_0|\exp\left[\int_{t_0}^t Ea(s,\omega)ds\right] \\ &+ \int_{t_0}^t \left\{|a(s,\omega) - E(a(s,\omega))|\exp\left[\int_s^t Ea(u,\omega)du\right]ds\right\}.\end{aligned} \qquad (5.9)$$

Taking expectation on both sides we get

$$\begin{aligned}E\,|y(t,\omega) - m(t)| &\leqslant E(|x_0(\omega) - m_0|)\exp\left[\int_{t_0}^t Ea(s,\omega)\mathrm{ds}\right] \\ &+ E\int_{t_0}^t \left\{|a(s,\omega) - Ea(s,\omega)|\exp\int_s^t E(a(u,\omega))du\right\}ds.\end{aligned} \qquad (5.10)$$

If $|\Phi(t,s,y(s,\omega))| \leqslant K$, from (5.10) we obtain

$$\begin{aligned}&E\,|y(t,\omega) - m(t)| \\ &\leqslant K\left[E\,|x_0(\omega) - m_0| + E\int_{t_0}^t |a(s,\omega) - Ea(s,\omega)|ds\right].\end{aligned} \qquad (5.11)$$

(ii) Suppose $V(t,x) = x^2$ and let $a(z) = z^2 = b(z)$. Then from (4.17) we get

$$\begin{aligned}&|V_x(t,x(t,s,y(s,\omega)) - x(t,s,m(s)))\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)| \qquad (5.12) \\ &\leqslant 2|x(t,s,y(s,\omega)) - x(t,s,m(s))|g(s,\omega),\end{aligned}$$

where

$$g(s,\omega) = |a(s,\omega) - E(a(s,\omega))|\,\exp\int_s^t E(a(u,\omega))du.$$

From Theorem 4.1 we get

$$|y(t,\omega) - m(t)|^2 \leqslant \left[|x(t,\omega) - m(t)| + \int_{t_0}^t g(s,\omega)ds\right]^2. \qquad (5.13)$$

Taking expectation on both sides of (5.13) we obtain

$$E\,|y(t,\omega) - m(t)|^2 \leqslant E\left[|x(t,\omega) - m(t)| + \int_{t_0}^t g(s,\omega)ds\right]^2. \qquad (5.14)$$

If $|\Phi(t,s,y(s,\omega))| \leqslant K$, (5.14) reduces to

$$\begin{aligned}&E\,|y(t,\omega) - m(t)|^2 \\ &\leqslant E\left[K\,|x_0(\omega) - m_0| + \int_{t_0}^t |a(s,\omega) - E(a(s,\omega))|ds\right]^2.\end{aligned} \qquad (5.15)$$

If the random processes $|x_0(\omega) - m_0|$ and $|a(s,\omega) - Ea(s,\omega)|$ are uncorrelated and $|a(s,\omega) - Ea(s,\omega)|$ is a stationary Gaussian process then (5.15) reduces to

$$\begin{aligned}&E\,|y(t,\omega) - m(t)|^2 \\ &\leqslant \left[K^2E\,|x_0(\omega) - m_0|^2 + \int_{t_0}^t\int_{t_0}^t C(s-u)dsdu\right],\end{aligned}$$

where $C(s-u) = E\,|(a(s,\omega) - Ea(s,\omega))(a(u,\omega) - Ea(u,\omega))|$.

Note that from (5.13) and (5.14) we can get estimates similar to (5.11) and (4.21).

In the next example we consider a nonlinear equation.

Example 5.2: Consider the following differential equations.

$$y'(t,\omega) = -a(t,\omega)y^3(t,\omega), \quad y(t_0,\omega) = y_0(\omega), \tag{5.16}$$

where $a(t,\omega) \in M[R,R[\Omega,R_+]]$ and satisfies enough regularity conditions for the existence of a solution process for $t \geqslant t_0$.

$$y'(t,\omega) = -E(a(t,\omega))y^3(t,\omega) + [E(a(t,\omega)) - a(t,\omega)] \times y^3(t,\omega), \quad y(t_0,\omega) = y_0(\omega), \tag{5.17}$$

$$m'(t) = -E(a(t,\omega))m^3(t), \quad m(t_0) = E(x_0(\omega)), \tag{5.18}$$

$$x' = -E(a((t,\omega))x^3, \quad x(t_0,\omega) = x_0(\omega), \tag{5.19}$$

with $x_0(\omega) = y_0(\omega)$.

We assume that the solution of (5.16) is bounded w.p. 1. That is, we assume that $y^2(t,\omega) \leqslant \alpha^{-2}$ w.p. 1, $\alpha \neq 0$. From (5.19) we obtain

$$x(t,\omega) = \frac{x_0(\omega)}{\left[1 + 2x_0^2(\omega)\int_{t_0}^{t} Ea(s,\omega)ds\right]^{1/2}}. \tag{5.20}$$

From (5.20) we get

$$\Phi(t,t_0,x_0(\omega)) = 1\Big/\left[1 + 2x_0^2(\omega)\int_{t_0}^{t} a(s,\omega)\right]^{3/2}. \tag{5.21}$$

From (5.17) and (5.21) we get the following inequality:

$$|\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)| \leqslant |a(s,\omega) - Ea(s,\omega)|\Big/\left[\alpha^2 + 2\int_s^t EA(u,\omega)du\right]^{3/2}, \tag{5.22}$$

since $y^2 \leqslant \alpha^{-2}$, w.p. 1. Further, we note that $\alpha^2 > 0$ and $2\int_s^t a(u,\omega)du$ is positive. Therefore, (5.22) reduces to

$$|\Phi(t,s,y(s,\omega))R(s,y(s,\omega),\omega)| \leqslant K_1|a(s,\omega) - Ea(s,\omega)|, \tag{5.23}$$

where K_1 is a suitable positive constant.

Now we obtain a few inequalities for a different choice of $V(t,x)$.

(i) Let $b(z) = z = a(z)$ and

$$V(t,x) = x. \tag{5.24}$$

Equation (5.24) together with (5.22) verifies the assumption (H_6). Moreover, we note that $|\Phi(t,s,y(x,\omega))| \leqslant 1$ and $H(s) = s$. From (4.25), (4.27), (5.22), and (5.24) we get

$$E|y(t,\omega) - m(t)| \leqslant E|x_0(\omega) - m_0| + E\left\{\int_{t_0}^{t} \frac{|a(s,\omega) - Ea(s,\omega)|ds}{[\alpha^2 + 2\int_s^t Ea(u,\omega)du]^{3/2}}\right\}. \tag{5.25}$$

In the above discussion, if instead of (5.22) we use (5.23) we get

$$E|y(t,\omega) - m(t)| \leqslant E|x_0(\omega) - m_0| + E\left\{\int_{t_0}^{t} |a(s,\omega) - Ea(s,\omega)|ds\right\}. \tag{5.26}$$

(ii) Let $b(z) = z^2 = a(z)$ and

$$V(t,x) = x^2. \tag{5.27}$$

Equation (5.27) together with (5.22) verifies the assumption (H_4). Since $|\Phi(t,s,y(s,\omega))| \leqslant 1$ and $H(s) = s^{1/2}$, from (4.17), (4.20), (5.22), and (5.27) we obtain

$$E|y(t,\omega) - m(t)|^2 \leqslant E\left[|x_0(\omega) - m_0| + \int_{t_0}^{t} \frac{|a(s,\omega) - Ea(s,\omega)|du}{[\alpha^2 + 2\int_s^t Ea(u,\omega)du]^{3/2}}\right]^2. \tag{5.28}$$

In the above discussion if instead of (5.22) we consider (5.23), we get

$$E|y(t,\omega) - m(t)|^2 \leqslant E\left[|x_0(\omega) - m_0| + K_1\int_{t_0}^{t} |a(s,\omega) - Ea(s,\omega)|du\right]^2. \tag{5.29}$$

In particular, if the random processes $|x_0(\omega) - m_0|$ and $|a(s,\omega) - Ea(s,\omega)|$ follow certain additional conditions we get the following interesting inequalities.

(a) If $|x_0(\omega) - m_0|$ and $|a(s,\omega) - Ea(s,\omega)|$ are independent processes, then (5.29) reduces to

$$E|y(t,\omega) - m(t)|^2 \leqslant E[x_0(\omega) - m_0]^2 + K_1E|x_0(\omega) - m_0| \times E\left[\int_{t_0}^{t} (|a(s,\omega) - Ea(s,\omega)|)du\right] + K^2E\int_{t_0}^{t}\int_{t_0}^{t} (|a(s,\omega) - Ea(s,\omega)||a(u,\omega) - Ea(u,\omega)|) \times duds. \tag{5.30}$$

(b) If $|a(s,\omega) - Ea(s,\omega)|$ and $|x_0(\omega) - m_0|$ are uncorrelated and $|a(s,\omega) - Ea(s,\omega)|$ is a stationary Gaussian process, (5.29) reduces to

$$E|y(t,\omega) - m(t)|^2 \leqslant E(x_0(\omega) - m_0)^2 + K^2\int_{t_0}^{t}\int_{t_0}^{t} c(s-u)duds, \tag{5.31}$$

where $c(s-u) = E|a(s,\omega) - Ea(s,\omega)||a(s,\omega) - Ea(s,\omega)|$.

(iii) Let

$$V(t,x) = x^p. \tag{5.32}$$

The $V(t,x)$ in (5.32) together with (5.22) verifies (4.33) with $C(r) = r^{(p-1)/p}$. Application of (4.35) and (5.22) gives

$$E|y(t,\omega) - m(t)|^p \leqslant E\left[|x_0(\omega) - m_0| + \int_{t_0}^{t} \frac{|a(s,\omega) - Ea(s,\omega)|}{[\alpha^2 + 2\int_s^t Ea(u,\omega)du]^{3/2}}ds\right]^p. \tag{5.33}$$

In the above discussions, if in the place of (5.22) we use (5.23) we obtain

$$E|y(t,\omega) - m(t)|^p \leqslant E\left[|x_0(\omega) - m_0| + K_1\int_{t_0}^{t} |a(s,\omega) - Ea(s,\omega)|ds\right]^p. \tag{5.34}$$

Using the known inequality $(a+b)^p \leqslant 2^p(a^p + b^p)$, (5.34) can be further simplified to the following form

$$E|y(t,\omega) - m(t)|^p \leqslant 2^p\left[E|x_0(\omega) - m_0|^p + K_1^p\,E\left\{\int_{t_0}^{t} |a(s,\omega) - Ea(s,\omega)|ds\right\}^p\right]. \tag{5.35}$$

[1] M. J. Christensen and A. T. Bharucha-Reid. Stability of the roots of random algebraic polynomials. Comm. Statist. B Simulation Comput. **9** 179–192 (1980).

[2] G. S. Ladde and M. Sambandham "An estimate for the roots of random algebraic polynomials with application" (unpublished).

[3] A. T. Bharucha-Reid, *Random Integral Equations* (Academic, New York, 1973).

[4] G. S. Ladde, V. Lakshmikantham, and S. Leela, "A technique in perturbation theory," Rocky Mountain J. Math. **6** 133–140, (1976).

[5] G. S. Ladde, "Variational comparison theorem and perturbations of nonlinear systems," Proc. Am. Math. Soc. **52** 181–187, (1975).

[6] V. Lakshmikantham, "A variation of constant formula and Bellman–Gronwall–Reid inequalities," J. Math. Anal. Appl. **41** 199–204, (1973).

[7] G. S. Ladde and V. Lakshmikantham, *Random Differential Inequalities* (Academic, New York, 1980).

[8] V. Lakshmikantham and S. Leela, *Differential and Integral Inequalities* (Academic, New York, 1969).

Explicit suspensions of diffeomorphisms—An inverse problem in classical dynamics

Paul J. Channell
Los Alamos National Laboratory, AT-6, MS-H811, Los Alamos, New Mexico 87545

(Received 22 June 1982; accepted for publication 26 November 1982)

Presented in this paper is a set of explicit prescriptions for associating with a given map of R^n, which is C^2-isotopic to the identity, a time-dependent vector field whose time-1 map is the given one. Also shown is how to apply additional restrictions to the vector field including that it be (1) periodic in time, (2) Hamiltonian, and (3) of potential form; several examples show numerical verification of the theory.

PACS numbers: 03.20. + i, 02.90. + p

I. INTRODUCTION

In this paper I am concerned with an inverse problem in dynamics; namely, given a map, find a time-dependent dynamical system whose time-one map is the required map. Given a map on a manifold of dimension n, it is easy[1] to construct a vector field on a manifold of one higher dimension, called a suspension, which has the given map as a cross section. The standard construction[1] is not very useful in applications because the manifold is constructed as a quotient and can thus, in general, only be embedded in a $(2n+2)$-, or higher-, dimensional Euclidean space. As a result, an equivalent Euclidean vector field can be globally defined only in $2n+2$, or higher, dimensions for many cases; this high dimensionality, together with a lack of prescription for embedding, makes this construction impractical for applications.

In this paper I will restrict my attention to maps of finite-dimensional vector spaces. By doing so, I can construct, explicitly, the required vector fields. It is usually desired to have the vector fields satisfy certain additional constraints; for example, one may require the vector fields to have a periodic time dependence. For some applications it may be necessary for the vector fields to be in Hamiltonian form, assuming of course that the map is symplectic. In addition, it may be necessary for a Hamiltonian system to be in potential form. Additional requirements, of course, make the suspension problem harder, but in this theory it remains tractable.

In the next section, I present the general theory and consider the requirement that the vector fields be periodic in time. In Sec. III of this paper, I consider linear maps to elucidate the general theory, to point out its generality, and to present alternative conditions for periodicity. In Sec. IV, I consider symplectic maps with the requirement that the vector fields be Hamiltonian and of potential form. In the last section, I discuss my results and make general comments on the mathematical questions that this work raises.

II. GENERAL THEORY

Let us assume that we are given a C^2-diffeomorphism of Euclidean n-space R^n;

$$f\colon R^n \to R^n . \tag{1}$$

We want to construct a vector field

$$X\colon R^n \times R \to R^n , \tag{2}$$

so that the solutions of the vector differential equation

$$\frac{dx}{dt} = X(x,t) \tag{3}$$

satisfy

$$x(1) = f[x(0)] . \tag{4}$$

To construct X, let us first construct a function

$$\phi\colon R^n \times R \to R^m , \tag{5}$$

where, in general, m may or may not be equal to n. Let us require that ϕ be a constant on each trajectory of Eq. (3). This condition gives

$$\frac{\partial\phi}{\partial t} + \frac{dx}{dt} \cdot \nabla\phi = 0 , \tag{6}$$

where the gradient, or Jacobian, is with respect to the first n arguments of ϕ and where the dot indicates the dot product. At a fixed (x,t), $\nabla\phi(x,t)$ is a linear map

$$\nabla\phi(x,t)\colon R^n \to R^m . \tag{7}$$

This map will have a unique inverse only if $m = n$ or unless additional conditions are applied (as in Sec. III). Let us, however, consider a generalized (nonunique) inverse $M(x,t)$, so that

$$\nabla\phi(x,t) \cdot M(x,t) = \mathrm{Id}_{R^n} . \tag{8}$$

We then find that

$$\frac{dx}{dt} = -\frac{\partial\phi}{\partial t} \cdot M . \tag{9}$$

We must now apply the condition that the solutions of Eq. (9) satisfy Eq. (4). Let us denote $\phi(x,1)$ by $\phi_1(x)$. Our condition then becomes

$$\phi[x(0),0] = \phi_1\{f[x(0)]\} . \tag{10}$$

Let us now restrict our attention to maps, f, C^2-isotopic to the identity. Such an isotopy is a map

$$F\colon R^n \times R \to R^n , \tag{11}$$

so that

$$F(x,0) = f(x) \tag{12}$$

and

$$F(x,1) = \mathrm{Id}_{R^n} , \tag{13}$$

0022-2488/83/040823-05$02.50

and so that F is a C^2-diffeomorphism for each t. Given ϕ_1 and F we can define ϕ by

$$\phi(x,t) = \phi_1[F(x,t)] . \tag{14}$$

This definition of ϕ satisfies Eq. (10). Thus, we have reduced our problem to that of finding a C^2-isotopy F. The most general form for such an isotopy in vector spaces is

$$F(x,t) = H_1(t) \cdot f(x) + H_2(t) \cdot x + H_3(x,t) , \tag{15}$$

where

$$H_i(t)\colon\ R^n \to R^n, \quad i = 1,2,$$
$$H_3\colon\ R^n \times R \to R^n \tag{16}$$

and where

$$\begin{aligned} &H_1(0) = \mathrm{Id}_{R^n}, \quad H_1(1) = 0, \quad H_2(0) = 0, \\ &H_2(1) = \mathrm{Id}_{R^n}, \quad H_3(x,0) = H_3(x,1) = 0 . \end{aligned} \tag{17}$$

Of course, the H_i's must all be C^2. The map H_3 can be set equal to zero unless there are additional restrictions such as the ones we consider subsequently.

Let us now make a few comments on nonuniqueness. First, although we have required that Eq. (4) be a solution of Eq. (10), it need not be unique, especially if $m < n$. This nonuniqueness is related to the nonuniqueness of the map M in Eq. (8). Thus, for $m < n$, additional conditions must be applied to ensure Eq. (4). We will see subsequently that this is sometimes possible. A second source of nonuniqueness lies in the map ϕ_1, which does not enter in any vital way. The final source of nonuniqueness lies in the functions H_i, which are determined only at $t = 0,1$. In fact, our problem cannot have a unique solution and the nonuniqueness in the H_i's is useful because in practical problems it is usually necessary to apply realizability constraints in addition to those already mentioned. Variations in the H_i's can allow one to satisfy these additional conditions.

We have now given a procedure for constructing $X(x,t)$. Because X is C^1, our differential equation has unique solutions; sometimes it may be necessary to require that

$$X(x,t+1) = X(x,t) . \tag{18}$$

To do this, we simply define

$$\{t\} = \text{fractional part of } t \tag{19}$$

and let

$$\widetilde{X}(x,t) \equiv X(x, \{t\}) . \tag{20}$$

The vector field $\widetilde{X}$ is now continuous and periodic and is C^1 everywhere except at $t = n$, where n is any integer. If, however, we require that

$$\frac{\partial H_1(0)}{\partial t} = \frac{\partial H_1(1)}{\partial t} = \frac{\partial H_2(0)}{\partial t} = 0 , \tag{21}$$

$$\frac{\partial H_3(x,0)}{\partial t} = \frac{\partial H_3(x,1)}{\partial t} = 0 , \tag{22}$$

then it is easy to check that $\widetilde{X}$ is Lipschitz at $t = n$ and, thus, that our equation still has unique solutions.

Thus, Eqs. (20)–(22) suffice to define a periodic vector field. We will see in the next section that these conditions are not necessary and that periodicity can result in other ways.

Lest our previous considerations seem too abstract, let us end this section with a particular example of an isotopy that satisfies Eqs. (17), (21), and (22).

$$F(x,t) = \tfrac{1}{2}(1 + \cos \pi t) f(x) + \tfrac{1}{2}(1 - \cos \pi t)\, x . \tag{23}$$

III. LINEAR MAPS

In this section, as an illustration of the general theory, let us consider linear maps; that is, we assume

$$f(x) = A \cdot x , \tag{24}$$

where A is an $n \times n$ matrix.

Let us first consider the case $m = n$. We can then assume, though it is not required, that

$$\phi(x,t) = \Phi(t) \cdot x , \tag{25}$$

where $\Phi(t)$ is an $n \times n$ matrix. The differential equation (9) then becomes

$$\frac{dx}{dt} = -\Phi(t)^{-1} \cdot \dot{\Phi}(t) \cdot x ; \tag{26}$$

that is, it is a linear equation. The dot indicates a time derivative. It is clear that the function $\Phi(t)$ must be nonsingular for all t.

Let us now observe that the conditions given in Sec. II for periodicity were sufficient, but not necessary. If we require that Eq. (26) be periodic, then

$$\dot{\Phi}(t) = \Phi(t) \cdot G(t) , \tag{27}$$

where $G(t)$ is some periodic $n \times n$ matrix. By the Floquet theorem[2] the general solution of Eq. (27) is

$$\Phi(t) = e^{Bt} \Psi(t) , \tag{28}$$

where Ψ is a periodic $n \times n$ matrix and where B is a constant matrix. Conversely, any Φ of the form of Eq. (28) will generate a periodic differential equation, but will, in general, not satisfy Eqs. (21) and (22). This is fortunate since it turns out that Eqs. (21) and (22) are usually inconsistent with the additional requirement that a Hamiltonian system be of potential form.

Let us now consider the case $m = 1$. We now assume, though again it is not required, that

$$\Phi(x,t) = \tfrac{1}{2}[x \cdot \Phi(t) \cdot x] , \tag{29}$$

where $\Phi(t)$ is again an $n \times n$ matrix that can be chosen to be symmetric.

If we additionally require that the differential equation be linear, that is,

$$X(x,t) = Y(t) \cdot x , \tag{30}$$

then Eq. (6) becomes

$$\tfrac{1}{2}(x \cdot \dot{\Phi} \cdot x) + [Y(t) \cdot x] \cdot [\Phi(t) \cdot x] = 0 . \tag{31}$$

Because Eq. (31) must hold for all x, we find

$$\tfrac{1}{2}\dot{\Phi} + (\Phi \cdot Y)_s = 0 , \tag{32}$$

where the subscript s denotes the symmetric part of the product matrix.

We note that by assuming both a quadratic invariant and a linear differential equation we have obtained a unique specification of the differential equation in the case $m < n$.

Note that in the case $m = n$, that is, Eq. (25), the matrix Φ is simply related to the transformation matrix A by

$$\Phi(0) = A\,. \tag{33}$$

In the case $m = 1$, however, the relation between Φ and A is

$$A^T \cdot \Phi(1) \cdot A = \Phi(0)\,, \tag{34}$$

where superscript T denotes the transpose.

IV. HAMILTONIAN SYSTEMS

Let us now assume that the space is even dimensional, $n = 2l$, and that R^{2l} is equipped with the standard[3] symplectic 2-form. We denote a point of R^{2l} by $(q_1,\ldots, q_l,\ p_1,\ldots, p_l)$.

If we define

$$J = \begin{pmatrix} 0 & \mathrm{Id}_{R^l} \\ -\mathrm{Id}_{R^l} & 0 \end{pmatrix}, \tag{35}$$

then f is said to be symplectic[4] if

$$[f'(x)]^T J f'(x) = J\,, \tag{36}$$

where $f'(x)$ is the Jacobian matrix of f at x. A vector field is said to be Hamiltonian[5] if there exists a function

$$H\colon\ R^{2l} \times R \to R\,, \tag{37}$$

such that

$$\frac{dx}{dt} = J \cdot \nabla H(x,t)\,, \tag{38}$$

where ∇ denotes the derivatives with respect to the first $2l$ arguments, that is, the q's and p's.

If we henceforth assume that f is symplectic, then it is natural to require that the generating vector field be Hamiltonian. Combining Eqs. (9) and (38), we find

$$\nabla H = J \cdot M \cdot \frac{\partial \phi}{\partial t}\,. \tag{39}$$

This substantially restricts the possible isotopies in Eq. (15) because relatively few vector fields are gradients of a function. A further restriction results if we require that

$$H = \tfrac{1}{2} p^2 + V(q,t)\,, \tag{40}$$

that is, that it be of potential form.

To illustrate these ideas and to develop some useful examples, let us restrict our attention to linear Hamiltonian systems of potential form with $l = 1$. The Hamiltonian can then be written as

$$H = \tfrac{1}{2} p^2 + \tfrac{1}{2} K(t) q^2\,. \tag{41}$$

With this set of restrictions our problem reduces to that of finding the single function $K(t)$.

Our map is of the form of Eq. (24) and the condition of Eq. (36) becomes simply

$$\det(A) = 1\,, \tag{42}$$

that is, area preservation. The cases $m = 2$ and $m = 1$ are quite different, and we consider each in turn.

Let us first consider $m = 1$. Equation (32) is then the appropriate one, and, when combined with Eqs. (38) and (41), we find

$$\frac{d}{dt}\begin{pmatrix} a & b \\ b & c \end{pmatrix} = -\begin{pmatrix} -2bk & a - ck \\ a - ck & 2b \end{pmatrix}, \tag{43}$$

where we have defined

$$\Phi \equiv \begin{pmatrix} a & b \\ b & c \end{pmatrix}. \tag{44}$$

Equation (43) is the same as the three differential equations

$$\dot a = 2bK(t)\,, \tag{45}$$

$$\dot b = K(t)c - a\,, \tag{46}$$

$$\dot c = -2b\,. \tag{47}$$

The function $K(t)$ is easily calculated once a, b, c are known, but, because $K(t)$ appears in both Eqs. (45) and (46), there is a consistency condition that can be written as

$$\tfrac{1}{2} \dot c \ddot c = \dot c a + \dot a c\,. \tag{48}$$

Equation (48) has an immediate first integral

$$\tfrac{1}{4} \dot c^2 = ca + D\,, \tag{49}$$

where D is a constant. We now observe that we have only one arbitrary function at our disposal, which we take to be $c(t)$. Given $c(t)$, we can calculate $b(t)$ from Eq. (47) and $a(t)$ is then easily calculated from either Eq. (46) or Eq. (49).

In many applications, especially related to particle accelerators, it is more convenient to work directly with the phase-space ellipse parameters, that is, the matrix Φ, rather than with the transformation A. If so, then there is no need to solve Eq. (34). Thus, the function $c(t)$ must be chosen so that $a(t)$, $b(t)$, and $c(t)$ assume the desired values at $t = 0$ and $t = 1$. Note that $\Phi(0)$ and $\Phi(1)$ are not completely independent because area is preserved and so we have five conditions to apply. Because we have the constant D in Eq. (49) at our disposal, the function $c(t)$ must contain at least four arbitrary constants to satisfy all the conditions.

Let us consider a particular example. Let us assume that

$$c(t) = c_1 + c_2 \sin \pi t + c_3 \cos \pi t + c_4 \sin 2\pi t\,, \tag{50}$$

where the c_i's are constants. It is then easy to see that

$$c_1 = [c(0) + c(1)]/2\,, \tag{51}$$

$$c_2 = [b(1) - b(0)]/\pi\,, \tag{52}$$

$$c_3 = [c(0) - c(1)]/2\,, \tag{53}$$

$$c_4 = [-b(0) - b(1)]/2\pi\,, \tag{54}$$

$$D = a(0)c(0) - b(0)^2\,. \tag{55}$$

In Fig. 1(a) we show an example, for the choice of $c(t)$ in Eq. (50) of initial and final ellipses; that is, $\phi(x,0) = \text{const} = \phi(x,1)$, and a single trajectory which started on the initial ellipse and obeyed the Hamiltonian equations with $K(t)$ as prescribed above. As can be seen, the trajectory does arrive at the correct final ellipse. In Fig. 1(b) we show the $K(t)$ that produces the trajectory of Fig. 1(a).

Though only four constants in $c(t)$ were required to satisfy initial and final conditions, it may be convenient to choose a $c(t)$ with more arbitrary constants if additional constraints are to be applied.

Let us now consider the case $m = 2$. Equation (26) is now the relevant one, and, combining it with Eqs. (38) and (41), we find

$$\frac{d}{dt}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} bK & -a \\ dK & -c \end{pmatrix}, \tag{56}$$

where we have now defined

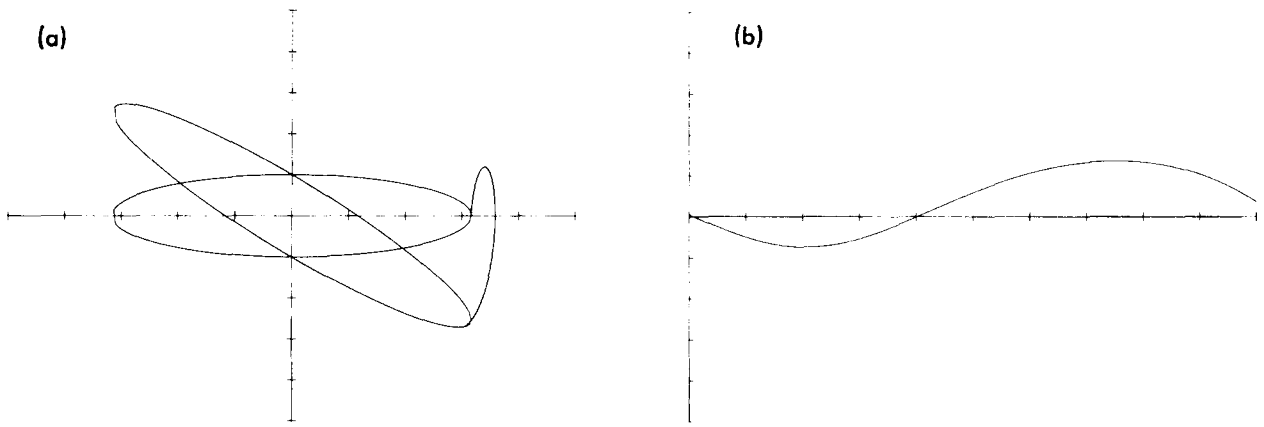


FIG. 1. (a) The initial and final ellipses and typical trajectory from one to the other for Eqs. (45)–(47). (b) The function $K(t)$, which produced the trajectory in Fig. 1(a).

$$\Phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{57}$$

Equation (56) is a set of four first-order differential equations, and we notice that the top and bottom rows decouple. It is then easy to obtain the two equations

$$\ddot{b} + K(t)b = 0\,, \tag{58}$$

$$\ddot{d} + K(t)d = 0\,. \tag{59}$$

Again it turns out that we have a single arbitrary function, which we take to be $d(t)$. Given $d(t)$, $K(t)$ follows immediately from Eq. (59) and $c(t)$ is simply

$$c(t) = -\dot{d}(t)\,. \tag{60}$$

The function $b(t)$ is then determined as the solution of Eq. (58) and $a(t)$ is given by

$$a(t) = -\dot{b}(t)\,. \tag{61}$$

Note that $b(t)$ and $(d(t)$ satisfy the same second-order equation, and thus, if $d(t)$ is given, $b(t)$ can be immediately written as

$$b(t) = \alpha d(t) \int^t \frac{dt'}{d(t')^2} + \beta d(t)\,, \tag{62}$$

where α and β are constants.

The function $d(t)$ must have enough arbitrary constants to fit the initial and final conditons, which in this case are

$$\Phi(0) = A\,, \tag{63}$$

$$\Phi(1) = \mathrm{Id}\,. \tag{64}$$

It is, of course, convenient to choose $d(t)$ in such a form that the indefinite integral in Eq. (62) can be done. As a specific example, let us choose

$$d(t) = (d_0 + d_1 \sin \pi t + d_2 \cos \pi t + d_3 \sin 2\pi t + d_4 \times \cos 2\pi t + d_5 \sin 3\pi t)^{-1/2}\,. \tag{65}$$

If we use the notation

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{66}$$

then the constants d_i, α, β can be determined from Eqs. (63) and (64) to be

$$\begin{aligned} d_0 &= -A_{12}/A_{22} + A_{21}/2\pi^2 A_{22}^3\,, \\ d_1 &= \tfrac{9}{8}(\pi A_{12}/A_{22} - 13A_{21}/36\pi A_{22}^3)\,, \\ d_2 &= \tfrac{1}{2}(1/A_{22}^2 - 1), \quad d_3 = A_{21}/2\pi A_{22}^3\,, \\ d_4 &= \tfrac{1}{2}(1/A_{22}^2 + 1) - d_0\,, \\ d_5 &= A_{21}/3\pi A_{22}^3 - d_1/3, \\ \alpha &= 1, \quad \beta = 0\,. \end{aligned} \tag{67}$$

Because of the negative power in Eq. (65), this choice of $d(t)$ tends to result in functions $K(t)$ that can take on large values. A typical phase trajectory and $K(t)$ for this example are shown in Figs. 2(a) and 2(b).

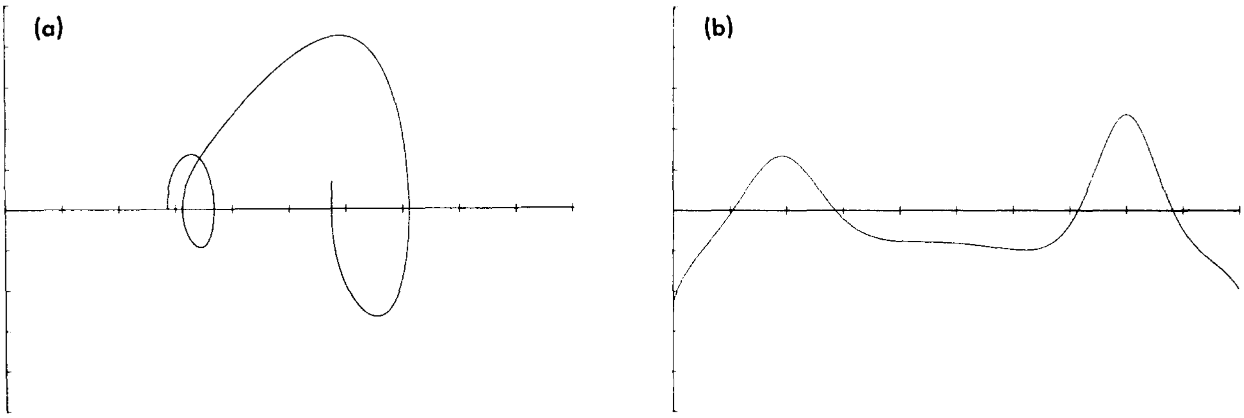


FIG. 2. (a) A typical trajectory for $K(t)$ given by Eq. (59). (b) The corresponding $K(t)$.

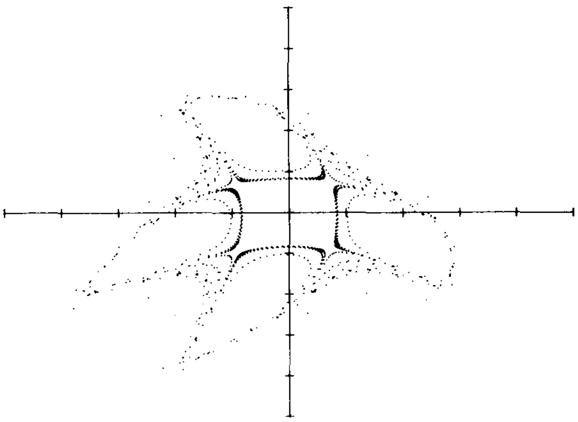


FIG. 3. 1100 iterations of a single initial condition with the map, Eq. (68). The parameters were $\alpha = 1.5$, $\theta = 1.6$, $x_0 = 0.4$, $y_0 = 0.35$.

V. MATHEMATICAL DISCUSSION

The examples in the previous section were linear. To give an example that is nonlinear we consider the symplectic mapping

$$\begin{aligned} x_1 &= (\cos\theta)x + (\sin\theta)y, \\ y_1 &= -(\sin\theta)x + (\cos\theta)y + \tfrac{1}{2}\alpha x_1^2, \end{aligned} \tag{68}$$

where θ and α are constants. If $\alpha = 0$, this map is simply a rotation by the angle θ. In Fig. 3 we show the result of 1100 iterations of Eq. (68) starting from a single initial point. The map is clearly nonintegrable.

We tested our theory in the case $m = 2$ using Eq. (9) with ϕ equal to the $F(x,t)$ in Eq. (23). Though the map, Eq. (68), is not a diffeomorphism except near the origin, our theory gave an equation that integrated to give the correct map for the region shown in Fig. 3.

This example raises a difficulty in that we have, by defining

$$\tilde{\phi}(x,t) = \phi(x, \{t\}), \tag{69}$$

as in Sec. II, a seeming constant for a system that is clearly nonintegrable. The resolution of this difficulty lies in a subtle mathematical point that we glossed over for pedagogical reasons in Sec. II. With the substitution $t \rightarrow \{t\}$, and, with the choice of Eq. (23), we do indeed satisfy Eqs. (21) and (22) and produce a periodic differential equation. However, we note that ϕ is not continuous, nor even defined at $t =$ integer, because

$$\phi(x, 1-\epsilon) \simeq x, \quad \phi(x, \{1+\epsilon\}) \simeq f(x). \tag{70}$$

Thus, $\partial\phi/\partial t$ is not defined at $t =$ integer, and the differential equation is not defined there. However, if we require Eqs. (21) and (22), we find

$$\frac{dx}{dt}(1-\epsilon) \simeq 0, \quad \frac{dx}{dt}(\{1+\epsilon\}) \simeq 0, \tag{71}$$

and thus we can define

$$\frac{dx}{dt}(t = \text{integer}) = 0, \tag{72}$$

and not only have a continuous equation but one that is Lipschitz as well. The function $\tilde{\phi}$, however, is no longer constant as t crosses integer values. Thus, we have not, in fact, constructed constants of the motion, but only an equation that produces the required map.

The general problem of uniquely specifying M in Eq. (8) for nonlinear systems for $m < n$ has not been solved and may not have a solution in many cases. It is not even known what conditions on the map and on M would suffice to insure a unique M in the general case.

In summary, we have given a set of prescriptions for explicit suspensions of diffeomorphisms of vector spaces C^2-isotopic to the identity. The theory applies to nonlinear systems and to systems with the additional constraints of periodicity, Hamiltonian form, and potential form. The nonuniqueness of the solutions provides sufficient flexibility to allow one to impose additional realizability constraints in practical applications.

---

[1] Z. Nitecki, *Differentiable Dynamics* (MIT Press, Cambridge, MA, 1971), p. 6.

[2] R. Bellman, *Stability Theory of Differential Equations* (Dover, New York, 1969), p. 28.

[3] R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Benjamin/Cummings, New York, 1978), 2nd ed., p. 176.

[4] R. Abraham and J. E. Marsden, Ref. 3, p. 177.

[5] R. Abraham and J. E. Marsden, Ref. 3, p. 187.

Inverse problem for the reduced wave equation with fixed incident field. III

V. H. Weston
Department of Mathematics, Purdue University, West Lafayette, Indiana 47907

(Received 28 August 1981; accepted for publication 8 January 1982)

The inverse problem for the reduced wave equation $\Delta\, u + k^2\, n^2(x)u = 0, x \in R^3$, is examined for the case where measurements of the amplitude of the scattered field (produced by a fixed incident field at a single frequency) are obtained at a finite number of points. A strategy is given for the recovering of the phase data through the minimization of a quadratic form involving comparison data. The problem is then reduced to the problem treated in previous papers where the complex-valued quantities $u^s(x_l)$ are known at a finite number of points. A relationship between the smallest eigenvalue of the "measurement" matrix and $\|\mathbb{K}\|_2$ is given.

PACS numbers: 03.40.Kf

I. INTRODUCTION

The inverse problem for the reduced wave equation

$$\Delta\, u + k^2 n^2(x)u = 0, \quad x \in R^3, \tag{1}$$

associated with time dependence $\exp(-i\omega t)$ was examined[1,2] for the case where measurements of the scattered field u^s (produced by an incident field u^i generated by a fixed source) were obtained at a finite number of points $\{x_j\}_{j=1}^N$. The index of refraction was taken to be unity outside and on the surface of a compact region D (enclosing the scattering object).

Here we not only simplify some of the previous notation and analysis, but amplify the previous results for the case where the real and imaginary parts of the scattered or total field are measured. Most important, however, we extend the results to include the case where only the amplitude (modulus) of the scattered or total field is measured.

Because the inverse problem with the sparse data as treated here is not well posed (there are infinite numbers of solutions), we need to impose additional constraints. With this in mind, we need to specify an *a priori* known comparison value $n_*(x)$ for the index of refraction, which is used to restrict the class of scatterers under consideration. For almost transparent objects, a natural choice of n_* would be unity. If, however, one has from other knowledge a rough guess for $n(x)$, then this could be used as the value for $n_*(x)$. The iteration techniques to be presented in the later section of this paper would then yield a correction to this initial approximation.

It is convenient to replace the unknown quantity $n(x)$ by $v(x)$ where

$$v(x) = n^2(x) - 1 \tag{2}$$

with a corresponding connotation for $v_*(x)$.

The set of scattered field measurements at the points $\{x_j\}_{j=1}^N$ outside D, then gives rise to the system of N nonlinear equations in the unknown quantity $v(x)$,

$$u(x_j) - u_*(x_j) = b_j, \quad j = 1,2,\ldots,N. \tag{3}$$

The complex numbers b_j correspond to the difference of the measured values of the total field $u(x_j) = u(x_j;v)$ and the calculated values (comparison data) of the total field $u_*(x_j) = u(x_j, v_*)$ associated with the index of refraction $n_*(x)$.

For the problem where only the amplitude

$$|u(x_j)| = \rho_j \tag{4}$$

of the total field is measured at the points $\{x_j\}_{j=1}^N$, the nonlinear equations corresponding to (3) are given by

$$|u_*(x_j) - b_j| = \rho_j, \quad j = 1,2,\ldots,N. \tag{5}$$

(Note if the amplitude ρ_j^s of the scattered field is measured, then in Eq. (5), u_* and ρ_j are replaced by u_*^s and ρ_j^s, respectively.)

To make the problem well posed, an additional constraint that corresponds to finding the solution that is closest to $n_*(x)$ will be imposed. The required constraint will be given initially by the following condition:

$$\min \int_D (n^2 - n_*^2)^2\, dx. \tag{6}$$

However, for practical purposes (as will be seen), certain modifications of this condition, such as

$$\min \int_D (n^2 - n_*^2)^2 |u/u_*|^2\, dx \tag{6a}$$

will be examined in more detail.

For the problem where only the amplitude is measured, additional constraints have to be imposed along with (6) or (6a). These will be given in a later section.

The assumptions that will be imposed on n and n_* are the same as given in Paper II,[2] namely that $n(x)$ be real bounded and continuous everywhere in $\bar{D}$ except for a finite number of surfaces across which n is discontinuous. In addition, n_* must be such that the total field $|u_*|$ is bounded from zero in $\bar{D}$, and the Green's function $G(x,y;v_*)$ exists, and as a kernel of an integral operator maps $\mathscr{L}_2(D)$ into $C(\bar{D})$.

With the application of the Green's function the set of nonlinear functional equations (3) involving the data has the explicit form

$$k^2 \int_D G(x_j,y,v_*)[v(y) - v_*(y)]\, u(y,v)\, dy = b_j, \tag{7}$$

with $j = 1,2,\ldots,N$, where the total field $u(x,v)$ satisfies the integral equation

$$u(x,v) = u(x,v_*) + k^2 \int_D G(x,y,v_*)[v(y) - v_*(y)]\, u(y,v)\, dy. \tag{8}$$

The inverse problem for the case where the real and imagi-

0022-2488/83/040828-06$02.50

nary parts of the total or scattered field is measured reduces to solving system (7) and (8) subject to constraint (6) or (6a).

For the case where only the amplitude of the field quantities are measured, the complex quantities $\{b_j\}_{j=1}^{N}$ are not explicitly known; the only information on them is given by Eq. (5) where the known quantities are $\{\rho_j\}_{j=1}^{N}$ (obtained from measurements) and $\{u_*(x_j)\}_{j=1}^{N}$ (obtained from calculations). The approach to the inverse problem for this case will be to impose an additional natural constraint (specified in a later section), and to use this coupled with Eq. (5) to obtain $\{b_j\}_{j=1}^{N}$. Thus the inverse problem (where only the field amplitude is measured) is thus reduced to the case where the real and imaginary parts of the field quantities are known.

II. THE INVERSE PROBLEM ASSOCIATED WITH PHASE AND AMPLITUDE MEASUREMENTS

The inverse problem associated with the measurements of the real and imaginary parts of the field quantities is treated first. With a change in some of the notation for simplification, a review of the previous[2] procedure and results are given. At the same time some additional results are presented.

The system of N nonlinear complex equations (7) is reduced to a system of $2N$ real linear functional equations through the introduction of the real functions $\varphi(x)$ and $\psi(x)$ defined by

$$(v - v_*)u(x,v) = [\varphi(x) + i\psi(x)]\, u_*(x). \tag{9}$$

(Note that φ and ψ are *defined* slightly differently than in the previous work.[2] With $(u,v) = \int_D u(x)v(x)dx$, system (7) becomes

$$(H_i,\varphi) = \sum_{j=1}^{2N} e_{ij}(H_j,\psi) + B_i \tag{10}$$

with $i = 1,2,\ldots,2N$. Here B_j corresponds to the real and imaginary parts of the data b_j through the relation

$$b_j = B_j + iB_{j+N}, \quad j = 1,2,\ldots,N, \tag{11}$$

and $H_j(y)$ corresponds in a similar manner to

$$H_j(y) + iH_{j+N}(y) = k^2 G(x_j, y, v_*)\, u_*(y, v_*) \tag{12}$$

with x_j corresponding to the points where the measurements were made. The numbers e_{ij} are related to the Kronecker delta by

$$e_{ij} = \begin{cases} \delta_{(i+N)j}, & i = 1,2,\ldots,N, \\ -\delta_{i(j+N)}, & i = N+1,\ldots,2N. \end{cases} \tag{13}$$

By noting that $v(x)$ is a real quantity, integral equation (8) is decomposed to yield two real expressions. One expression, relating φ and ψ through a quadratic integral equation, is given by

$$\begin{aligned} &\psi = S(\psi,\varphi), \\ &S(\psi,\varphi) = \psi(L_I\psi - L_R\varphi) + \varphi(L_R\psi + L_I\varphi), \end{aligned} \tag{14}$$

where L_R and L_I are the real and imaginary parts of the integral operator with kernel

$$k^2 G(x,y,v_*)\, u_*(y)/u_*(x), \quad x, y \in D. \tag{14'}$$

The second expression

$$v(x) - v_*(x) = \varphi(x)/[1 + L_R\varphi - L_I\psi] \tag{15}$$

relates $v(x)$ to φ and ψ and allows one to recoup the value of $v(x)$ from knowledge of φ and ψ.

It will be assumed that the measured position $\{x_j\}_{j=1}^{N}$ will be selected so that $\{H_j\}_{j=1}^{2N}$ forms an independent set, in which case we will let $\mathscr{M}$ denote the subspace of $\mathscr{L}_2(D)$ spanned by $\{H_j\}_{j=1}^{2N}$.

Let H be the real symmetric matrix with elements H_{ij} given by

$$H_{ij} = (H_i, H_j) \tag{16}$$

and denote its inverse by $\tilde{H}$ with associated elements $\tilde{H}_{ij}$, i.e.,

$$\sum_{k=1}^{2N} H_{ik}\, \tilde{H}_{kj} = \delta_{ij}.$$

Note that the matrix H is positive definite, since if $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_{2N}) \in R^{2N}$, the quadratic form

$$\sum_{ij=1}^{2N} \sigma_i H_{ij} \sigma_j = \int_D \Big(\sum_{i=1}^{2N} \sigma_i H_i(y) \Big)^2 dy \geqslant 0$$

vanishes only if $\sigma \equiv 0$. Hence the eigenvalues λ_i of H can be ordered as follows:

$$0 < \lambda_1 \leqslant \lambda_2 \cdots \leqslant \lambda_{2N}.$$

If we set

$$\xi_i(x) = \sum_{j=1}^{2N} \tilde{H}_{ij}\, H_j(x) \tag{17}$$

and note that

$$(\xi_i, H_j) = \delta_{ij}, \tag{18}$$

the system of Eq. (10) can be solved (nonuniquely) for φ in terms of ψ, giving

$$\varphi(x) = \Phi(x) + \mathbb{K}\psi + \varphi^{\perp}(x), \tag{19}$$

where the data term is given by

$$\Phi(x) = \sum_{j=1}^{2N} B_j\, \xi_j(x), \tag{20}$$

and

$$\mathbb{K}\psi = \sum_{j,k=1}^{2N} \xi_j(x) e_{jk} (H_k, \psi). \tag{21}$$

The unknown function $\varphi^{\perp}(x)$ belongs to $\mathscr{M}^{\perp}$ the orthogonal complement of $\mathscr{M}$. As a consequence, note that

$$(\Phi, \varphi^{\perp}) = (\mathbb{K}\psi, \varphi^{\perp}) = 0.$$

The operator $\mathbb{K}$ has the property that

$$-\mathbb{K}^2 u = \sum_{j=1}^{2N} \xi_j(x)(H_j, u) = \mathbb{P}u, \tag{22}$$

where $\mathbb{P}$ is the projection operator on the subspace spanned by $\{H_j\}$. The norm $\|\mathbb{K}\|_2$, which plays an important role in subsequent analysis, depends upon the measurement positions $\{x_j\}_{j=1}^{2N}$, through the smallest eigenvalue of the measurement matrix $H = \{H_{ij}\}$. The precise relationship is given as follows.

Lemma: $\|\mathbb{K}\|_2 = (\lambda_1)^{-1/2}$, where λ_1 is the smallest eigenvalue of the matrix H.

Proof: Set $u = \Sigma_{i=1}^{2N} c_i \xi_i(x) + u^{\perp}(x)$, where $u^{\perp}$ is orthogonal to the space $\mathscr{M}$. It then follows that

$$\mathbb{K}u = \sum_{j,k=1}^{2N} \xi_j(x) e_{jk}\, c_k.$$

Using the result that $(\xi_i,\xi_j) = \tilde{H}_{ij}$, it is seen

$$(\mathbb{K}u,\mathbb{K}u) = c^T E^T \tilde{H} E c,$$

$$(u,u) = c^T \tilde{H} c + (u^\perp,u^\perp),$$

where c is the $2n \times 1$ matrix with components $c_1,c_2,\cdots$ and E is the matrix with coefficients e_{ij}. Since $\tilde{H}$ is a symmetric positive definite matrix, there exists a nonsingular transformation Q such that $Q^T \tilde{H} Q = I$. Hence, setting $c = Q\sigma$, the following is obtained:

$$(\mathbb{K}u,\mathbb{K}u)/(u,u) < \sigma^T T\sigma/\sigma^T\sigma,$$

where the matrix $T = Q^T E^T \tilde{H} E Q$ is similar to the matrix $\tilde{H} = H^{-1}$ and hence has the same eigenvalues. Thus

$$(\mathbb{K}u,\mathbb{K}u)/(u,u) < \text{largest eigenvalue of } \tilde{H} = 1/\lambda_1.$$

The result follows.

The inverse problem is now reduced to solving a single nonlinear equation, namely Eq. (14) [with φ replaced by expression (19)]

$$\psi = S(\psi, \Phi + \mathbb{K}\psi + \varphi^\perp)$$

for the two unknowns ψ and $\varphi^\perp$. This yields a solution of the form $\psi = \psi(\varphi^\perp)$. As was pointed out, a constraint is needed to give an additional relationship between ψ and $\varphi^\perp$. The constraint (requiring n close to n_*) given by Eq. (6) reduces to the form

$$\min_{\varphi^\perp \in \mathscr{M}^\perp} \int_D \frac{\varphi^2}{(1 + L_R\varphi - L_I\psi)^2}\, dy.$$

Because of the complications arising from the denominator in this integral, a simpler and more useful version of this constraint, employed and examined in detail previously,[2] is given by

$$\min \int_D (n^2 - n_*^2)^2 (R_e u/u_*)^2\, dy = \min \int_D \varphi^2\, dy. \qquad (6b)$$

An alternative choice, which may be of more physical interest, is given by Eq. (6a), which reduces to

$$\min_{\varphi^\perp \in \mathscr{M}^\perp} \int_D [(\Phi + \mathbb{K}\psi)^2 + \psi^2 + (\varphi^\perp)^2]\, dy. \qquad (6a)$$

Since all three choices of constraints have the form

$$\min_{\varphi^\perp \in \mathscr{M}^\perp} \int_D F(\varphi^\perp)\, dy,$$

the minima (or minimum) are found by selecting the stationary solutions (given by setting the first Gateaux derivative to zero):

$$\delta \int_D F(\varphi^\perp)\, dy = 0$$

In all cases, the problem of finding the stationary solution reduces to solving a nonlinear integral equation of the form

$$\varphi^\perp = -(I - \mathbb{P}) S^*_{\varphi^\perp} (I - S^*_\psi)^{-1} V(\psi), \qquad (23)$$

$$\text{where} \quad V(\psi) = \begin{cases} \mathbb{K}^*(\Phi + \mathbb{K}\psi) + \psi & \text{for case (6a),} \\ \mathbb{K}^*(\Phi + \mathbb{K}\psi) & \text{for case (6b).} \end{cases}$$

Here $\mathbb{P} = -\mathbb{K}^2$ is the projection operator on $\mathscr{M}$, and $\mathbb{K}^*, S^*_{\varphi^\perp}$, and S^*_ψ are the adjoint operators of $\mathbb{K}$, $S_{\varphi^\perp}$, and S_ψ. Note that the explicit expressions given by Eqs. (26) and (27) of Paper II may be used for the latter two operators, except that for this paper the appropriate kernels of the integral operator L_R and L_I are given by Eq. (14a) of this paper.

The resulting system (14) and (23) of nonlinear integral equations for ψ and $\varphi^\perp$ has the important property that it possesses the trivial solution $\psi = \varphi^\perp = 0$, when the data term Φ vanishes. Hence small norm $\|\psi\|_2 + \|\varphi^\perp\|_2$ solutions are sought. It is shown[2] that the system corresponding to constraint (6b) could be solved by the method of successive approximations, starting from the initial approximation $\psi_0 = \varphi_0^\perp = 0$, and the solution yields a (local) minimum, provided that the data term is not too large. To be precise, a unique local minimum is obtained if the data satisfies a constraint of the form

$$\sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j = (\|\Phi\|_2)^2 \leqslant 0.0036\, \tilde{k}^{-4}(\|L_I\| + \|L_R\|)^{-2} \qquad (24)$$

where $\tilde{k} = \max[1,\|\mathbb{K}\|_2] = \max[1,\lambda_1^{-1/2}]$ and λ_1 is the smallest (positive) eigenvalue of the matrix H.

It is interesting to note that the first iterated solution of system (14) and (23), starting from $\psi_0 = \varphi^\perp = 0$, is the same for either constraint (6a) or (6b), and, if we retain only terms up to second order in Φ, the corresponding value for v [obtained from Eq. (25)] is given by

$$v \sim v_* + \Phi - \Phi L_R \Phi + \mathbb{K}(\Phi L_I \Phi) + \varphi_1^\perp,$$

where $\varphi_1^\perp = -(I - \mathbb{P})[L_I^*(\Phi \mathbb{K}^* \Phi) + (L_I \Phi)(\mathbb{K}^* \Phi)]$.

This result is equivalent to using the modified Born approximation solution of Eq. (8).

III. PROCEDURE FOR RECOVERING PHASE FROM AMPLITUDE MEASUREMENTS

Here, given data $|u(x_l)| = \rho_l$, $l = 1,2,\ldots,N$, the problem is to obtain the real quantities $\{\mathrm{B}_i\}_{i=1}^{2N}$ defined by Eqs. (3) and (11). With this in mind set

$$\left.\begin{aligned} u(x_l) &= \eta_l + i\eta_{l+N} \\ u_*(x_l) &= \eta_l^* + in_{l+N}^* \end{aligned}\right\}, \quad l = 1,2,\ldots,N$$

yielding $B_l = \eta_l - \eta_l^*$, $l = 1,2,\ldots,N$.

The relationship between the data ρ_l and B_l is given by the following expression:

$$(B_l + \eta_l^*)^2 + (B_{l+N} + \eta_{l+N}^*)^2 = \rho_l^2, \quad l = 1,2,\ldots,N, \qquad (25)$$

where $\{\eta_l^*\}$ are known quantities obtained from calculations. The problem reduces to the determination of the quantities $\{B_l\}_{l=1}^{2N}$ from system (25). Since the solution of this system is not unique, we need to impose an additional constraint so as to select a unique solution. Here, the choice of this additional constraint will be one that automatically fits in with the theory and analysis for the inverse problem under investigation here.

In the previous section, it was shown that the inverse problem where the real and imaginary parts of the field quantities are known ($\{B_i\}_{i=1}^{2N}$ are given) has a unique solution in a certain ball of function space, provided that the data satisfies a condition of the form $\Sigma_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j <$ const. This then suggests that a natural condition to impose for uniqueness would be to select the solution that

$$\min \sum_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j. \tag{26}$$

For subsequent analysis we shall qualify this condition further. Rather than select the solution that gives the absolute minimum, we will investigate here the conditions for existence of a local minimum in a neighborhood of the point $B_i = 0, i = 1, \ldots, 2N$, and develop a method to obtain this solution.

The method that will be employed here will be to look for stationary solution of the quadratic form [given by expression (26)] satisfying relations given by Eq. (25), and then check to see if and when stationary solutions in the neighborhood of the point $B_i = 0$ yield a local minimum.

The stationary points of $f = \Sigma_{i,j=1}^{2N} B_i \tilde{H}_{ij} B_j$ subject to condition (2) are obtained by setting $\partial f / \partial B_i = 0$, $i = 1,2, \ldots, N$, where $B_i, i = 1,2, \ldots, N$, are treated as independent variables and $B_i, i = N+1, \ldots, 2N$, are treated as dependent variables as defined by relation (25). This yields the following system expressed in compact form:

$$P(B) = 0. \tag{27}$$

Here $P(B)$ is a vector-valued function of $B = (B_1, B_2, \ldots, B_{2N})$ with components

$$p_i(B) = \sum_{j=1}^{2N} \left[(B_{i+N} + \eta^*_{i+N}) \tilde{H}_{ij} - (B_i + \eta^*_i) \tilde{H}_{i+Nj} \right] B_j, \tag{28a}$$

for $i = 1,2, \ldots, N$,

$$p_i(B) = \tfrac{1}{2}(B_i + \eta^*_i)^2 + \tfrac{1}{2}(B_{i-N} + \eta^*_{i-N})^2 - \tfrac{1}{2}\rho^2_{i-N}, \tag{28b}$$

for $i = N+1, \ldots, 2N$.

System (27) can be rewritten in the following form, demonstrating its quadratic nature:

$$P(0) = P'(0)B + \tfrac{1}{2} P''(0)BB = 0. \tag{29}$$

Here $P'(0)$ is the Fréchet derivative of $P(B)$ at $B = 0$, with a similar connotation for $P''(0)$. The linear operator $P'(0)$ mapping R^{2N} into itself can be written in the matrix

$$P'(0) = \begin{bmatrix} F & G \\ 0 & D \end{bmatrix} \begin{bmatrix} D_2 & -D_1 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}, \tag{30}$$

where D_1 and D_2 are diagonal $N \times N$ matrices with diagonal elements $\eta^*_1, \eta^*_2, \ldots, \eta^*_N$ and $\eta^*_{N+1}, \ldots, \eta_{2N}$, respectively. The matrix D, defined by

$$D = D_1^2 + D_2^2, \tag{31}$$

is a diagonal matrix with diagonal elements $(\rho^*_1)^2, \ldots, (\rho^*_N)^2$ [where $|u_*(x_l)|^2 = (\rho^*_l)^2$]. It is assumed that these diagonal elements are bounded from zero; hence the inverse matrix D^{-1} exists. The N-square matrices F and G are defined by

$$F = [D_2 - D_1] \tilde{H} \begin{bmatrix} D_2 \\ -D_1 \end{bmatrix} \quad \text{and} \\ G = [D_2 - D_1] \tilde{H} \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}. \tag{32}$$

The importance of the block representation for $P'(0)$ is that it can be easily deduced that since F^{-1} exists (see Appendix), then $[P'(0)]^{-1}$ exists and is given by

$$[P'(0)]^{-1} = \begin{bmatrix} D_2 & D_1 \\ -D_1 & D_2 \end{bmatrix} \begin{bmatrix} F^{-1} & -F^{-1}GD^{-1} \\ 0 & D^{-1} \end{bmatrix}. \tag{33}$$

The symmetric bilinear operator $P''(0)$ mapping $R^{2N} \otimes R^{2N}$ in R^{2N} takes the form

$$z = P''(0)\sigma\alpha, \tag{34}$$

where z, σ, α are vectors in R^{2N}. The components of the transformation are given explicitly for $i = 1,2, \ldots, N$ by

$$z_i = \sum_{j=1}^{2N} \tilde{H}_{ij} (\alpha_{i+N} \sigma_j + \sigma_{i+N} \alpha_j) \\ - \sum_{j=1}^{2N} \tilde{H}_{i+Nj} (\alpha_i \sigma_j + \sigma_i \alpha_j) \tag{35a}$$

and for $i = N+1, \ldots, 2N$ by

$$z_i = \alpha_i \sigma_i + \alpha_{i-N} \sigma_{i-N}. \tag{35b}$$

Equation (29) may now be placed in the following form:

$$B = X(B) \\ = -[P'(0)]^{-1} P(0) - \tfrac{1}{2} [P'(0)]^{-1} P''(0)BB, \tag{36}$$

which suggests that the method of successive approximations may be used to solve it. In fact, from Rall,[3,4] we can state the following. Provided that $h_0 < \frac{1}{4}$, where

$$h_0 = \| \tfrac{1}{2}[P'(0)]^{-1} P''(0) \|_2 \, \eta_0 \tag{37a}$$

and

$$\eta_0 = \| [P'(0)]^{-1} P(0) \|_2, \tag{37b}$$

the sequence

$$B^0 = X(0), \\ B^{n+1} = X(B^n) \tag{38}$$

converges to solution B^{**} of Eq. (36), lying in the ball

$$\|B^{**}\|_2 \leqslant [1 - (1 - 4h_0)^{1/2}] \, \eta_0 / 2h_0.$$

To check the condition $h_0 < \frac{1}{4}$, we should note that $\|P(0)\|_2 = [\frac{1}{4} \Sigma_{i=1}^N (\rho^{*2}_i - \rho^2_1)^2]^{1/2}$; hence, using the estimates for $\|[P'(0)]^{-1}\|_2$ and $\|P''(0)\|_2$ given in the Appendix, we obtain

$$h_0 \leqslant \frac{1}{2} \frac{(1 + 4/\lambda_1^2)^{1/2}}{(\max \rho^*_i)^2} [\lambda^2_{2N} + (\lambda_{2N} \Lambda / \lambda_1)^2 + 1] \Lambda^2 \|P(0)\|_2,$$

where $\Lambda = (\max \rho^*_i / \min \rho^*_i)^2$, and λ_1 and λ_{2N} are, respectively, the smallest and largest eigenvalues of the matrix H. This condition can be placed in the form

$$\sum_{i=1}^{N} (\rho^{*2}_i - \rho^2_i)^2 \leqslant C(\lambda_1, \lambda_{2N}, \max \rho^*_i, \min \rho^*_i), \tag{39}$$

which says that the method of successive approximations starting from the indicated initial approximation converges, provided that the measured amplitude data set $\{\rho_i\}$ is sufficiently close to $\{\rho^*_i\}$ as is expected.

With the initial approximation $B^{(0)} = 0$, the first and second iterates are given by

$$B^{(1)} = -[P'(0)]^{-1} P(0), \\ B^{(2)} = B^{(1)} - \tfrac{1}{2} [P'(0)]^{-1} P''(0) B^{(1)} B^{(1)}. \tag{40}$$

We now need to determine the circumstance for which the stationary solution of Eq. (27) yields a minimum. Since we have assumed that $\rho_i \neq 0$ for $i = 1,2,\ldots,N$, it follows that from Eq. (25) that for each i ($i = 1,2,\ldots,N$) one of the pair $(B_i + \eta_i^*)$, $(B_{i+N} + \eta_{i+N}^*)$ does not vanish. For the following analysis only, we shall assume that $(B_{i+N} + \eta_{i+N}^*) \neq 0$ for each i, and will take $B_1, B_2, \ldots, R_N$ as independent variables. (If one or more of $B_{i+N} + \eta_{i+N}^*$ vanishes, then the corresponding independent variable will be changed to B_{i+N} and the analysis can be suitably modified.) In any event, set $f(B_1,\ldots,B_N) = \sum_{i,j=1}^{N} B_i \tilde{H}_{ij} B_j$ and, using condition (25),

$$\frac{\partial B_{i+N}}{\partial B_i} = -\frac{(B_i + \eta_i^*)}{(B_{i+N} + \eta_{i+N}^*)} = q_i,$$

the resulting stationary points are given by ($i = 1,2,\ldots,N$)

$$\frac{1}{2}\frac{\partial f}{\partial B_i} = \sum_{j=1}^{2N} (\tilde{H}_{ij} + q_i \tilde{H}_{i+Nj}) B_j = 0. \tag{41}$$

In addition it follows for $i,j = 1,2,\ldots,N$,

$$\frac{1}{2}\frac{\partial^2 f}{\partial B_i \partial B_j} = \tilde{H}_{ij} + q_j \tilde{H}_{ij+N} + q_i \tilde{H}_{i+Nj} + \tilde{H}_{i+Nj+N} q_i q_j - \delta_{ij}(1 + q_i^2) \sum_{k=1}^{2N} H_{i+Nk} B_k/(B_{i+N} + \eta_{i+N}^*),$$

where δ_{ij} is the Kronecker delta.

To show that the stationary solution is indeed a minimum, we need to demonstrate that the quadratic form

$$\frac{1}{2}\sum_{i,j=1}^{N} x_i \frac{\partial^2 f}{\partial B_i \partial B_j} x_j$$

is positive definite. If y is a vector in R^{2N} with components $y_i = x_i$, $y_{i+N} = q_i x_i$ for $i = 1,2,\ldots,N$, then the quadratic form can be written in matrix notation

$$y^T(\tilde{H} - R)y, \tag{42}$$

where R is a square $2N$ matrix with diagonal elements r_i given by

$$r_i = r_{i+N} = \sum_{k=1}^{2N} \frac{\tilde{H}_{i+Nk} B_k}{B_{i+N} + \eta_{i+N}^*} \tag{43a}$$

$$= \sum_{k=1}^{2N} \frac{\tilde{H}_{ik} B_k}{B_i + \eta_i^*}. \tag{43b}$$

The last equality holds since $\{B_k\}$ is a stationary point satisfying Eq. (41). Since $\tilde{H}$ is a positive-definite matrix, it is easily seen that the quadratic form (42) will be positive definite if the smallest eigenvalue of $\tilde{H}$, namely $1/\lambda_{2N}$, is greater than $\max_{i+1,\ldots,N} |r_i|$.

To estimate $\max|r_i|$, we will need to select from Eq. (43a) or (43b) that expression for which the term $(B_i + \eta_i^*)$ or $(B_{i+N} + \eta_{i+N}^*)$ is the larger. Hence, if $(B_{i+n} + \eta_{i+N}^*)$ is the larger of the two terms, it follows from equality (25) that $(B_{i+N} + \eta_{i+N}^*)^2 > \frac{1}{2}\rho_i^2$. Hence, using such an estimate, it follows from Eq. (43a)

$$|r_i| \leqslant \left| \sum_{k=1}^{2N} \frac{\tilde{H}_{i+Nk} B_k 2^{1/2}}{\rho_i} \right| \leqslant \frac{\|\tilde{H}\|_2 \|B\|_2 2^{1/2}}{\rho_i} = \frac{2^{1/2} \|B\|_2}{\lambda_1 \rho_i}.$$

If $B_i + \eta_i^*$ is the larger term, we obtain a similar estimate using Eq. (43b). Thus we obtain

$$\max_{i=1,\ldots,N} |r_i| \leqslant 2^{1/2} \|B\|_2/(\lambda_1 \min \rho_i).$$

Consequently, a sufficient (but not necessary) condition for the stationary point to be a minimum is that

$$2^{1/2} \|B\|_2/(\lambda_1 \min \rho_i) < 1/\lambda_{2N}$$

or that the solution lies in the ball center $B = 0$, with radius

$$\|B\|_2 \leqslant \lambda_1 \min \rho_i/(2^{1/2} \lambda_{2N}). \tag{44}$$

Summarizing, if the measured data is sufficiently close to the comparison data, then the quadratic system of $2N$ Eqs. (36) may be solved for $B_1 \cdots B_{2N}$ by the method of successive approximations, yielding a local minimum of expression (26). The first two iterates are given by (40). The phase and amplitude are recovered from the amplitude measurements.

IV. SUMMARY AND COMMENTS ON THE "MEASUREMENT" MATRIX

The problem where only the amplitude of the scattered field $|u^s(x_j)| = \rho_j$ is measured at a finite number N of points is reduced to the corresponding inverse problem associated with measurements of the phase and amplitude of the scattered field at these points. This is achieved by choosing the real and imaginary parts so as to minimize a quadratic form Eq. (26) in the $2N$ real variables B_j (representing the difference between the real parts of the measured and comparison data, and the corresponding imaginary parts). Here, the comparison data is associated with a known or prescribed value of n_*. The stationary solutions of the quadratic form give rise to a quadratic system of equations in R^{2N}. This system, Eq. (27), is transformed to a simpler standardized form, Eq. (36). It is shown that if the measured data ρ_j is sufficiently close to the comparison data ρ_j^* [see Eq. (39)], the method of successive approximations applied to Eq. (36) yields a solution in the neighborhood of the point $B_i = 0$, this solution being a local minimum of the quadratic form. The first two iterates are given by Eq. (40).

The previous approach[2] to the inverse problem associated with knowledge of both the real and imaginary parts of the scattered fields at a finite set of points is reviewed together with a simplification of method and a slight extension of the results. The second-order correction terms to $v(x)$, arising from the method of successive approximations of the resulting nonlinear system of equations, is given at the end of Sec. II. Again these results are valid provided that the measurement data are sufficiently close to the comparison data, as indicated by Eq. (24).

The conditions that the measured data be sufficiently close to the comparison data, as expressed by Eqs. (24) and (39), contain the parameters λ_1 and λ_{2N} the smallest (positive) and largest eigenvalues of the "measurement" matrix H, whose elements H_{ij} are defined by Eq. (16). What this means [especially see Eq. (24)] is that the measurement positions $\{x_j\}$ must be judiciously chosen so that the matrix H does not become ill-conditioned, i.e., λ_1 is too small. This will also imply that $\|\mathbb{K}\|_2$ is not too large.

Finally, there remains to be investigated the case where the measured data is not close enough to the comparison data for the above results to hold.

The application of the results in these studies to nondestructive testing (differentiating voids and cracks), and the consequences of these results on presently employed techniques, such as linearization of the problem, will be detailed elsewhere.

APPENDIX

Lemma A.1:

$$\| F^{-1}\|_2 \leqslant \lambda_{2N} \left[\min_{i=1,\dots,N} (\rho_i^*)^2 \right]^{-1}.$$

Proof: Let ν and σ be vectors in R^N and R^{2N}, respectively, and set

$$\sigma = \begin{bmatrix} D_2 \\ -D_1 \end{bmatrix} \nu.$$

Then from Eq. (32)

$$\nu^T F\nu/\nu^T\nu = (\sigma^T \tilde{H}\sigma/\sigma^T\sigma)\,(\nu^T D\nu/\nu^T\nu)$$

$$\geqslant(\lambda_{2N})^{-1} \min_{i=1,\dots,N} (\rho_i^*)^2 > 0,$$

where $(\lambda_{2N})^{-1}$ is the smallest eigenvalue of the positive definite matrix $\tilde{H}$. Since, as is shown, F is positive definite, F^{-1} exists, and, since $\|F\|_2 \geqslant (\lambda_{2N})^{-1} \min_{i=1,\dots,N}(\rho_i^*)^2$, the estimate for the norm $\|F^{-1}\|_2$ follows.

Lemma A.2:

$$\left\| \begin{matrix} D_2 & D_1 \\ -D_1 & D_2 \end{matrix} \right\|_2 = \max_{i=1,\dots,N} \rho_i^* .$$

Proof:

$$z = \begin{bmatrix} D_2 & D_1 \\ -D_1 & D_2 \end{bmatrix} y,$$

where z and y are vectors in R^{2N}. Then a quick calculation yields the results

$$(\|z\|_2)^2 = \sum_{i=1}^{N} (\rho_i^*)^2\,[y_i^2 + y_{i+N}^2]$$

$$\leqslant \max_{i=1,\dots,N} (\rho_i^*)^2\,(\|y\|_2)^2.$$

Lemma A.3:

$$\|[P'(0)]^{-1}\|_2 \leqslant \Lambda\,[(\lambda_{2N})^2 + (\lambda_{2N}\,\Lambda/\lambda_1)^2 + 1]^{1/2}/\max \rho_i^*$$

where $\Lambda = (\max \rho_i^*/\min \rho_i^*)^2$.

Proof: Let A be the partitioned matrix

$$A = \begin{bmatrix} F^{-1} & -F^{-1}GD^{-1} \\ 0 & D^{-1} \end{bmatrix}$$

and y a column vector in R^{2N} partitioned into two vectors y_1 and y_2 in R^N; then

$$(\|Ay\|_2)^2$$

$$= (\|F^{-1}y_1 - F^{-1}GD^{-1}y_2\|_2)^2 + (\|D^{-1}y_2\|_2)^2$$

$$\leqslant (\|F^{-1}y_1\|_2 + \|F^{-1}GD^{-1}y_2\|_2)^2 + (\|D^{-1}y_2\|_2)^2$$

$$\leqslant \{(\|F^{-1}\|_2)^2[1 + (\|GD^{-1}\|_2)^2] + (\|D^{-1}\|_2)^2\}\,\|y\|_2)^2 .$$

Since D is a diagonal matrix, it is easily seen that $\|D^{-1}\|_2 = 1/\min_{i=1,\dots,N}(\rho_i^*)^2$. It can be shown that the norms of the partitioned matrices $[D_2\ D_1]$ and $[D_1\ D_2]$ are bounded above by $\max_{i=1,\dots,N} \rho_i^*$; hence from Eq. (32) it follows that

$$\|G\|_2 \leqslant (1/\lambda_1)\,(\max \rho_i^*)^2 .$$

Thus we can obtain the following estimate:

$$\|A\|_2 \leqslant [(\lambda_{2N})^2 + (\lambda_{2N}\,\Lambda/\lambda_1)^2 + 1]^{1/2}/(\min \rho_i^*)^2$$

$$\text{where}\quad \Lambda = (\max \rho_i^*/\min \rho_i^*)^2.$$

Hence from Eq. (33) we obtain the desired result.

Lemma A.4: $\|\mathrm{P}''(0)\|_2 \leqslant (1 + 4/\lambda_1^2)^{1/2}$, where λ_1 is the smallest eigenvalue of H.

Proof: Using the inequality for components of vectors a and b in R^{2N}

$$\sum_{i=1}^{N} (a_i\,b_i + a_{i+N}\,b_{i+N})^2 \leqslant \sum_{i=1}^{2N} a_i^2 \sum_{j=1}^{2N} b_j^2 = (\|a\|_2\,\|b\|_2)^2$$

it is seen from Eq. (35b) that

$$\sum_{i=N+1}^{2N} z_i^2 \leqslant (\|\alpha\|_2\,\|\sigma\|_2)^2.$$

Using the triangle law for norms and the above inequality, it follows from Eq. (35a)

$$\sum_{i=1}^{N} z_i^2 \leqslant (\|\alpha\|_2\,\|\tilde{H}\sigma\|_2 + \|\sigma\|_2\,\|\tilde{H}\alpha\|_2)^2$$

$$\leqslant (2\|\tilde{H}\|_2\,\|\alpha\|_2\,\|\sigma\|_2)^2 = [(2/\lambda_1)\|\alpha\|_2\,\|\sigma\|_2]^2.$$

Thus we have

$$(\|z\|_2)^2 \leqslant (1 + 4/\lambda_1^2)\,\|\alpha\|_2^2\,\|\sigma\|_2^2.$$

---

[1] V. H. Weston, "Inverse problem for the reduced wave equation with fixed incident field," J. Math Phys. **21**, 758–64 (1980).

[2] V. H. Weston, J. Math Phys. **22**, 2523 (1981).

[3] L. B. Rall, *Non-Linear Functional Analysis and Applications* (Academic, New York, 1971), pp. 383–94.

[4] L. B. Rall, "Quadratic Equations in Banach Space," Rend. Circulo Math. Palermo **10**, 314–32 (1961).

Operator formalism equivalent to the Feynman quantization technique

V. Janiš and J. Souček
Mathematical Institute, Czechoslovak Academy of Sciences, Žitná 25, 115 67 Prague, Czechoslovakia
V. Souček
Department of Mathematical Analysis, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 00 Prague, Czechoslovakia

(Received 22 January 1982; accepted for publication 5 February 1982)

The concept of a Fock–Stueckelberg space of quantum states and a procedure of an operator quantization using only Lagrangians (kinematical quantization) are introduced. A propagator operator $\mathscr{K}$, matrix elements of which are Green's functions, is used, and an equation of motion for it is derived. We prove that kinematical quantization is an operator (coordinate-free) form of the Feynman quantization technique. The Feynman path integral (FPI) is obtained as a spectral representation of the operator $\mathscr{K}$ in a coordinate basis. The connection of a representation of commutation relations in this scheme, the domain of integration in FPI, and causality is mentioned.

PACS numbers: 03.65. – w

I. INTRODUCTION

In recent years functional methods in quantum field theory were used very often. This technique is a very useful tool for going beyond perturbative calculations in many physically interesting situations such as spontaneously broken symmetries,[1] non-abelian gauge theories,[2] quantum gravity,[3] etc. Using functional methods, we have Green's functions as basic constituents in the theory without mentioning the Hilbert space of quantum states. Quantization, i.e., the way the vacuum expectation values are obtained, is supposed to be FPI quantization. This construction of quantum theory is suitable for functional calculations, but there are still some difficulties, connected with it, left. Let us point out three of them.

A mathematically rigorous definition of FPI is missing due to the nonexistence of the corresponding functional measure. The usual definition is a perturbative one.[4]

The second problem lies in the fact that Feynman quantization, unlike canonical quantum field theory (QFT), has not the important property of unitary invariance. In Dirac's theory the commutation relations of canonically conjugated field variables, the Hamiltonian $H = H(\varphi, \pi)$, and equations of motion are defined without the use of any special basis in the state space. It is described by the representation theory in quantum mechanics. The corresponding unitary invariance of the Feynman integral is not known.

The last problem is somewhat hidden in the present state of the theory. The quantization based on FPI does not form a complete theory alone; it is rather a convenient auxiliary computational tool.[5] It needs well-established relations to canonical QFT. FPI does not give the physical interpretation of Green's functions, it is completely based on canonical QFT (LSZ formulas). There are situations such as quantization on a general curved spacetime (global Poincaré invariance is broken), quantization of the nonlinear σ model and auxiliary fields[6] (the space of quantum states is not well defined), and others, where the corresponding canonical QFT does not exist. Then the notions like a state of a system, an observable, etc., are missing.

In this paper, we introduce another quantization procedure which generalizes Feynman quantization. In this way we give a solution to the second problem, connected with FPI quantization and a possible recipe for curing the other two. The usual state space is too small for this new procedure. We introduced[7] a larger Fock space called Fock–Stueckelberg space (FS space) fitted for kinematical quantization. The FS space is the Fock extension of the space of a particle which is not constrained to be on the mass shell. Field operators $\phi_x, x \in M$ (= Minkowski space), which correspond to a classical field $\varphi(x)$, are defined in the FS space. Then the quantum action operator $\mathscr{A} = S[\phi_x]$ and a quantum propagator operator $\mathscr{K} = \exp i\mathscr{A}$ are defined ($S[\varphi(x)]$ is a classical action). This procedure depends on kinematics only, the conjugated momentum variable is not involved. The matrix element

$$\langle 0|\phi_{x_1} \cdots \phi_{x_n} \mathscr{K}|0\rangle = G_n(x_1, ..., x_n) \tag{1.1}$$

of the operator $\mathscr{K}$ is the usual Green's function.

Let us point out now the connection of this procedure with FPI quantization. The operators ϕ_x, $x \in M$, commute one with another and form a complete set (in Dirac's sense) in FS space $\mathscr{H}$. Hence we can find a basis (we call it Feynman basis) in which they are diagonal. Elements of this basis are parametrized by eigenvalues $\varphi(x)$ of the operators ϕ_x, $x \in M$. Writing the formula (1.1) in the Feynman basis we obtain the expression for Green's functions in FPI representation.

There still remain problems with a proper mathematical definition of the operators ϕ_x, $\mathscr{A}$, $\mathscr{K}$, but they are the same as in canonical QFT (the regularization and renormalization are needed). Kinematical quantization is not more pathological than the canonical one. The substantial pathology comes only with the introduction of the Feynman basis. Hence a part of difficulties with FPI can be overcome by writing all formulas using the Feynman integral in the operator form, which is independent of the concept of the functional integration.

We want to stress here that our operator approach to quantum theory does not rely on the Hamiltonian descrip-

0022-2488/83/040834-05$02.50

tion of the dynamics as does Dirac's quantum theory. Kinematical quantization is based (as FPI in final form) on classical action, which is the most fundamental advantage of FPI and functional methods. Relation of this "Lagrange" operator quantization and canonical "Hamilton" quantum theory, and the construction of the physical subspace in the FS space, will be given in the forthcoming paper.

The paper is organized as follows. In Sec. II we introduce the FS space and kinematical quantization. The Feynman basis, the vacuum wave function, and Feynman integrals are described in Sec. III. In Sec. IV the equation of motion for the operator $\mathscr{K}$ is derived. It is shown that the equation for $\mathscr{K}$ is a compact form of equations for Green's functions (the functional integration is not used in the derivation), and the standard perturbative expansion for their solution is established. The spectral or operator definition of FPI is given in Sec. V. In Sec. VI we discuss in brief the problem of symmetry breakdown, a representation of commutation relations, causality, and the domain of integration in FPI in the framework of the proposed formalism.

II. KINEMATICAL QUANTIZATION

The Stueckelberg's space $\mathscr{H}^{(1)}$ is the set of complex functions defined on the Minkowski space M which are normalized to the 4-interval

$$\mathscr{H}^{(1)} = \left\{ \psi \left| \int_M \right| \psi(x)|^2 d^4x < \infty \right\}. \qquad (2.1)$$

The FS space $\mathscr{H}$ is then the Fock space of $\mathscr{H}^{(1)}$; each element of it can be written in the form

$$|\psi\rangle = \{ \psi_n(x_1, ..., x_n) \}_{n=0}^{\infty}, \qquad (2.2)$$

where functions ψ_n are symmetric (we shall consider for simplicity only one Bose scalar field). The scalar product is defined by

$$\langle \psi | \psi' \rangle = \sum_{n=0}^{\infty} \int \psi_n^*(x_1, ..., x_n) \times \psi_n'(x_1, ..., x_n) d^4x_1 \cdots d^4x_n . \qquad (2.3)$$

Let us introduce a (distributional) creation operator[8] a_x^+, $x \in M$, by

$$a_f^+ = \int_M d^4x f(x) a_x^+ , \quad f \in L_2(M)$$

$$a_f^+ |\psi\rangle = \{ \mathrm{Sym} [\psi_n(x_1, ..., x_n) f(x_{n+1})] \}_{n=0}^{\infty}, \qquad (2.4)$$

with the 4-dimensional commutation relations

$$[a_x, a_y^+] = \delta^{(4)}(x-y). \qquad (2.4')$$

Thus we can write

$$|\psi\rangle = \sum_{n=0}^{\infty} \int dx_1 \cdots dx_n \psi_n(x_1 \cdots x_n)(n!)^{-1/2} \times a_{x_1}^+ \cdots a_{x_n}^+ |0\rangle, \qquad (2.5)$$

where $|0\rangle$ is the Fock vacuum, $a_x|0\rangle = 0$, $x \in M$. A function from $\mathscr{H}^{(1)}$ describes the 1-particle state outside the mass shell. The physical states form a (distributional) subspace of $\mathscr{H}^{(1)}$ defined by the dispersion relation; it can be seen immediately from the Fourier decomposition of such a function, the 4-momenta used in it are not constrained to be on the mass shell. Similarly vectors from the FS space describe virtual many-particle (many-time) states. These states are quantum field analogs of virtual trajectories $q_i(t)$ known from classical mechanics (they do not fulfill equations of motion). The operator a_x^+ is the creation operator outside the mass shell. The FS space is substantially larger than the usual Fock space.

We shall present kinematical quantization for the scalar field with the Lagrangian

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}' + \mathscr{L}_{ex} = \tfrac{1}{2}(\partial_\mu \varphi \partial_\mu \varphi - m^2\varphi^2) - (\lambda/4)\varphi^4 + J(x)\varphi \qquad (2.6)$$

(other fields can be treated in a similar way, see Ref. 7). The procedure consists of the following steps.

(i) To a classical field $\varphi(x)$ corresponds an operator-valued function (distribution) ϕ_x with another operator-valued distribution π_x satisfying the commutation relation

$$[\phi_x, \pi_y] = i\delta^{(4)}(x-y), \quad x, y \in M. \qquad (2.7)$$

(ii) The representation of this commutation relation must be chosen; we choose the Fock representation (i.e., no symmetry breakdown appears)

$$\phi_x = \frac{1}{(\sqrt{2})\mu_0}(a_x + a_x^+), \quad x \in M \qquad (2.8)$$

$$\pi_x = \frac{i\mu_0}{\sqrt{2}}(a_x^+ - a_x), \quad x \in M \qquad (2.8')$$

where a_x is the annihilation operator in the FS space and μ_0 is an arbitrary chosen scale of mass. The meaning of μ_0 will be clarified below.[9]

(iii) Substituting ϕ_x into the classical action,[10] we obtain a quantum operator $\mathscr{A}$ and a propagator operator $\mathscr{K}$:

$$\mathscr{A} = \int \mathscr{L}(\phi_x) d^4x, \quad \mathscr{K} = \exp i\mathscr{A}. \qquad (2.9)$$

(iv) We pick up the Fock vacuum $|0\rangle$ (defined by the condition $a_x|0\rangle = 0$) and define Green's functions as vacuum expectation values

$$G_n(x_1, ..., x_n) \equiv \langle 0 | \phi_{x_1} \cdots \phi_{x_n} \mathscr{K} | 0 \rangle. \qquad (2.10)$$

G_n is the n-point unnormalized Green's function. The equivalent way to say it is that the matrix element

$$\langle 0 | a_{y_1} \cdots a_{y_n} \mathscr{K} a_{x_1}^+ \cdots a_{x_n}^+ | 0 \rangle$$

is the full unnormalized transition amplitude of virtual particles from points $x_1, ..., x_n$ to points $y_1, ..., y_n$.

The operator $\mathscr{K}$ is then S matrix outside the mass shell; the physical S matrix can be obtained by contracting $\mathscr{K}$ with appropriate plane waves. Formula (1.1) is the analog of the well-known expression $\langle 0 | T[\hat{\varphi}(x_1) \cdots \hat{\varphi}(x_n) S] | 0 \rangle$ from the canonical QFT (in the interaction representation).

We shall now show by two different methods that kinematical quantization gives usual Green's functions—using the "formal" functional integration (in the next section) and the correct operator method (in Sec. IV).

III. THE FEYNMAN BASIS

The most direct way to "calculate" the operator $\mathscr{K}$ is to diagonalize it. The Feynman basis (just suitable for this purpose) diagonalizes the complete set $\{\phi_x; x\in M\}$ of commuting operators. We shall denote by $|\varphi\rangle$ (φ is a function on M) their common eigenvector with the eigenvalues

$$\phi_x|\varphi\rangle = \varphi(x)|\varphi\rangle, \quad x\in M. \tag{3.1}$$

The Feynman basis is hence parametrized by classical field configurations. The state $|\psi\rangle\in\mathscr{H}$ can be decomposed in the Feynman basis as

$$|\psi\rangle = \int d\varphi\, \psi(\varphi)|\varphi\rangle, \quad \psi(\varphi) = \langle\varphi|\psi\rangle. \tag{3.2}$$

Phases of vectors $|\varphi\rangle$ can be chosen in such a way that

$$\phi_x|\psi\rangle = \int d\varphi\, \varphi(x)\psi(\varphi)|\varphi\rangle, \tag{3.3}$$

$$\pi_x|\psi\rangle = \int d\varphi\left(-i\frac{\delta}{\delta\varphi(x)}\psi(\varphi)\right)|\varphi\rangle \tag{3.4}$$

hold. It follows from the commutation relation (2.7).

The relation (3.4) gives the phases of $|\varphi\rangle$'s in the same way as in the usual coordinate representation in quantum mechanics.

The Feynman representation is similar to the Schroedinger (or field) representation for a quantum field with one important difference, that φ's are defined on the whole M instead of $\mathbb{R}^3$. The completeness relations have the form

$$\mathbb{1} = \int d\varphi\,|\varphi\rangle\langle\varphi|, \quad \langle\varphi|\varphi'\rangle = \delta(\varphi|\varphi'), \tag{3.5}$$

where a δ-functional is defined by the condition $\int d\varphi'\delta(\varphi|\varphi')\psi(\varphi') = \psi(\varphi)$. Let us find the wave function of the Fock vacuum. We have

$$|0\rangle = \int d\varphi\psi_0(\varphi)|\varphi\rangle, \quad \psi_0(\varphi) = \langle\varphi|0\rangle. \tag{3.6}$$

The vacuum is annihilated by the operators a_x and thus in the Feynman representation we have the functional equation

$$\left(\varphi(x) + \mu_0^{-2}\frac{\delta}{\delta\varphi(x)}\right)\psi_0(\varphi) = 0. \tag{3.7}$$

The solution of it is

$$\psi_0(\varphi) = e^{-\mu_0^2/2}\int\varphi^2(x)d^4x \tag{3.8}$$

(where the normalization factor is supposed to be contained in the measure $d\varphi$).

The quantum action is diagonal in this representation $\mathscr{A}|\varphi\rangle = \mathscr{A}(\varphi)|\varphi\rangle$, where $\mathscr{A}(\varphi)$ is the classical action corresponding to the configuration $\varphi(x)$. The same is true for the operator $\mathscr{K}$:

$$\begin{aligned}\mathscr{K} &= \int d\varphi\,\mathscr{K}(\varphi)|\varphi\rangle\langle\varphi|,\\ \mathscr{K}(\varphi) &= \exp i\mathscr{A}(\varphi).\end{aligned} \tag{3.9}$$

The Green's functions can be expressed as

$$\begin{aligned}&\langle 0|\phi_{x_1}\cdots\phi_{x_n}\mathscr{K}|0\rangle\\ &= \int d\varphi\,\langle 0|\phi_{x_1}\cdots\phi_{x_n}\mathscr{K}|\varphi\rangle\langle\varphi|0\rangle\\ &= \int d\varphi\,\varphi(x_1)\cdots\varphi(x_n)\mathscr{K}(\varphi)|\psi_0(\varphi)|^2 \qquad (3.10)\\ &= \int d\varphi\,\varphi(x_1)\cdots\varphi(x_n)\exp\left[i\mathscr{A}(\varphi) - \mu_0^2\int\varphi^2(x)d^4x\right].\end{aligned}$$

This is the well-known Feynman integral formula for the Green's function in which $\epsilon = 2\mu_0^2$ plays clearly the role of the Feynman causal ϵ. Thus the limit $\mu_0^2\to 0$ should be performed after the calculation.

IV. EQUATIONS OF MOTION

In this section we shall use operator methods instead of the formal functional integrals. From formulas (2.4′) and (2.8) we obtain the relation

$$[a_x^+, e^{i\mathscr{A}}] = [a_x^+, i\mathscr{A}]e^{i\mathscr{A}}, \tag{4.1}$$

since the commutator

$$[a_x^+, i\mathscr{A}] = -\frac{i}{(\surd 2)\mu_0}\frac{\delta\mathscr{L}}{\delta\varphi}(\phi_x),$$

$$\frac{\delta\mathscr{L}}{\delta\varphi}(\phi_x) = -(\Box + m^2)\phi_x - \lambda\phi_x{}^3 + J(x)$$

commutes with the operator $\mathscr{A}$.

The formula (4.1) can be written in the equivalent form

$$\begin{aligned}&i(\surd 2)\mu_0(a_x^+\mathscr{K} + \mathscr{K}a_x)\\ &\quad = \left(\frac{\delta\mathscr{L}}{\delta\varphi}(\phi_x) + 2i\mu_0^2\phi_x\right)\mathscr{K};\end{aligned} \tag{4.2}$$

the right-hand side should be understood in the sense of distributions. This is the equation of motion. Multiplying it by the operators $\phi_{x_1}, \ldots, \phi_{x_n}$ from the right and taking the vacuum expectation value we obtain

$$\begin{aligned}&i\sum_{i=1}^{n}\delta(x - x_i)G_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)\\ &\quad = -(\Box_{(x)} + m^2 - 2i\mu_0^2)G_{n+1}(x, x_1, \ldots, x_n) \qquad (4.3)\\ &\qquad - \lambda G_{n+3}(x,x,x,x_1,\ldots,x_n) + J(x)G_n(x_1,\ldots,x_n).\end{aligned}$$

These are equations for Green's functions[11] already containing the Feynman epsilon: $m^2\to m^2 - i\epsilon$, $\epsilon = 2\mu_0^2$; thus the right boundary conditions are satisfied and the complete equivalence with the canonical QFT is proved.

Equation (4.2) (together with the condition $[\phi_x, \mathscr{K}] = 0$) may be considered as the defining equation of the operator $\mathscr{K}$ (instead of the formula $\mathscr{K} = \exp i\mathscr{A}$).

Equation (4.2) is in fact the operator form of the Schwinger equation. The operator $\mathscr{K}$ depends on the external field $J(x)$ and we have

$$\frac{\delta}{\delta J(x)}\mathscr{K} = \phi_x\mathscr{K}. \tag{4.4}$$

Substituting this relation into Eq. (4.2) and taking the vacuum expectation value we obtain the Schwinger equation for the Green's functional $G(J) = \langle 0|\mathscr{K}|0\rangle$,

$$\left[\frac{\delta\mathscr{L}}{\delta\varphi}\left(\frac{\delta}{\delta J(x)}\right) + 2i\mu_0^2\frac{\delta}{\delta J(x)}\right]G(J) = 0. \tag{4.5}$$

The standard perturbation expansion now follows easily. We shall write the action (2.9) and (2.6) in the form $\mathscr{A} = \mathscr{A}_0 + \mathscr{A}_I$, $\mathscr{A}_0 = \mathscr{A}(\lambda = 0)$, and we have similarly $\mathscr{K} = \mathscr{K}_0\mathscr{K}_I$. From the relation (4.4) we have $\mathscr{K} = (\exp i\mathscr{A}_I(\delta/\delta J(x)))\mathscr{K}_0$ and thus the formula

$$G(J) = \exp\left[i\mathscr{A}_I\left(\frac{\delta}{\delta J(x)}\right)\right]G_0(J) \tag{4.6}$$

holds for the Green's functional G. The free field Green's functional $G_0(J) = \langle 0|\mathscr{K}_0|0\rangle$ is given by the standard formula (see Sec. V), where the right boundary conditions are ensured by the Feynman term $2i\mu_0^2$ in Eq. (4.3).

V. THE SPECTRAL DEFINITION OF THE FEYNMAN INTEGRAL

Let us consider the Feynman integral in the form

$$\int \psi(\varphi) e^{-\mu_0^2 \int \varphi^2 dx} d\varphi. \tag{5.1}$$

The exponential term gives the regularization of the Feynman integral and defines, in a sense, the integration domain of it. The limit $\mu_0^2 \to 0$ should be performed at the end of all calculations. Let us suppose that we are able to define the operator $\psi(\phi)$, where the operators ϕ_x are replaced instead of $\varphi(x)$. Then we define the integral (5.1) to be equal to the vacuum expectation value of the operator $\psi(\phi)$:

$$\int \psi(\varphi) e^{-\mu_0^2 \int \varphi^2 dx} d\varphi \equiv \langle 0|\psi(\phi)|0\rangle.$$

As an example, we shall show how the Gaussian integral can be obtained in this way. Let us suppose generally that $\psi(\varphi) = \exp i\mathscr{A}(\phi)$ and that the operator $\mathscr{A}(\varphi)$ can be defined. Then the relation (4.1) holds with the commutator

$$[a_x, i\mathscr{A}] = \frac{i}{(\surd 2)\mu_0} \frac{\delta \mathscr{A}}{\delta \varphi(x)} (\phi). \tag{5.2}$$

The formula

$$i(\surd 2)\mu_0 (a_x^+ \psi(\phi) + \psi(\phi) a_x)$$
$$= \left(\frac{\delta \mathscr{A}}{\delta \varphi(x)} (\phi) + 2i\mu_0^2 \phi_x \right) \psi(\phi) \tag{5.3}$$

[of the type of (4.2)] holds. Assuming that the functional $\mathscr{A}(\varphi)$ contains the term $\int \varphi(x) J(x) dx$, the Schwinger equation

$$\left(\frac{\delta \mathscr{A}}{\delta \varphi(x)} \left(\frac{\delta}{\delta J(x)} \right) + 2i\mu_0^2 \frac{\delta}{\delta J(x)} \right) \langle 0|\psi(\phi)|0\rangle = 0 \tag{5.4}$$

can be derived. In the Gaussian case

$$\mathscr{A}(\varphi) = \frac{1}{2} \int \varphi(x) A_{xy} \varphi(y) dx dy + \int \varphi(x) J(x) dx, \tag{5.5}$$

and the Schwinger equation reads

$$\left(\int (A_{xy} + 2i\mu_0^2 \delta_{xy}) \frac{\delta}{\delta J(y)} dy + J(x) \right) \langle 0|\psi(\phi)|0\rangle = 0. \tag{5.6}$$

The matrix $(A_{xy} + 2i\mu_0^2 \delta_{xy})$ can be inverted (because of the regularization term $2i\mu_0^2 \delta_{xy}$) and then this functional equation can be easily integrated to

$$\langle 0|\psi(\phi)|0\rangle = N \exp\left[- \int \tfrac{1}{2} J(x) \Delta_{xy} J(y)\, dx dy \right], \tag{5.7}$$

where Δ is the inverse of the mentioned matrix and N is an integration constant.

We conjecture that all meaningful calculations using the Feynman integral can be "translated" into the operator language in a similar manner to the spectral definition of the Feynman integral. Moreover, we conjecture that if the correct meaning of the corresponding operator $\psi(\phi)$ cannot be given, the Feynman integral cannot be defined in any rigorous way. Thus in this approach the Feynman integral is only a representation of the operator $\mathscr{K}$ (or its matrix elements) in a special coordinate basis, as is the Schroedinger equation in quantum mechanics.

From our point of view the question of the existence of the "correct integral" (5.1) is, in fact, that of the existence and the meaning of the Feynman basis $\{ |\varphi\rangle \}$, and this is a very interesting but difficult mathematical problem, which we do not solve here.

VI. DISCUSSION AND CONCLUSIONS

In this paper we have proposed a possible operator formalism leading to the functional calculus and to FPI. The basic ingredients in this approach are operators ϕ_x, π_x with commutation relations (2.7). The conjugated momentum π_x is only an auxiliary object serving for defining the operator character of ϕ_x, i.e., ϕ_x and π_x define in some combination the creation and the annihilation operators of virtual states of a quantum field (that define Hermitian conjugation on the FS space). Kinematical quantization represents then a transition from a classical virtual field to an operator field expressed in a combination of the creation and the annihilation operators on the FS space. Specifying the combination [e.g., by Eq. (2.8)] we pick up the representation of the relation (2.7) or a subspace of the whole functional space, i.e., we choose the integration domain in FPI. The choice of a representation of (2.7) and the domain of integration in the functional space are related by the definition of the vacuum state since the vacuum is annihilated by the operator a_x and $|\langle 0|\varphi\rangle|^2$ defines the weight of the functional measure in Eq. (3.10).

Freedom in a representation of the relation (2.7) gives one a possibility to incorporate quantization around classical solutions or spontaneous symmetry breaking. We can introduce a new parameter φ_0 (which we shall specify from the condition that the effective potential evaluated in φ_0 is minimal[1]) and define

$$\phi_x = \frac{1}{(\surd 2)\mu_0} (a_x + a_x^+) - \varphi_0, \tag{6.1}$$

$$\pi_x = i \frac{\mu_0}{(\surd 2)} (a_x^+ - a_x). \tag{6.1'}$$

Then the formula (3.8) turns out to be

$$\psi_0'(\varphi) = \exp\left[- \frac{\mu_0^2}{2} \int (\varphi - \varphi_0)^2 d^4x \right]. \tag{6.2}$$

Equation (6.2) changes the domain of the functional integration in (3.10) (it fixes the values of classical fields at infinity) and makes clear the way to obtain symmetry breakdown in the FPI quantization scheme.[12] From the equation of motion (4.3), the meaning of the parameter μ_0^2 is clear—it fixes the boundary values of the free propagator Δ, i.e., it specifies causality. Thus we see that a change in a representation of (2.8) leading to symmetry breakdown is equivalent to breaking of the Feynman causality prescription defined by Eq. (3.8). From this example we can see advantages of this operator formalism. In it, we naturally incorporate symmetry breaking and its connection to causality and to FPI.

From the construction of quantum theory in the proposed way, we see that the quantization procedure here is larger than the canonical one since we do not have specified the representation of commutation relation (2.7) from the beginning. The freedom in a choice of a representation of (2.7) is removed postulating that any realistic theory [i.e., any acceptable representation of (2.7)] must have a solution (vacuum) minimizing the effective potential.

The Fock-Stueckelberg space, the kinematical operators ϕ_x, and the propagator operator $\mathscr{K}$ are the basic tools for coordinate (basis)—free description of the Feynman quantization technique. From it we can deduce that the lack of a rigorous definition of FPI can be divided into two parts. The first one, before we introduce the Feynman basis, is connected with correctness of multiplication of kinematical operators and is essentially equivalent to the corresponding difficulties in canonical QFT (i.e., the problem of renormalization). The second part lies in the concept of the Feynman functional basis which is highly distributive (singular), and brings itself new mathematical complications rendering the meaning of FPI more perplexed. The spectral definition, here suggested, could help in resolving this second difficulty.

We expect that this line of considerations could clarify the true meaning of FPI, formulate Feynman quantization independently on canonical formalism, and be useful in the unification of QFT and General Relativity.

---

[1] S. Coleman and E. Weinberg, Phys. Rev. D **7**, 1888 (1973).

[2] J. C. Taylor, *Gauge Theories of Weak Interactions* (Cambridge U.P., Cambridge, 1976).

[3] S. W. Hawking, in *Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1979).

[4] See, e.g., L. D. Faddeev and A. A. Slavnov, *Gauge Fields, Introduction to Quantum Theory* (Benjamin–Cummings, New York, 1980).

[5] S. Coleman, in *Laws of Hadronic Matter*, edited by A. Zichichi (Academic, New York, 1975).

[6] For auxiliary fields or Lagrange multipliers see, e.g., T. Eguchi, Phys. Rev. D **17**, 611 (1978); see also D. J. Gross and A. Neveu, *ibid.* **10**, 3235 (1974).

[7] J. Souček, V. Janiš, and V. Souček, preprint KMA 1/1980.

[8] The mathematics used in the paper is on a formal level, but it can be put on a rigorous basis using the spectral theory of infinite many variable functions; e.g., Ju. M. Berezansky, *Selfadjoint Operators in Functional Spaces with Infinite Many Variables* (in Russian) (Naukova Dumka, Kiev, 1978).

[9] The operator ϕ_x is known from another context as Segal's field operator, cf. M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1975), Vol. 2.

[10] Expression $\partial_\mu \phi_x$ is understood in the sense of Schwartz' generalized functions.

[11] These equations define Green's functions only formally since they are not renormalized. But they are the starting point for defining the perturbative expansion or renormalized equations.

[12] Symmetry breaking in FPI was described by H. Matsumoto, N. J. Papastamatiou, and H. Umezawa, Phys. Lett. B **46**, 73 (1973).

Weak dispersion-free states and the hidden variables hypothesis

Pavel Pták

Technical University of Prague, 166 27-Prague 6, Suchbatarova 2, Czechoslovakia

(Received 22 December 1981; accepted for publication 12 February 1982)

We investigate dispersion-free states which are additive only on the pairs containing a central element (central-absolutely compatible). We show that any logic possesses plenty of such states, in fact, as many as a certain Boolean algebra. The latter result matches the hidden variables conjecture.

PACS numbers: 03.65. — w

According to the proponents of the hidden variables theory, there are certain hidden variables in the quantum mechanical system which would accurately determine the behavior of the system—if we knew them. The probabilistic character of the occurrences of events would therefore disappear (see Refs. 1–3). In the mathematical model, the presence of hidden variables would yield the existence of dispersion-free states on the logic in question, which means that the existence of such states which would have only the values 0 and 1 (see Ref. 4 for more detailed exposition). It is known that, e.g., that the Hilbert space logic $L(H)$ for quantum mechanics does not admit hidden variables in the usual sense, and there are other examples (see Refs. 5–9).

We show in this paper that if we weaken the notion of a state, we obtain the dispersion-free states on [any logic L. A state in our sense is a mapping $s:L\rightarrow\{0,1\}$ such that: (1) $s(a')=1-s(a)$ for any $a\in L$; (2) if $s(b)=0$ and $a<b$, then $s(a)=0$; (3) $s(a\vee b)=s(a)+s(b)$ whenever $a<b'$ and b is a central element. Any dispersion-free state in the usual sense fulfills the latter conditions, and if L is a Boolean algebra, then our "weak" state means exactly the state in the usual sense. Obviously, the third condition of our definition vanishes as soon as L has a trivial center (e.g., for the Hilbert space logic). There are, however, many important logics which have considerably rich centers and yet do not have any dispersion-free states. If we, for instance, take $L=L(H)^{\omega_0}$, where $L(H)^{\omega_0}$ means the countable power of $L(H)$ (see Refs. 4 and 10), we obtain a logic with the center isomorphic to the Boolean algebra of all subsets of a countable set. One can show easily (see Ref. 10) that such a logic possesses no dispersion-free states in the usual sense. On the other hand, for the logics with rich centers, the suggested generalization of a state is meaningful because the intrinsic structure of the logic is very much involved.

The terminology and basic facts are taken from Refs. 4 and 11. The content of the paper overlaps slightly with Refs. 11 and 12.

Let us first introduce the notions which are used throughout the paper.

Definition 1: A logic is a partially ordered set $(L,<)$ with the least and greatest elements 0 and 1 and with a unitary operation $'$ satisfying the properties:

(i) $(a')'=a$ for any $a\in L$;

(ii) $a\vee a'=1$ for any $a\in L$, where the symbol $\vee$ (and dually $\wedge$) means the lattice-theoretic operation induced by $<$;

(iii) if $a<b, a,b\in L$, then $a'>b'$ and $b=(a\vee(b\wedge a')$.

In what follows, we shall reserve the letter L for a logic. One can show easily (see Ref. 11) that if $a<b'$, then $a\vee b$ and $a\wedge b$ exist in L.

Definition 2: Two elements $a,b\in L$ are called compatible (abbreviated $a\leftrightarrow b$) if there are elements $c,d,e\in L$ such that $c<d'$, $d<e'$, $e<c'$ and $a=c\vee d$, $b=c\vee e$.

Obviously, if $a\leftrightarrow b$, then $a\vee b=c\vee d\vee e$ and $a\wedge b=c$.

Definition 3: The center of a logic L is the set of all $a\in L$ such that $a\leftrightarrow b$ for any $b\in L$. We denote the center of L by $C(L)$.

Proposition 1: The set $C(L)$ constitutes a Boolean subalgebra of L. The logic L is a Boolean algebra if and only if $L=C(L)$.

Proof: See Refs. 11, 13, or others.

Definition 4: A weak (dispersion-free) state is a mapping $s:L\rightarrow\{0,1\}$ such that:

(i) $s(a')=1-s(a)$ for any $a\in L$;

(ii) if $s(b)=0$ and $a<b$, then $s(a)=0$;

(iii) $s(a\vee b)=s(a)+s(b)$ whenever $a<b'$ and $b\in C(L)$.

Proposition 2: Suppose that s is a weak state on L. If $b\in C(L)$, then $s(a\vee b)<s(a)+s(b)$ for any $a\in L$.

Proof: Since we can write $a\vee b=(a\wedge b')\vee b$, we have $s(a\vee b)=s(a\wedge b')+s(b)$. If $s(a\vee b)=1$ and $s(a)=0$, then $s(a\wedge b')=0$ and therefore $s(b)=1$. The rest is obvious.

Theorem 1: If $a\in L$, $a\neq 0$, then there exists a weak state on L such that $s(a)=1$.

Proof: We shall first state a lemma. To simplify the argument, let us introduce an auxiliary notion. A subset I of L is said to be absorbing if I fulfills the following properties:

(1) if $a\in I$ and $b<a$, then $b\in I$;

(2) if $a\in I$ and $b\in I\cap C(L)$, then $a\vee b\in I$;

(3) if $a\in I$ then $a'\notin I$.

Lemma 1: Let I be an absorbing subset of L and let $I\cap\{c,c'\}=\emptyset$ for some element $c\in L$. Then there exists an absorbing set J which contains the set $I\cup\{c\}$.

Proof of Lemma 1: Let us set $I_c=\{x\in L\mid x<c\}$. Put $J=\{x\in L\mid$ there are elements $m\in I\cap C(L)$, $n\in I_c\cap C(L)$ and $k\in J\cup I_c$ such that $x<m\vee n\vee k\}$. Obviously, $c\in J$. We shall prove now that J is absorbing. If $a\in J$ and $b<a$, then $b\in J$ by the definition of J. Suppose that $a\in J$ and $b\in J\cap C(L)$. Then $a<m\vee n\vee k$ and $b<p\vee r\vee s$, where $m, p\in I\cap C(L)$, $n,r\in I_c\cap C(L)$ and $k, s\in J\cup I_c$. Observe first that there exist two elements $u,v\in L$, $u\in I\cap C(L)$, $v\in I_c\cap C(L)$ such that $b<u\vee v$. To see that, let us notice that we may assume the elements p,r,s to be mutually orthogonal [for $p,r\in C(L)$, therefore

$p \vee r \vee s = p \vee (r \wedge p') \vee t$ for a $t \in L$ orthogonal to $p \vee r$, and hence $t < s$]. Since $b \in C(L)$, we have the equality $b = (p \wedge b) \vee (r \wedge b) \vee (s \wedge b)$ and therefore $s \wedge b \in C(L)$. Suppose that $s \in I$. Then $s \wedge b \in I \cap C(L)$, and we put $u = (p \wedge b) \vee (s \wedge b), v = r \wedge b$. (The case $s \in I_c$ argues similarly.) We see that $a \vee b < (m \vee n \vee k) \vee (u \vee v) = (m \vee u) \vee (n \vee v) \vee k \in J$.

Let us check the condition (3). Suppose that $a \in J$ and $a' \in J$. We may assume without any loss of generality that $a < m \vee k, a' < n \vee h$, where $m \in I \cap C(L)$, $n \in I_c \cap C(L)$, and $k \in I_c$, $h \in I$, $m < k'$, $n < h'$. Since m,n are central, we may write $a = (m \wedge a) \vee (m' \wedge a)$, $a' = (n \wedge a') \vee (n' \wedge a')$. It follows that $1 = a \vee a' = (m \wedge a) \vee (m' \wedge a) \vee (n \wedge a') \vee (n' \wedge a')$ and the right side is a supremum of four mutually orthogonal elements. Since $a' \leqslant n \vee h$, we obtain that $a' \wedge n' < (n \vee h) \wedge n' = h \wedge n' < h$ and therefore $a' \wedge n' \in I$. Analogously $m' \wedge a \in I_c$. We obtain that $(m' \wedge a) \vee (n \wedge a') \leqslant c$ and therefore $((m' \wedge a) \vee (n \wedge a'))' = (m \wedge a) \vee (n' \wedge a') \geqslant c'$. Since $(m \wedge a) \vee (n' \wedge a') \in I$, it follows that $c' \in I$, which is a contradiction. The proof of Lemma 1 is finished.

Let us continue with the proof of Theorem 1. Denote by $\mathscr{A}$ the collection of all absorbing subsets of L which contain the element a'. Since $I_{a'} = \{x \in L \mid x < a'\}$ belongs to $\mathscr{A}$, we see that $\mathscr{A} \neq \emptyset$. It is obvious (Zorn lemma) that $\mathscr{A}$ has maximal elements if we order $\mathscr{A}$ by inclusion. Take a maximal element of $\mathscr{A}$, some A. By Lemma 1, if $c \in L$, then either c or c' belongs to A. We may thus define a mapping $s: L \to \{0,1\}$ by setting $s(x) = 0$ if and only if $x \in A$. One can check easily that s is a weak state with the required properties. The proof of Theorem 1 is finished.

The latter theorem asserts that any logic possesses weak states. One must notice that the proof went via a nonconstructive argument (Zorn lemma). It would be desirable to remove any such reasonings when dealing with the physical theories but we doubt that this is possible.

The following result refines Theorem 1 and connects together weak states on logics with the states on Boolean algebras. Let us denote by $\mathscr{S}(L)$ the set of all weak (dispersion-free) states on L and recall that a weak state on a Boolean algebra is a state.

Theorem 2: Let L be a logic. Then there is a mapping $f: L \to B$ to a Boolean algebra B such that:

(i) $f(0) = 0$;

(ii) $f(a') = f(a)'$ for any $a \in L$;

(iii) $f(a) < f(b)$ if and only if $a < b$;

(iv) $f(a \vee b) = f(a) \vee f(b)$ whenever $a \in L$, $b \in C(L)$;

(v) if $s \in \mathscr{S}(B)$, then $sf \in \mathscr{S}(L)$ and the latter assignment is injective.

Proof: Let B be the Boolean algebra of subsets of $\mathscr{S}(L)$ generated by the sets of the type $A_a = \{s \in \mathscr{S}(L) | s(a) = 1\}$, $a \in L$. Put $f(a) = A_a$, $a \in L$. The properties (i) and (ii) are then evidently fulfilled. The property (iii) requires one to show that $f(a) < f(b)$ implies $a < b$. We employ the method which we have used for proving Theorem 1. Suppose that $a \nless b$. Consider the set $\mathscr{A}_{a,b}$ of all absorbing sets which contain b but do not contain a. Take a maximal element of $\mathscr{A}_{a,b}$ and denote it by D. According to Lemma 1, if $c \in L$, then either c or c' belongs to D. The weak state $s: L \to \{0,1\}$ such that $s(x) = 0$ if and only if $x \in D$ then belongs to $f(a)$ but does not belong to $f(b)$. Therefore, $f(a) \nless f(b)$ and the statement (iii) follows.

The statement (iv) follows immediately from Proposition 2, for if $s(a \vee b) = 1$ for $a \in L$, $b \in C(L)$, then either $s(a) = 1$ or $s(b) = 1$.

Finally, if $s \in \mathscr{S}(B)$ then $sf \in \mathscr{S}(L)$ according to the latter statement (iv). It is easy to see that if two dispersion-free states on a Boolean algebra B agree on the generators, they have to agree on the entire B. This yields that the assignment $s \to sf$ is injective. The proof of Theorem 2 is finished.

[1] D. Bohm and J. Bub, "A Proposed Solution of the Measurement Problem in Quantum Mechanics by Hidden Variables," Rev. Mod. Phys. **38**, 453–69 (1966).

[2] D. Bohm and J. Bub, "A Refutation of the Proof by Jauch and Piron That Hidden Variables Can Be Excluded in Quantum Mechanics," Rev. Mod. Phys. **38**, 470–5 (1966).

[3] A. Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Reality Be Considered Complete?," Phys. Rev. **47**, 777–80 (1935).

[4] S. Gudder, *Stochastic Methods in Quantum Mechanics* (Elsevier-North-Holland, New York, 1979).

[5] V. Alda, "On 0 – 1 measures for projectors, II," Apl. Mat. **26**, 57–8 (1981).

[6] R. Greechie, "Orthomodular lattices admitting no states," J. Comb. Theory **10**, 1198–32 (1971).

[7] J. Jauch and C. Piron, "Can hidden variables be excluded from quantum mechanics?" Helv. Phys. Acta **36**, 827–37 (1965).

[8] F. Shultz, "A characterization of state spaces of orthomodular lattices," J. Comb. Theory A **17**, 317–28 (1974).

[9] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P. Princeton, N.J., 1944).

[10] V. Maňasová and P. Pták, "On states on the product of logics," Int. J. Theor. Phys. (to be published).

[11] N. Zierler and M. Schlessinger, "Boolean embeddings of orthomodular sets and quantum logics," Duke J. Math. **32**, 251–62 (1965).

[12] P. Pták, "Categories of orthomodular posets," submitted for publication.

[13] J. Brabec and P. Pták, "On compatibility in quantum logics," Found. Phys. (to be published).

A generalization of Cohen's theorem

Roman Groblicki
Department of Mathematics, Monash University, Clayton, Victoria, Australia 3168

(Received 10 December 1981; accepted for publication 19 February 1982)

The theorem due to L. Cohen, which implies that quantum mechanics cannot be formulated as a stochastic theory in phase space, is generalized. The assumption that the phase-space representatives of the density operators satisfy the quantum mechanical marginals is replaced by the weaker condition of Ω-representability.

PACS numbers: 03.65.Ca

I. INTRODUCTION

The possibility of expressing quantum mechanical expectation values as averages over phase-space distribution functions has been widely discussed. The majority of such representations of quantum mechanics considered in the literature[1–4] can be characterized by two mappings, one which maps density operators $\hat{\rho}$ into so-called quasidistribution functions $f(q,p)$ on phase space, and the other which maps operators $\hat{g}$ representing observables into their phase-space counterparts g. The basic requirement imposed on these maps is

$$\mathrm{Tr}(\hat{\rho}\hat{g}) = \int\int f(q,p)\, g(q,p)\, dq\, dp\,. \qquad (1.1)$$

Among other specific conditions they are usually also taken to be one-to-one and linear.

Now Wigner's theorem[5] states that for a certain class of such representations in which the quasidistribution functions satisfy the quantum mechanical marginals, these functions cannot in general be true probability distributions, i.e., they take on some negative values. This is often taken[6,7] as demonstrating the impossibility of formulating quantum mechanics as a classical stochastic theory in phase space. A similar conclusion based on similar assumptions can be provided by Cohen's theorem.[2,6] He considers the phase-space representative of the observable $F(\hat{g})$ and demonstrates that in general it is not $F(g)$.

Both theorems are based on specific assumptions about phase-space representations, the foremost being linearity and the requirement that quasidistribution functions satisfy quantum mechanical marginals. Both assumptions have been criticized[8–10], and this leads one to question the general conclusion drawn from these theorems. For it might be the case that in dropping either or both of these assumptions, Cohen's and Wigner's theorems would not hold in general, so that a representation could be found which did formulate quantum mechanics as a classical stochastic theory. No one seems to have considered dropping the assumption of linearity, due probably to mathematical difficulties. However, in the case of the quantum mechanical marginals condition it is well known[4] that there are many (linear) representations with non-negative distribution functions. This condition is thus essential for Wigner's theorem. On the other hand, this is not the case of Cohen's theorem. It is the purpose of this paper to show that Cohen's theorem is actually true for the large class of all Ω-representations (which encompass all those of practical interest). Thus, the possibility of a classical stochastic formulation of quantum mechanics is left open only to nonlinear and linear non-Ω-representations.

In Sec. II I give a brief review of the phase-space representations of quantum mechanics as formulated by Srinivas and Wolf.[4] These, termed Δ-representations, are the most general kind considered in the literature. Finally, in Sec. III Cohen's theorem and its generalization are stated in this framework, and the latter proved.

II. PHASE-SPACE REPRESENTATIONS OF QUANTUM MECHANICS

In the following, the notation and approach of Srinivas and Wolf[4] to the phase-space representation of quantum mechanics is adopted. Attention will be restricted to a system consisting of one particle, and hence observables will be functions of the usual position and momentum operators (these are denoted by $\hat{q},\hat{p}$ but it should be emphasized that they are not necessarily operators which correspond via some representation to the phase-space functions q,p, as the notation of Srinivas and Wolf may suggest).

Each Δ-representation is characterized by a pair of operators $\Delta(\hat{q},\hat{p};q,p)$, $\bar{\Delta}(\hat{q},\hat{p};q,p)$ (parametrized by q,p), which satisfy

$$\Delta^{+}(\hat{q},\hat{p};q,p) = \Delta(\hat{q},\hat{p};q,p)\,, \qquad (2.1)$$

$$\mathrm{Tr}[\bar{\Delta}(\hat{q},\hat{p};q,p)\Delta(\hat{q},\hat{p};q',p')] = \delta(q-q')\delta(p-p')\,, \qquad (2.2)$$

$$\int\int \Delta(\hat{q},\hat{p};q,p)dqdp = I\,. \qquad (2.3)$$

The mapping (as outlined in Sec. I) which relates functions in phase space to operators representing observables is then constructed by postulating that the function $\delta(q-q')\delta(p-p')$ of q,p be mapped to the operator $\Delta(\hat{q},\hat{p};q',p')$ (for all q',p'). The images [denoted by $\hat{g}(\hat{q},\hat{p})$] of other functions $g(q,p)$ are then obtained by the assumption of linearity,

$$\hat{g}(\hat{q},\hat{p}) = \int\int g(q',p')\Delta(\hat{q},\hat{p};q',p')dq'dp'\,. \qquad (2.4)$$

The invertibility of this map is assured by (2.2), which explicitly gives

$$g(q,p) = \mathrm{Tr}[\hat{g}(\hat{q},\hat{p})\bar{\Delta}(\hat{q},\hat{p};q,p)]\,. \qquad (2.5)$$

On the other hand, the mapping which relates density opera-

0022-2488/83/040841-03$02.50

tors $\hat{\rho}$ to quasidistribution functions $f(q,p)$ is defined by

$$f(q,p) = \mathrm{Tr}[\hat{\rho}\Delta(\hat{q},\hat{p};q,p)]\,, \tag{2.6}$$

likewise it is linear and invertible.

The basic relation (1.1) then follows directly from (2.5), (2.6), and (2.2).

Now (2.1) is the "reality condition" which ensures real functions are mapped onto self-adjoint operators and vice-versa (for both mappings of the representation), while in the case of the mapping given by (2.4), (2.3) implies that the identity function corresponds to the unit operator [this need not be the case for the mapping given by (2.6)]. Details can be found in the cited paper.[4]

An important subclass of the Δ-representations is the class of Ω-representations. These have been studied in detail by Agarwal and Wolf.[3] They are generated by taking a complex valued function Ω from phase space satisfying

$$\Omega(0,0) = 1 \tag{2.7}$$

and

$$\Omega^*(-\xi,-\eta) = \Omega(\xi,\eta) \tag{2.8}$$

(Ω is also usually assumed to be a boundary value of an entire analytic function of two complex variables [3(a)]), and setting

$$\Delta_\Omega(\hat{q},\hat{p};q,p) = \frac{1}{(2\pi)^2}\int\int \Omega(\xi,\eta)e^{i[\xi(\hat{q}-q)+\eta(\hat{p}-p)]}d\xi\, d\eta, \tag{2.9}$$

$$\overline{\Delta}_\Omega(\hat{q},\hat{p};q,p) = 2\pi\hbar\Delta_{\tilde{\Omega}}(\hat{q},\hat{p};q,p)\,, \tag{2.10}$$

where

$$\tilde{\Omega}(\xi,\eta) = [\Omega(-\xi,-\eta)]^{-1}\,.$$

In this case formula (2.4) reduces to

$$\hat{g}(\hat{q},\hat{p}) = \frac{1}{(2\pi)^2}\int\int\int\int g(q,p)\Omega(\xi,\eta)e^{i[\xi(\hat{q}-q)+\eta(\hat{p}-p)]}\times d\xi d\eta dq dp \tag{2.11}$$

and formula (2.5) to

$$g(q,p) = \frac{\hbar}{2\pi}\int\int \mathrm{Tr}(\hat{g}e^{-i[\xi\hat{q}+\eta\hat{p}]})[\Omega(\xi,\eta)]^{-1}e^{i[\xi q+\eta p]}\times d\xi\, d\eta\,. \tag{2.12}$$

In Ref. 4, $\hbar$ has been inadvertently omitted from the right-hand side of (2.10) and (2.12).

Now, condition (2.7) is introduced so that the identity operator be mapped via (2.12) to the unit function, and (2.8) is required so that real-valued functions on phase space are mapped to self-adjoint operators and vice-versa, i.e., (2.7) and (2.8) are, respectively, special cases of (2.3) and (2.1).

It is interesting to note that these two conditions will play no part in the proof of the generalized Cohen's theorem (Sec. III).

III. COHEN'S THEOREM AND ITS GENERALIZATION

Ω-representations are still a more general class than the one considered by Cohen.[2,6] However, if we impose the condition

$$\Omega(\xi,0) = \Omega(0,\eta) = 1\,, \tag{3.1}$$

the resulting class of Ω-representations is identical to his [$\Omega(\xi,\eta)$ is $f(\theta,\tau)$ in Cohen's notation[6,7]]. Which means, as demonstrated by Cohen,[2,6] that in the context of Ω-representations, (3.1) is equivalent to

(i) For all density operators $\hat{\rho}$ and their phase-space representatives f,

$$\int f(q,p)dp = \langle q|\hat{\rho}|q\rangle\,,$$

$$\int f(q,p)dq = \langle p|\hat{\rho}|p\rangle$$

(i.e., the quasidistribution functions are required to satisfy the quantum mechanical marginals).

Noting the above, Cohen's theorem can then be stated as

Theorem 1 (Cohen's theorem[2,6]): There exists no Ω-representation satisfying (i) such that (ii). For arbitrary real-valued functions g on $\mathbb{R}^2$, if g is the phase-space representative of the observable $\hat{g}$, then g^2 is the representative of $\hat{g}^2$.

Now the generalization of Theorem 1 consists in simply leaving condition (i) out, i.e.,

Theorem 2 (Generalization of Theorem 1): There exists no Ω-representation which satisfies (ii).

Proof: As a preliminary we need the following simple lemma:

Lemma: If $R(q,p;q_1,p_1,q_2,p_2)$ is a complex valued function from $\mathbb{R}^2$ satisfying

(iii) For arbitrary functions g from $\mathbb{R}^2$ into $\mathbb{R}$

$$[g(q,p)]^2 = \int\int\int\int R(q,p;q_1,p_1,q_2,p_2)g(q_1,p_1)g(q_2,p_2)\times dq_1dp_1dq_2dp_2\,.$$

Then

$$R(q,p;q_1,p_1,q_2,p_2) + R(q,p;q_2,p_2,q_1,p_1) - 2\delta(q-q_1)\delta(p-p_1)\times\delta(q-q_2)\delta(p-p_2) = 0\,. \tag{3.2}$$

Proof: Interchanging q_1 with q_2 and p_1 with p_2 in (iii) demonstrates that $R(q,p;q_2,p_2,q_1,p_1)$ is also a solution of (iii). But clearly $\delta(q-q_1)\delta(p-p_1)\delta(q-q_2)\delta(p-p_2)$ is likewise a solution of (iii), hence

$$\int\int\int\int [R(q,p;q_1,p_1,q_2,p_2) + R(q,p;q_2,p_2,q_1,p_1) - 2\delta(q-q_1)\delta(p-p_1)\delta(q-q_2)\delta(p-p_2)]\times g(q_1,p_1)g(q_2,p_2)dq_1dp_1dq_2dp_2 = 0$$

for arbitrary functions g, which immediately gives (3.2), Q.E.D.

The proof of Theorem 2 is by contradiction. So suppose we have an Ω-representation satisfying (ii), characterized by the function $\Omega(\xi,\eta)$. It can be shown[4] that if $g\otimes h$ is the phase-space representative used to calculate the expectation values of the operator $\hat{g}\hat{h}$, then

$$g(q,p)\otimes h(q,p) = \int\int\int\int R_\Omega(q,p;q_1,p_1,q_2,p_2)g(q_1,p_1)h(q_2,p_2)\times dq_1dp_1dq_2dp_2\,, \tag{3.3}$$

where

$$R_\Omega(q,p;q_1,p_1,q_2,p_2)$$
$$=\frac{1}{(2\pi)^4}\int\int\int\int[\Omega(\xi_1+\xi_2,\eta_1+\eta_2)]^{-1}\Omega(\xi_1,\eta_1)\Omega(\xi_2,\eta_2)$$
$$\times e^{-(i\hbar/2)(\xi_1\eta_2-\xi_2\eta_1)}\, e^{i[\xi_1(q-q_1)+\eta_1(p-p_1)+\xi_2(q-q_2)+\eta_2(p-p_2)]}$$
$$\times d\xi_1 d\eta_1 d\xi_2 d\eta_2\,. \tag{3.4}$$

in Ref. 4, $1/(2\pi)^4$ has been inadvertently left out and the sign changed in the first exponential term.

Now noting that condition (ii) requires $g\otimes g=g^2$ for all functions, it follows using (3.3) that $\Omega(\xi,\eta)$ satisfies

$$[g(q,p)]^2$$
$$=\int\int\int\int R(q,p;q_1,p_1,q_2,p_2)g(q_1,p_1)g(q_2,p_2)$$
$$\times dq_1dp_1dq_2dp_2 \tag{3.5}$$

for arbitrary $g(q,p)$. But this is just the hypothesis of the lemma in relation to R_Ω, hence

$$R_\Omega(q,p;q_1,p_1,q_2,p_2)+R_\Omega(q,p;q_2,p_2,q_1,p_1)-2\delta(q-q_1)$$
$$\times\delta(p-p_1)\delta(q-q_2)\delta(p-p_2)=0\,. \tag{3.6}$$

From this we can derive a contradiction. To do this write the left-hand side of (3.6) explicitly in terms of $\Omega(\xi,\eta)$. Noting that by interchanging ξ_1 with ξ_2 and η_1 with η_2 in (3.4), $R_\Omega(q,p;q_2,p_2,q_1,p_1)$ is just $R_\Omega(q,p;q_1,p_1,q_2,p_2)$ with $e^{-(i\hbar/2)(\xi_1\eta_2-\xi_2\eta_1)}$ replaced by $e^{(i\hbar/2)(\xi_1\eta_2-\xi_2\eta_1)}$, (3.6) thus becomes

$$\frac{1}{(2\pi)^4}\int\int\int\int 2\Big\{[\Omega(\xi_1+\xi_2,\eta_1+\eta_2)]^{-1}\Omega(\xi_1,\eta_1)\Omega(\xi_2,\eta_2)$$
$$\times\cos\left(\frac{\hbar}{2}(\xi_2\eta_1-\xi_1\eta_2)\right)\ -1\Big\}$$
$$\times e^{i[\xi_1(q-q_1)+\eta_1(p-p_1)+\xi_2(q-q_2)+\eta_2(p-p_2)]}\,d\xi_1d\eta_1d\xi_2d\eta_2$$
$$=0\,.$$

Hence by the one-to-one nature of Fourier transforms we must have

$$\Omega(\xi_1,\eta_1)\Omega(\xi_2,\eta_2)\cos\left(\frac{\hbar}{2}(\xi_1\eta_2-\xi_2\eta_1)\right)$$
$$=\Omega(\xi_1+\xi_2,\eta_1+\eta_2)\,. \tag{3.7}$$

This gives

$$\Omega(\xi_1,0)\Omega(\xi_2,0)=\Omega(\xi_1+\xi_2,0)\,,$$
$$\Omega(0,\eta_1)\Omega(0,\eta_2)=\Omega(0,\eta_1+\eta_2)\,,$$
$$\Omega(\xi,0)\Omega(0,\eta)\cos\left(\frac{\hbar}{2}\xi\eta\right)=\Omega(\xi,\eta)\,.$$

Thus

$$\Omega(\xi_1,\eta_1)\Omega(\xi_2,\eta_2)=\Omega(\xi_1,0)\Omega(0,\eta_1)\cos\left(\frac{\hbar}{2}\xi_1\eta_1\right)\Omega(\xi_2,0)$$
$$\times\Omega(0,\eta_2)\cos\left(\frac{\hbar}{2}\xi_2\eta_2\right)$$
$$=\Omega(\xi_1+\xi_2,0)\Omega(0,\eta_1+\eta_2)\cos\left(\frac{\hbar}{2}\xi_1\eta_1\right)\cos\left(\frac{\hbar}{2}\xi_2\eta_2\right)$$
$$=\frac{\Omega(\xi_1+\xi_2,\eta_1+\eta_2)}{\cos\left(\frac{\hbar}{2}(\xi_1+\xi_2)(\eta_1+\eta_2)\right)}\cos\left(\frac{\hbar}{2}\xi_1\eta_1\right)\cos\left(\frac{\hbar}{2}\xi_2\eta_2\right)$$

and therefore

$$\cos\left(\frac{\hbar}{2}(\xi_1\eta_2-\xi_2\eta_1)\right)\cos\left(\frac{\hbar}{2}(\xi_1+\xi_2)(\eta_1+\eta_2)\right)$$
$$=\cos\left(\frac{\hbar}{2}\xi_1\eta_1\right)\cos\left(\frac{\hbar}{2}\xi_2\eta_2\right),$$

which is not an identity and hence we have a contradiction. Q.E.D.

Theorem 2, proved above, is a bona-fide generalization of Cohen's theorem, since functions $\Omega(\xi,\eta)$, which satisfy (2.7) and (2.8) but not (3.1), give rise to Ω-representations which necessarily do not satisfy (i) [as pointed out in the paragraph following (3.1)].

A bonus of the above proof is that $\Omega(\xi,\eta)$ need not satisfy (2.7), and (2.8), so the argument holds even for Ω-representations which do not necessarily map real-valued functions to self-adjoint operators and vice-versa, or the identity operators via (2.12) to the unit function.

In view of Theorem 2, the possibility of a classical stochastic formulation of quantum mechanics is left open only to nonlinear and non-Ω-representations.

It would be interesting to see whether Theorem 2 can be further generalized, e.g., to all Δ-representations, and if it cannot be generalized then to see for what kind of representation it does not hold.

ACKNOWLEDGMENT

I thank Professor P. D. Finch for clarifying discussions and, in particular, for pointing out that a contradiction can be derived from (3.7) without further assumptions.

[1] J. E. Moyal, Proc. Camb. Philos. Soc. **45**, 99 (1949).

[2] L. Cohen, J. Math. Phys. **7**, 781 (1966).

[3] G. S. Agarwal and E. Wolf, (a) Phys. Rev. D **2**, 2161 (1970); (b) **2**, 2182 (1970); (c) **2**, 2206 (1970).

[4] M. D. Srinivas and E. Wolf, Phys. Rev. D **11**, 1477 (1975).

[5] E. P. Wigner, in *Perspectives in Quantum Theory*, edited by W. Yourgrau and A. Vander Merwe (M. I. T., Cambridge, Mass., 1971), pp. 25–36.

[6] H. Margenau and L. Cohen, in *Quantum Theory and Reality*, edited by M. Bunge (Springer, Berlin, 1968).

[7] L. Cohen, Philos. Sci. **33**, 317 (1966).

[8] W. M. de Muynck, P. A. E. Janssen, and A. Santman, Found. Phys. **9**, 123 (1979).

[9] M. Mugur-Schächter, Found Phys. **9**, 389 (1979).

[10] V. V. Kuryshkin, in *The Uncertainty Principle and Foundations of Quantum Mechanics* (Wiley, London, 1977).

Path integration of an action related to an electron gas in a random potential

A. K. Dhara, D. C. Khandekar, and S. V. Lawande
Theoretical Reactor Physics Section, Bhabha Atomic Research Centre, Bombay 400085, India

(Received 21 August 1981; accepted for publication 12 February 1982)

Path integration of an action related to an electron gas in a random potential is performed within the framework of Feynman's polygonal path approach. The exact propagator obtained is simply related to the harmonic oscillator propagator. The integration is direct and does not require the knowledge of an auxiliary measure or the artificial coupling of the system to the external forces.

PACS numbers: 03.65.Db, 03.65.Ca

I. INTRODUCTION

The path integral formulation of Feynman[1] offers a global approach for solving quantum mechanical problems. In this formulation the usual time-dependent Schrödinger equation is replaced by an integral equation

$$\psi(x, T) = \int K(x, T; x_0, 0)\psi(x_0, 0)\, dx_0, \tag{1}$$

which expresses the wavefunction $\psi(x, T)$ at the time T in terms of the wavefunction $\psi(x_0, 0)$ at the initial time $t = 0$. The propagator or the kernel K is defined by a path integral

$$K(x, T; x_0, 0) = \int_{x(0) = x_0}^{x(T) = x} \exp\{(i/\hbar)S[x(t)]\}\, \mathscr{D}x(t), \tag{2}$$

where the symbol $\mathscr{D}x(t)$ implies that integrations are over all possible paths starting at $x(0) = x_0$ and terminating at $x(T) = x$. The functional $S[x(t)]$ in the integral is the classical action defined by

$$S[x(t)] = \int_0^T L(x, \dot{x}, t)\, dt, \tag{3}$$

$L(x, \dot{x}, t)$ being the Lagrangian of the system considered.

Although this approach is intuitively appealing, the problem of obtaining the propagator is, in general, nontrivial, because of the associated analytical difficulties. A simple prescription for computing the path integral K involves the assumption of polygonal paths. In this scheme, first proposed by Feynman and subsequently by Cameron,[2] the propagator is obtained as the limit of a multiple Riemann integral of order N when $N \to \infty$. A rigorous justification of the polygonal path formulation has been given by Truman in his recent papers.[3] Path integration without the limiting procedure has been discussed by Mizrahi.[4] Albeverio and Hoegh-Krohn[5] base the mathematical definition of Feynman path integrals on a general theory of oscillatory integrals on real Hilbert spaces.

In a recent review,[6] DeWitt-Morette, Maheshwari, and Nelson have discussed extensively a new method of path integration based on the theory of promeasures. Although this technique of global integration on function spaces is mathematically elegant and potentially powerful, the examples treated in Ref. 6 are precisely those which one can comfortably handle with the polygonal approach. A somewhat nontrivial instance, where the theory of promeasures was exploited, has been considered by Maheshwari.[7] This case involved the path integration of a system characterized by the action

$$S[x(t)] = \int_0^T dt \left\{ \frac{m}{2}\dot{x}^2 - \frac{m\Omega^2}{4T}\int_0^T [x(t) - x(t')]^2\, dt' \right\}. \tag{4}$$

This action has been considered by Bezak[8] in connection with a path integral theory of an electron gas in a random potential. Bezak, however, had to use imaginary time $-i\beta$ (β is the inverse of temperature) for obtaining the partition function and arrived at only an approximate evaluation of the path integral. Subsequently, Papadopoulos[9] evaluated the path integral (in imaginary time) in an exact closed form by coupling the system to auxiliary external forces.

The main purpose of the present paper is to show that the path integration of the action (4) can be carried out within the framework of Feynman's polygonal approach and in the spirit of some of our previous work.[10] Section II outlines this derivation, which shows how the propagator depends on the solution of the classical harmonic oscillator of frequency Ω acted on by a constant force. We present in Sec. III an alternative, shorter derivation of the same result, which brings out explicitly the relation between the present propagator and the well-known harmonic oscillator propagator.

II. DERIVATION OF THE PROPAGATOR

The polygonal paths formulation for the propagator $K(x, T; x_0, 0)$ is based on a partition P_N if the time interval $[0, T]$ into N subintervals. Assuming for simplicity that all subintervals are of equal length, we characterize P_N by

$$P_N: t_0 = 0, t_1, t_2, \ldots, t_{N-1}, t_N = T, \tag{5a}$$

$$t_j - t_{j-1} = \epsilon, \quad N\epsilon = T, \quad j = 1, 2, \ldots, N,$$

and the corresponding discretization of a path $x(t)$ by

$$x_j = x(t_j), \quad x_0 = x(0), \quad x_N = x(T) = x. \tag{5b}$$

The action $S[x(t)]$ of Eq. (4) is next expressed in the discrete form as

$$S_N[x_j] = \sum_{j=1}^{N} \epsilon\left[\frac{m}{2}\left(\frac{x_j - x_{j-1}}{\epsilon}\right)^2 - \frac{m\Omega^2\epsilon}{4T}\sum_{k=1}^{N}(x_j - x_k)^2\right]. \tag{6a}$$

Here, we have replaced the integrals in Eq. (4) by sums following the standard prescriptions of Feynman[1] for approximating the kinetic and potential energy terms over the jth subinterval $[t_{j-1}, t_j]$. Next the path differential measure takes the form

$$\mathscr{D}x(t) \mapsto \frac{1}{\Delta} \prod_{j=1}^{N-1} \frac{dx_j}{\Delta}, \tag{6b}$$

where $\Delta = (2\pi i\hbar\epsilon/m)^{1/2}$. One then writes the approximate propagator as a multiple Riemann integral

$$K_N(x, T; x_0, 0) = \frac{1}{\Delta} \int \cdots \int \exp\left(\frac{i}{\hbar} S_N[x_j]\right) \prod_{j=1}^{N-1} \frac{dx_j}{\Delta}. \tag{7}$$

Finally, in the limit of infinite refinement of the partition $P_N(\epsilon \to 0)$ of the time interval $[0, T]$, one expects that K_N goes over into the exact propagator $K(x, T; x_0, 0)$ defined by Eq. (2). We may, therefore, write

$$K(x, T; x_0, 0) = \lim_{\substack{N\to\infty \\ \epsilon \to 0 \\ (N\epsilon = T)}} K_N(x, T; x_0, 0). \tag{8}$$

Note that the introduction of Δ as defined above in the path differential measures is necessary to obtain the desired limit of the path integral (7) and the correct free-particle normalization[1] as $N \to \infty$. It may also be mentioned here that for one-dimensional problems (e.g., Refs. 1, 10) the polygonal path approach yields correct propagators. We shall see below that, even in the present case, this approach leads to an exact evaluation of the propagator in a closed form.

Next, in order to evaluate K_N, we have to substitute expression (6a) for S_N in Eq. (7) and carry out the integrations over x_j ($j = 1 \cdots N-1$) successively. This appears to be a formidable task at first. However, repeated use of the following identity for a one-dimensional Gaussian integral derived in the Appendix,

$$\int_{-\infty}^{\infty} \exp\left[\pm i \sum_{j=1}^{M} a_j(\xi_j - x)^2\right] dx = \left(\frac{\pi}{A}\right)^{1/2} e^{\pm i\pi/4} \exp\left[\pm \frac{i}{2A} \sum_{j,k=1}^{M} a_k a_j (\xi_j - \xi_k)^2\right], \tag{9}$$

where $A = \Sigma_{j=1}^{M} a_j$ and a_j are positive real numbers, simplifies the task considerably. Note that formula (9) is handy to use since at every stage j the integral to be evaluated is of the form given in lhs of Eq. (9) when one identifies the variable x with x_j and ξ_k as the x_k ($N-1 > k > j$) besides x_0 and x_N. We then obtain

$$K_N(x, T; x_0, 0) = \left(\frac{m}{2\pi i\hbar}\right)^{1/2} \frac{1}{q_N^{1/2}} \exp\left[\frac{im}{2\hbar} p_N (x - x_0)^2\right], \tag{10}$$

where

$$p_N = B_N/\epsilon, \quad q_N = \epsilon \prod_{k=1}^{N-1} A_k. \tag{11}$$

The coefficients A_K and B_K are determined from the following system of recursion relations:

$$A_k = 1 + B_k + (N - k)C_k, \tag{12}$$

$$B_1 = 1, \quad B_k = \frac{B_{k-1}}{A_{k-1}} + \sum_{j=1}^{k-1} \frac{B_j C_j}{A_j}, \quad k \geqslant 2, \tag{13}$$

$$C_1 = -\frac{\Omega^2 \epsilon^3}{T}, \quad C_k = C_1 + \frac{C_{k-1}}{A_{k-1}} + \sum_{j=1}^{k-1} \frac{C_j^2}{A_j}. \tag{14}$$

The system of Eqs. (12)–(14) represent a set of nonlinear coupled difference equations, and hence is difficult to treat analytically. However, according to Eq. (8), in order to obtain the propagator $K(x, T; x_0, 0)$ we need merely to evaluate p_N and q_N as $N \to \infty$ ($\epsilon \to 0$). For this purpose, it is more appropriate to derive a set of differential equations equivalent to the set (12)–(14) by taking the limit $\epsilon \to 0$. To this end, it is convenient to introduce the quantities λ_k, P_k, and Q_k by writing

$$\begin{aligned} A_k &= \lambda_{k+1}/\lambda_k, \\ B_k &= \epsilon P_k/\lambda_k, \\ C_k &= \epsilon^2 Q_k/\lambda_k. \end{aligned} \tag{15}$$

The recursion relations (12)–(14) now take the form

$$\lambda_{k+1} = \lambda_k + \epsilon P_k + (N - k)\epsilon^2 Q_k, \tag{16}$$

$$P_k = P_{k-1} + \epsilon^2 \lambda_k \sum_{j=1}^{k-1} \frac{P_j Q_j}{\lambda_j \lambda_{j+1}}, \tag{17}$$

$$Q_k = Q_{k-1} - \frac{\Omega^2 \epsilon}{T} \lambda_k + \epsilon^2 \lambda_k \sum_{j=1}^{k-1} \frac{Q_j^2}{\lambda_j \lambda_{j+1}}. \tag{18}$$

It is easy to see that Eqs. (16)–(18) reduce to the following system of equations in the limit $\epsilon \to 0$, $N \to \infty$ ($N\epsilon = T$):

$$\dot\lambda = P + (T - t)Q, \tag{19}$$

$$\dot P = \lambda \int_0^t \frac{P(\tau)Q(\tau)}{\lambda^2(\tau)} d\tau, \tag{20}$$

$$\dot Q = -\frac{\Omega^2}{T}\lambda + \lambda \int_0^t \frac{Q^2(\tau)}{\lambda^2(\tau)} d\tau, \tag{21}$$

with the initial conditions

$$\begin{aligned} &\lambda(0) = 0, \quad Q(0) = \dot Q(0) = 0, \\ &P(0) = \dot\lambda(0), \quad \dot P(0) = 0. \end{aligned} \tag{22}$$

After some algebra, it is possible to cast Eqs. (19)–(21) in the form

$$\ddot\lambda + \Omega^2 \lambda = -2Q, \tag{23}$$

$$\dot P = Q, \tag{24}$$

$$\ddot Q + \Omega^2 Q = -(\Omega^2/T)\dot\lambda(0). \tag{25}$$

Note that Eq. (25) indeed represents the equation of motion of a classical oscillator of frequency Ω acted on by a constant force.

Equations (23)–(25) together with the initial conditions (22) may be readily solved to obtain

$$\lambda(t) = [\dot\lambda(0)/\Omega T][(T - t)\sin \Omega t + (2/\Omega)(1 - \cos \Omega t)], \tag{26}$$

$$P(t) = [\dot\lambda(0)/T][(T - t) + (1/\Omega)\sin \Omega t], \tag{27}$$

$$Q(t) = -[\dot\lambda(0)/T](1 - \cos \Omega t). \tag{28}$$

It is now straightforward to obtain the limiting values of p_N and q_N as $N \to \infty$ $(\epsilon \to 0)$. We first note that

$$q_N = \epsilon \prod_{k=1}^{N-1} A_k = \frac{\epsilon \lambda_N}{\lambda_1} = \frac{\epsilon \lambda (T)}{\lambda (\epsilon)}.$$

Hence

$$\lim_{N \to \infty} q_N = \lambda (T) \lim_{\epsilon \to 0} \frac{\epsilon}{\lambda (\epsilon)} = \frac{\lambda (T)}{\dot{\lambda} (0)}$$
$$= \left(\frac{\sin \Omega T}{\Omega}\right)\left(\frac{\tan \Omega T/2}{\Omega T/2}\right). \tag{29}$$

Next, it is easy to see that

$$\lim_{N \to \infty} p_N = \lim_{N \to \infty} \frac{p_N}{\lambda_N} = \frac{p(T)}{\lambda (T)}$$
$$= \frac{\Omega}{2} \cot \frac{\Omega T}{2}. \tag{30}$$

Substituting these limiting values of q_N and p_N in Eq. (10) and using Eq. (8), we arrive at the exact expression for the propagator

$$K(x, T; x_0, 0) = \left(\frac{\Omega T}{\sin \Omega T}\right)^{1/2}\left(\frac{m\Omega}{4\pi i\hbar} \cot \frac{\Omega T}{2}\right)^{1/2}$$
$$\times \exp\left[\frac{im\Omega}{4\hbar} \cot \frac{\Omega T}{2} (x - x_0)^2\right]. \tag{31}$$

We recover from Eq. (31) the propagator obtained by Papadopoulos[9] and Maheshwari[7] when T is replaced by $-i\beta$. It is interesting to note that the propagator of Eq. (31) looks like that of a free particle with an "effective mass" $m^* = \frac{1}{2} m\Omega T \cot(\frac{1}{2}\Omega T)$ while the normalization factor contains an additional term $(\Omega T/\sin \Omega T)^{1/2}$ apart from the free particle normalization $(m^*/2\pi i\hbar T)^{1/2}$. In fact, the action (4) admits the classical equation of motion

$$m\ddot{x} + m\Omega^2 x = (m\Omega^2/T)\int_0^T x(t')\, dt', \tag{32}$$

which can be solved with the conditions $x(0) = x_0, x(T) = x$ to yield the classical path

$$x(t) = \frac{x + x_0}{2} + \frac{x - x_0}{2} \frac{\sin[\frac{1}{2}\Omega(2t - T)]}{\sin(\frac{1}{2}\Omega T)}. \tag{33}$$

It is now easy to evaluate the contribution of the classical path to the action S_{cl}, which turns out to be

$$S_{\text{cl}} = [\tfrac{1}{4} m\Omega \cot(\tfrac{1}{2}\Omega T)](x - x_0)^2. \tag{34}$$

The propagator of Eq. (31) is then essentially given by the Van Vleck–Pauli formula

$$C_f\left(\frac{1}{2\pi i\hbar}\left|\frac{\partial^2 S_{\text{cl}}}{\partial x \partial x_0}\right|\right)^{1/2} \exp\left(\frac{i}{\hbar} S_{\text{cl}}\right), \tag{35}$$

apart from the correction factor $C_f = (\Omega T/\sin \Omega T)^{1/2}$, which represents the sum of contributions arising out of the deviations from the classical path.

III. RELATION TO THE HARMONIC OSCILLATOR PROPAGATOR

In this section, we consider an alternative derivation of the propagator of Eq. (31) which avoids the integrations over successive $x_j (j = 1,2,...,N-1)$. First note that by letting $M = N, a_j = m\Omega^2\epsilon/2\hbar$, and $\xi_j = x_j$ $(j = 1 \cdots N)$ in the identity (9) and taking the negative sign, we can write

$$\exp\left[-\frac{im\Omega^2\epsilon^2}{4\hbar T} \sum_{j,k=1}^{N} (x_j - x_k)^2\right] = \Omega\left(\frac{mT}{2\pi\hbar}\right)^{1/2} e^{i\pi/4}$$
$$\times \int_{-\infty}^{\infty} dy \exp\left[-\frac{im\Omega^2\epsilon}{2\hbar} \sum_{j=1}^{N} (y - x_j)^2\right]. \tag{36}$$

The propagator defined through Eqs. (7) and (8) then takes the form

$$K(x,T;x_0,0) = \lim_{\substack{N \to \infty \\ \epsilon \to 0 \\ (N\epsilon = T)}} \Omega\left(\frac{mTi}{2\pi\hbar}\right)^{1/2} \int_{-\infty}^{\infty} dy \int \cdots \int \prod_{j=1}^{N-1} \frac{dx_j}{\Delta}$$
$$\times \exp\left\{\frac{im}{2\hbar\epsilon}\left[(x_j - x_{j-1})^2 - \Omega^2\epsilon^2 \sum_{j=1}^{N} (y - x_j)^2\right]\right\}. \tag{37}$$

Next a change of variables from x_j to $X_j = x_j - y$ reduces Eq. (37) to

$$K(x,T;x_0,0)$$
$$= \Omega\left(\frac{mTi}{2\pi\hbar}\right)^{1/2} \lim_{N \to \infty} \left[\int_{-\infty}^{\infty} dy\, K_N^{\text{HO}}(x + y, T; x_0 + y, 0)\right], \tag{38}$$

where K_N^{HO} is the propagator for harmonic oscillator in the Nth approximation. Assuming for simplicity[11] that the limit $N \to \infty$ can be carried out under the integral sign, we obtain

$$K(x,T;x_0,0) = \Omega\left(\frac{mTi}{2\pi\hbar}\right)^{1/2} \int_{-\infty}^{\infty} dy\, K^{\text{HO}}(x + y, T; x_0 + y, 0). \tag{39}$$

Finally using the well-known expression for K^{HO} given by

$$K^{\text{HO}}(x, T; x_0, 0) = \left(\frac{m\Omega}{2\pi i\hbar \sin \Omega T}\right)^{1/2}$$
$$\times \exp\left\{\frac{im\Omega}{2\hbar \sin \Omega T} [(x^2 + x_0^2)\cos \Omega T - 2xx_0]\right\}, \tag{40}$$

in Eq. (39) and carrying out the integration over y, we again arrive at Eq. (31) for the propagator. Equation (39) yields an interesting explicit relation between the propagator for the present problem and that for a harmonic oscillator.

IV. CONCLUSIONS

The main contribution of this paper is to show that the action considered by Bezak,[8] Papadopoulos,[9] and Maheshwari[7] may be path-integrated within the polygonal path approach of Feynman without much ado. The present derivation is self-contained and does not require the knowledge of an auxiliary measure[7] or an artificial coupling to the external forces.[9] An explicit relation between the propagator for the present problem and the harmonic oscillator propagator has been obtained, leading to a "back of an envelope" calculation of the path integral considered.

APPENDIX

In this appendix, we give a derivation of the identity (9) of Sec. II. Choosing first the positive sign in Eq. (9), we have

$$\begin{aligned}\int_{-\infty}^{\infty} \exp\left[i\sum_{j=1}^{M} a_j(\xi_j - x)^2\right] dx &= \exp\left[i\left(\sum_{j=1}^{M} a_j\xi_j^2 - \frac{1}{A}\sum_{j,k=1}^{M} a_j a_k \xi_j \xi_k\right)\right]\\ &\quad\times\int_{-\infty}^{\infty}\exp\left[iA\left(x-\frac{1}{A}\sum_{j=1}^{M}a_j\xi_j\right)^2\right]dx\\ &= \exp\left[\frac{i}{2A}\sum_{j,k=1}^{M}a_j a_k(\xi_j-\xi_k)^2\right]\\ &\quad\times\int_{-\infty}^{\infty}dy\,\exp(iAy^2), \qquad \text{(A1)}\end{aligned}$$

where $A = \Sigma_{j=1}^{M} a_j > 0$.

It now remains to show that the improper Gaussian integral in Eq. (A1) may be assigned a value $(\pi/A)^{1/2} \times \exp(i\pi/4)$. Consider the integral

$$\oint_C \exp(iAz^2)\,dz,$$

where C is a contour shown in Fig. 1. Next

$$\begin{aligned}\oint_{-\infty}^{\infty} \exp(iAz^2)\,dz &= \int_{-R}^{R}\exp(iAx^2)\,dx + \int_{C_1}\exp(iAz^2)\,dz\\ &\quad + e^{i\pi/4}\int_{-R}^{R}\exp(-Ar^2)\,dr + \int_{C_2}\exp(iAz^2)\,dz = 0. \qquad \text{(A2)}\end{aligned}$$

The contributions from the arcs C_1 and C_2 vanish as $R\rightarrow\infty$. Hence

$$\begin{aligned}\int_{-\infty}^{\infty}\exp(iAx^2)\,dx &= e^{i\pi/4}\int_{-\infty}^{\infty}\exp(-Ar^2)\,dr = (\pi/A)^{1/2}e^{i\pi/4}. \qquad \text{(A3)}\end{aligned}$$

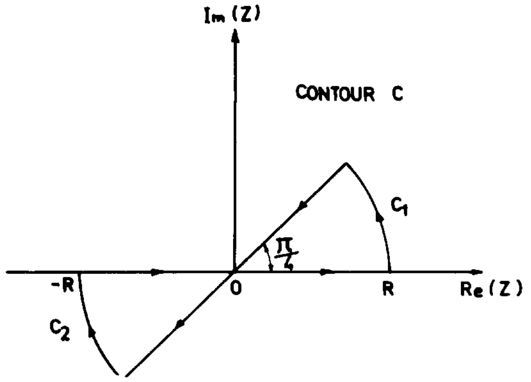


FIG. 1.

The result involving the negative sign in Eq. (9) may be proved similarly by choosing an appropriate contour.

[1] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[2] R. H. Cameron, J. Math. Phys. **39**, 126 (1960); J. Anal. Math. **10**, 287 (1962).

[3] A. Truman, J. Math. Phys. **17**, 1852 (1976); **18**, 1499, 2308 (1977); **19**, 1742 (1978).

[4] M. M. Mizrahi, J. Math. Phys. **17**, 566 (1976).

[5] S. Albverio and R. Hoegh-Krohn, *Mathematical Theory of Feynman Path Integrals*, Lecture Notes in Mathematics 532 (Springer-Verlag, Berlin, 1976).

[6] C. DeWitt-Morette, A. Maheshwari, and B. Nelson, Phys. Rep. **50**, 255 (1979).

[7] A. Maheshwari, J. Phys. A **8**, 1019 (1975).

[8] V. Bezak, Proc. R. Soc. London Ser. A **315**, 339 (1970).

[9] G. Papadopoulos, J. Phys. A **7**, 183 (1974).

[10] D. C. Khandekar and S. V. Lawande, J. Math. Phys. **16**, 384 (1975); **20**, 1870 (1979).

[11] This assumption is not necessary. In fact, one can rigorously carry out the evaluation of $K_N^{\rm HO}$ in Eq. (38) and, subsequently, the integration over y. Taking further the limit $N\rightarrow\infty$, we arrive at Eq. (31).

A partial inner product space of analytic functions for resonances[a)]

L. P. Horwitz and E. Katznelson
Department of Physics and Astronomy, Tel Aviv University, Ramat Aviv, Israel

(Received 4 September 1980; accepted for publication 19 February 1982)

Generalized Hilbert spaces $D(\alpha,\beta)$ are defined using analytic continuation of Hardy class functions into a wedge bounded by the angles α,β. Eigenfunctions of *isolated* complex eigenvalues may be found in $D(\alpha,\beta)$ for operators that have a self-adjoint representation in L^2. These eigenvalues correspond to resonances in the associated decay problem. A bilinear form between $D(\alpha,\beta)$ and $D(-\beta,-\alpha)$ is defined, which has some of the properties of a Hilbert space scalar product, and it is shown that this form can be used to define a variational principle to obtain the eigenvalue equations.

PACS numbers: 03.65.Db, 03.65.Nk

INTRODUCTION

As pointed out in the excellent discussion of quasistationary state theory by Baz', Zel'dovich, and Perelomov,[1] Gamow's theory of the alpha decay of heavy nuclei (1928) was the first successful application of quantum mechanics to the atomic nucleus. Although effective methods have been in use for the approximate calculation of various properties of unstable states for many years, the mathematical foundation for these procedures has not been very clear. Unstable states belong to the continuous spectrum; the separation between the part of the continuum to be associated with such states and the part which should be associated with the continuum was not well defined. Wavefunctions with complex energy eigenvalues which may be defined from the Schrödinger equation, such as the Gamow wavefunctions, are not elements of the Hilbert space, and the usual notions of completeness and orthogonality do not apply. Nevertheless, the notion of associating an unstable state with a well-defined functional of some type seems essential in order to characterize the properties of the state.

Berggren,[2] using Zel'dovich's method of regularization, defined a generalized inner product which enabled him to discuss the problems of completeness and orthogonality of resonant states. A later paper by Romo,[3] following independently some of Berggren's procedures, emphasized the technique of carrying out analytic continuation of matrix elements. In a similar way, Fuller[4] worked with analytic continuations of the Lippmann–Schwinger equation.

As first pointed out by Grossmann,[5] it is possible to study the resonance problem and define a resonant state systematically by weak analytic continuation (a procedure actually used by Romo[3]), which he realizes by constructing mappings into subspaces of analytic functions. Although he utilized nested Hilbert spaces, he remarked that one could use sequences such as Gel'fand triples as well. Aguilar, Balslev, and Combes[6] have carried out a program of this type by using dilatation analytic potentials. Howland and Baumgärtel[7] achieved general results for the perturbation theory of eigenvalues imbedded into the continuous spectrum. Simon[8] obtained these results using the procedures of ABC. Horwitz and Sigal[9] used these approaches to study perturbation theory and the resonant state as an element of a Gel'fand triple, both through weak analytic continuation and through the use of a dilatation analytic subset of the Hilbert space. Sudarshan, Chiu, Gorini, and Parravicini and Bailey and Schieve[10] studied the analytic deformation of the real continuous spectrum, and Bohm[11] has investigated an explicit pole formalism. Katznelson[12] has discussed the time dependence of the decay law associated with arbitrary degeneracy of the complex poles of the resolvent.

In techniques which use analytic continuation of the continuous spectrum to the lower half-plane to construct the functional which is an eigenfunction of the Hamiltonian with complex eigenvalue coinciding with the position of the resonance pole, this eigenvalue lies in a continuous sea of eigenvalues[9]; one cannot, therefore, construct a variational principle. One sees, however, that it is appropriate to construct a theory which works with complex canonical variables.[13] We have adopted a point of view in which we use complex variables from the beginning in the framework of Hilbert spaces of analytic functions (Hardy class) as developed by Van Winter,[14,15] in close relation to the work of ABC. In a given angular wedge of the complex plane for the canonical variables, one finds a quantum theory parallel to the usual real canonical theory,[13] that is, every expectation value remains the same under the transformation leading to the new representations. One can find, however, complex eigenvalues for the Hamiltonian in the new representation, with eigenvectors which belong to the space defined by the complex wedge. These eigenvectors do not have an L^2 norm, since the existence of such a norm would preclude a complex eigenvalue. Nevertheless, one can define a scalar product (with the same L^2 measure) between elements of the space defined by the complex wedge and elements of a dual space, of which some elements are also L^2. Structures of this type, involving two spaces with the property that the scalar product is not defined between any pair of elements, but only between elements belonging to distinct spaces, and a third space, which is the intersection between them, compatible with both, were first proposed by Antoine and Grossmann,[16] and given the name of partial inner product spaces. It appears, therefore, that partial inner product spaces provide a suitable structure for the description of the states of unstable

---

[a)] Supported in part by the Binational Science Foundation (BSF) Jerusalem.

0022-2488/83/040848-12$02.50

systems. In this paper, we shall utilize the results of Van Winter to establish a connection between the complex singularities of the resolvent associated with resonance phenomena, and partial inner product spaces.

In Sec. I, we define $D(\alpha,\beta)$, a Hilbert space of analytic functions in a wedge.[15] In Secs. II and III some properties of these spaces are discussed. The notion of a space and its dual space is developed, and operators constructed in these two spaces are defined. A class of Hamiltonians, their associated resolvents, and their isolated singularities, are discussed in Secs. IV and V. In Sec VI, a notion of adjoint, appropriate to partial inner product spaces (based on a version of the Reisz theorem applicable to such spaces) is introduced, and in Sec. VII a variational principle for the isolated singularities of the resolvent is studied.

I. SPACES $D(\alpha,\beta)$ AND RESONANCES

We wish to study the resonance problem in the context of a Schrödinger equation with self-adjoint Hamiltonian of the form

$$H = p^2 + V(r), \tag{1.1}$$

where $V(r)$ is a real multiplication operator defined on $0 \leqslant r < \infty$. We shall work in one dimension to establish a set of basic results. These will apply directly to the spherically symmetrical case in higher dimensionality which we shall use later. We shall utilize a Hilbert space of analytic functions of the type discussed by Van Winter,[15] corresponding to a conformal mapping of a complete set of Hardy class functions. Functions of this type are analytic, regular in a sector $\alpha < \varphi < \beta$ and have the property that

$$\int_0^\infty |f(ke^{i\varphi})|^2 \, dk \tag{1.2}$$

exists and is bounded uniformly in φ for $\alpha < \varphi < \beta$. We shall denote such a set of functions by $D(\alpha,\beta)$. There are boundary functions[15] $f(ke^{i\alpha})$ and $f(ke^{i\beta})$ [also satisfying (1.2)] such that $(\psi = \alpha, \beta)$

$$\lim_{\varphi \to \psi} \int_0^\infty |f(ke^{i\varphi}) - f(ke^{i\psi})|^2 \, dk = 0. \tag{1.3}$$

With the inner product

$$(f,g)_{D(\alpha,\beta)} = \frac{1}{2} \left[\int_0^\infty \bar{f}(ke^{i\alpha}) g(ke^{i\alpha}) \, dk + \int_0^\infty \bar{f}(ke^{i\beta}) g(ke^{i\beta}) \, dk \right], \tag{1.4}$$

this set of functions, closed in the norm $\|f\|_{D(\alpha,\beta)}$, becomes a Hilbert space. The variable k can denote momentum or position; in fact, as we shall discuss later, $D(\alpha,\beta) \mapsto D(-\beta,-\alpha)$ under Fourier transform.

If we take $V(r)$ to be the restriction to $\varphi = 0$ of a function $v(re^{-i\varphi})$ in $D(-\gamma,\gamma)$ for some $\gamma > 0$, then there may exist solutions of the equation (for E real).

$$Hf = Ef, \tag{1.5}$$

which are elements of $D(\alpha,\beta)$ for $-\gamma < \alpha, \beta < \gamma$.

Furthermore, if we consider spaces $D(\alpha,\beta)$ for which α, β have the same sign, we may find solutions to Eq. (1.5) for E complex-valued with f an element of such a space. Such solutions are associated with complex poles of the resolvent, and, as will be demonstrated at the end of Sec. VI, lead to the exponential decay law in time, which is characteristic of resonant or unstable states of the system described by the Hamiltonian H. We shall therefore interpret these solutions as the quantum mechanical representations of the corresponding unstable states. These solutions cannot be continued back to $\varphi = 0$ and still remain integrable in the sense of Eq. (1.2). For every solution, in the space $D(\alpha,\beta)$, corresponding to a complex eigenvalue E there is a solution in $D(-\beta,-\alpha)$ corresponding to complex eigenvalue $\bar{E}$ (these eigenvalues lie between the wedge and the real axis).

II. MELLIN TRANSFORM AND PROPERTIES OF THE BOUNDED KERNEL K

The φ dependence of the relations which we wish to study is simplified by the introduction of the Mellin transform. We define the Mellin transform by

$$\mathbf{f}(u,\varphi) = [1/(2\pi)^{1/2}] \int_0^\infty f(ke^{i\varphi}) k^{iu-1/2} \, dk. \tag{2.1}$$

The weight factor is determined by the conformal mapping between functions in $D(\alpha,\beta)$ and functions of w, where $e^w = ke^{i\phi}$, analytic in a strip.[14] A complex-valued function f defined on the sector $\alpha < \varphi < \beta$ belongs to $D(\alpha,\beta)$ if and only if[14]

$$\mathbf{f}(u,\varphi) = -e^{\varphi u - i\varphi/2} \mathbf{f}(u) \tag{2.2}$$

for some function $\mathbf{f}(u)$ satisfying

$$\int_{-\infty}^\infty (e^{2\alpha u} + e^{2\beta u}) |\mathbf{f}(u)|^2 \, du < \infty. \tag{2.3}$$

The inverse Mellin transform is given by

$$f(ke^{i\varphi}) = [1/(2\pi)^{1/2}] \int_{-\infty}^\infty \mathbf{f}(u) (ke^{i\varphi}) e^{-iu-1/2} \, du. \tag{2.4}$$

Let us define a class $\mathscr{K}$ of integral kernels $K(ke^{i\varphi}, k'e^{-i\varphi})$ (we choose to represent the phase of the second argument as $e^{-i\varphi}$ for later convenience), where

$$\int_0^\infty dk \int_0^\infty dk' |K(ke^{i\varphi}, k'e^{-i\varphi})|^2 \tag{2.5}$$

exists and is bounded uniformly in φ, for $\alpha < \varphi < \beta$; and, furthermore

$$(Kf)(ke^{i\varphi}) = \int_0^\infty K(ke^{i\varphi}, k'e^{-i\varphi}) f(k'e^{i\varphi}) e^{i\varphi} \, dk' \tag{2.6}$$

is in $D(\alpha,\beta)$ if f is in $D(\alpha,\beta)$. The Mellin transform of the kernel K is defined by[14]

$$\mathbf{K}(u,u',\varphi) = (1/2\pi) \int_0^\infty dk \int_0^\infty dk' \, k^{iu-1/2} K(ke^{i\varphi}, k'e^{-i\varphi}) \times \overline{(k')^{iu'-1/2}} = e^{\varphi u} e^{-\varphi u' - i\varphi} \mathbf{K}(u,u'), \tag{2.7}$$

with $\mathbf{K}(u,u')$ satisfying

$$\int_{-\infty}^\infty du \int_{-\infty}^\infty du' (e^{2\alpha(u-u')} + e^{2\beta(u-u')}) |\mathbf{K}(u,u')|^2 < \infty. \tag{2.8}$$

The transform inverse to (2.7) is

$$K(ke^{i\varphi},k'e^{-i\varphi})$$
$$= (1/2\pi)\int_{-\infty}^{\infty} du \int_{-\infty}^{\infty} du'\, \mathbf{K}(u,u')(ke^{i\varphi})^{-iu-1/2}$$
$$\times\, \overline{(k'e^{-i\varphi})^{-iu'-1/2}}. \tag{2.9}$$

The Mellin transform of Eq. (2.6) is

$$(\mathbf{K}f)(u,\varphi)$$
$$= [1/(2\pi)^{1/2}]\int_0^{\infty} (Kf)(ke^{i\varphi})k^{iu-1/2}\,dk$$
$$= (1/2\pi)\int_0^{\infty} dk \int_0^{\infty} dk'e^{i\varphi}k^{iu-1/2}K(ke^{i\varphi},k'e^{-i\varphi})$$
$$\times \int_{-\infty}^{\infty} k'^{-iu'-1/2}e^{\varphi u'-i\varphi/2}\mathbf{f}(u')\,du'$$
$$= e^{\varphi u - i\varphi/2}\int_0^{\infty} du'\,\mathbf{K}(u,u')\mathbf{f}(u'). \tag{2.10}$$

Now consider the integral

$$\int_0^{\infty} \overline{g(ke^{-i\varphi})}K(ke^{i\varphi},k'e^{-i\varphi})e^{i\varphi}\,dk', \tag{2.11}$$

where $g(ke^{-i\varphi})$, for $\alpha<\varphi<\beta$, lies in $D(-\beta,-\alpha)$. We shall show that this integral is a function of $k'e^{-i\varphi}$ and that it lies in $D(-\beta,-\alpha)$. To see this, we substitute the definition (2.4) into (2.11), and take the Mellin transform. Equation (2.11) then becomes

$$(1/2\pi)\int_{-\infty}^{\infty} du \int_0^{\infty} dk \int_0^{\infty} dk'\bar{\mathbf{g}}(u)(ke^{i\varphi})^{iu-1/2}$$
$$\times K(ke^{i\varphi},k'e^{-i\varphi})e^{i\varphi}\,k'^{-iu'-1/2}$$
$$= e^{-\varphi u'-i\varphi/2}\int_{-\infty}^{\infty} \bar{g}(u)\mathbf{K}(u,u')\,du. \tag{2.12}$$

Comparing with Eq. (2.2), one sees that (2.12) corresponds to the complex conjugate of a function $\mathbf{f}(u',-\varphi)$, *for* $\alpha<\varphi<\beta$; in the integral we can multiply by $e^{-\varphi u}e^{\varphi u}$ and apply the Schwarz inequality which, along with (2.7) and (2.8), shows that (2.12) lies in $D(-\beta,-\alpha)$. We shall say that a kernel of this type is constructed in the two spaces $D(\alpha,\beta)$ and $D(-\beta,-\alpha)$, in contrast with the usual type of kernel in a Hilbert space, which is constructed in just one space.

III. RELATIONS BETWEEN THE SPACES $D(\alpha,\beta)$ AND $D(-\beta,-\alpha)$

We shall use the notation $D(0,0)$ for the usual L^2 space defined by square integrable functions on the real half-line. For any $g(k)\in L^2$ and any fixed ψ in $\alpha\leqslant\psi\leqslant\beta$, and a positive ϵ, there is a function $f(ke^{i\varphi})$ in $D(\alpha,\beta)$ such that

$$\int_0^{\infty} |f(ke^{i\psi})-g(k)|^2\,dk<\epsilon, \tag{3.1}$$

i.e., the functions in $D(\alpha,\beta)$ and their boundary functions are dense in $D(0,0)$.[15] In a similar way we shall show that the set of functions that can be continued analytically to a wedge containing the real half-line, say $D(-\gamma,\gamma)$, is dense in a wedge contained in $(-\gamma,\gamma)$ in the following sense:

For any fixed ψ in $\alpha\leqslant\psi\leqslant\beta$ and a positive ϵ, for any $g(ke^{i\varphi})\in D(\alpha,\beta)$, where $-\gamma\leqslant\alpha<\beta<\gamma$, there exists a function $f(ke^{i\theta})\in D(-\gamma,\gamma)$, for each θ, such that

$$\int_0^{\infty} |f(ke^{i\theta})-g(ke^{i\psi})|^2dk<\epsilon, \tag{3.2}$$

i.e., the set of elements associated with any given θ in $D(-\gamma,\gamma)$ is dense in $D(\alpha,\beta)$. The proof is as follows. The set of functions, for fixed $\theta, f(ke^{i\theta})\in D(-\gamma,\gamma)$ is dense in $D(0,0)$. Since $f(ke^{i\theta})$ is a continuation of a function $f(k)\in D(0,0)$, this statement implies that the functions in $D(0,0)$ that can be continued analytically to a wedge $(-\gamma,\gamma)$ are dense in $D(0,0)$ and remain dense after the continuation. Now consider $g(ke^{i\psi})$ as a function in L^2; since $f(ke^{i\varphi})$ is dense in L^2, the result (3.2) follows.

Let us consider the class of operators $\mathscr{U}$ defined on $D(\alpha,\beta)$, such that

$$\mathscr{A}(A,\varphi)=\sup_{f\in D(\alpha,\beta)}\left[\int_0^{\infty} |(Af)(ke^{i\varphi})|^2dk\right]^{1/2}$$
$$\times\left[\int_0^{\infty} |f(ke^{i\varphi})|^2dk\right]^{-1/2} \tag{3.3}$$

exists and is uniformly bounded in $\alpha<\varphi<\beta$. If A satisfies Eq. (2.5), then, by the Schwarz inequality, it satisfies (3.3) also. Now, considering $f(ke^{i\varphi})$ as a function in L^2, then $Af(ke^{i\varphi})$, for all f in $D(\alpha,\beta)$, defines a family of operators, $A(\varphi)$ from L^2 to L^2, i.e.,

$$(A(\varphi)f)(ke^{i\varphi})=(Af)(ke^{i\varphi}), \tag{3.4}$$

and Eq. (3.3) is the L^2 norm $\|A(\varphi)\|$ of $A(\varphi)$. We remark that[14]

$$\sup_{\alpha<\varphi<\beta}\|A(\varphi)\|=\max\{\|A(\alpha)\|,\|A(\beta)\|\}, \tag{3.5}$$

which provides a norm for operators A in $\mathscr{U}$.

IV. THE HAMILTONIAN AND ITS ASSOCIATED RESOLVENT

Let $\mathscr{D}(H_0)$ be the set of functions $f(ke^{i\varphi})$ in $D(\alpha,\beta)$ with the property that $k^2e^{2i\varphi}f(ke^{i\varphi})$ is in $D(\alpha,\beta)$. Let the free Hamiltonian H_0 have domain $\mathscr{D}(H_0)$ and let it act according to

$$H_0f=k^2e^{2i\varphi}f \tag{4.1}$$

for all f in $D(H_0)$. This definition results in a closed operator [in the norm defined by Eq. (1.4)]. The resolvent of the free Hamiltonian,

$$R_0(\lambda)=(H_0-\lambda)^{-1}, \tag{4.2}$$

is bounded as an operator on $D(\alpha,\beta)$ if λ is in the sector

$$2\beta<\arg\lambda<2\pi+2\alpha,$$

where $-\pi/2<\alpha\leqslant\varphi\leqslant\beta<\pi/2$. In particular, $R_0(\lambda,\varphi)$ [an operator defined on L^2 as in Eq. (3.4)] is bounded for all λ for which

$$2\varphi<\arg\lambda<2\pi+2\varphi. \tag{4.3}$$

Now suppose the potential V of Eq. (1.1) to be the restriction to $\varphi=0$ of a function $V(re^{-i\varphi})$ in $D(-\gamma,\gamma)$ for some positive γ. Its Fourier transform is defined by

$$(FV)(\vec{k},\varphi)=[1/(2\pi)^{3/2}]\int_0^{\infty}\int e^{i\vec{k}\cdot\vec{r}}V(re^{-i\varphi})$$
$$\times e^{-i\varphi}(re^{-i\varphi})^2\,dr\,d\omega. \tag{4.4}$$

In the following we shall discuss the three-dimensional (one-particle) case for the sake of simplicity in working with the

Fourier transform; ω corresponds to angular variables, i.e., polar and azimuthal angles. It is shown by Van Winter[15] that $FV(\vec{k},\varphi)$ is a function of $ke^{i\varphi}$ only. In fact, if we consider the action of F on $\hat{f}(\vec{r}e^{-i\varphi}) \in D(-\beta,-\alpha)$, then $f(\vec{k},\varphi) = F\hat{f}(\vec{k},\varphi)$ is a function of $\vec{k}e^{i\varphi}$ and lies in $D(\alpha,\beta)$. We shall denote it as $f(\vec{k}e^{i\varphi})$. F is therefore an isometric mapping of $D(-\beta,-\alpha)$ into $D(\alpha,\beta)$ and a unitary mapping of $D(-\gamma,\gamma)$ into $D(-\gamma,\gamma)$. We define

$$W(\vec{k}e^{i\varphi}) = [1/(2\pi)^{3/2}]FV(\vec{k},\varphi). \tag{4.5}$$

We define the action of $V(re^{-i\varphi})$ on $\hat{f}(\vec{r}e^{-i\varphi}) \in D(-\beta,-\alpha)$, as $V(re^{-i\varphi})\hat{f}(\vec{r}e^{-i\varphi})$. In the Fourier transformed (momentum) space, FV acts an an integral kernel in the following way:

$$(FV\hat{f})(\vec{k}e^{i\varphi}) = \int_0^\infty \int W(|k-k'|e^{i\varphi}) f(\vec{k}'e^{i\varphi}) \times e^{i\varphi}(k'e^{i\varphi})^2\, dk'\, d\omega. \tag{4.6}$$

We remark that, although $V\hat{f}$ is in $D(-\beta,-\alpha)$, $\|V\hat{f}\|/\|\hat{f}\|$ [the usual $L^2(\mathbb{R}_+,d^2k)$ norm, where $\mathbb{R}_+$ is the positive half-line, and d^3k the Lebesque measure in momentum space) may have no supremum, and therefore the operator $V(\varphi)$ on $L^2(\mathbb{R}_+,d^3k)$ defined by (4.6) may be unbounded.

We shall define $V(\varphi)$ as follows. Let $g(k)$ run through a dense set in $L^2(\mathbb{R}_+,d^3k)$; then $V(\varphi)$ is the operator for fixed φ in $-\gamma \leqslant \varphi \leqslant \gamma$ on $L^2(\mathbb{R}_+,d^3k)$, which acts according to

$$(V(\varphi)g)(\vec{k}) = \int W(|\vec{k}-\vec{k}'|e^{i\varphi})g(\vec{k}')e^{3i\varphi}d^3k'. \tag{4.7}$$

The operator $R_0(\lambda,\varphi)$ on $L^2(\mathbb{R}_+,d^3k)$ consisting of multiplication by $(k^2e^{2i\varphi}-\lambda)^{-1}$ is the resolvent of a closed operator $H_0(\varphi)$ on L^2 with domain $\mathscr{D}(H_0(\varphi))$ consisting of all $f(k,\omega)$ such that $k^2f(k,\omega)$ is in L^2. The spectrum $\sigma[H_0(\varphi)]$ is the half line $0<|\lambda|<\infty$, $\arg\lambda = 2\varphi$ and the resolvent set $\rho[H_0(\varphi)]$ is given by (4.3). For φ fixed in $-\gamma \leqslant \varphi \leqslant \gamma$ and $\lambda \in \rho[H_0(\varphi)]$ the operator $V(\varphi)R_0(\lambda,\varphi)$ belongs to the Schmidt class on $L^2(\mathbb{R}_+,d^3k)$. Its Schmidt norm (9) satisfies

$$S[V(\varphi)R_0(\lambda,\varphi)] \leqslant \text{const}[\text{Im}(\lambda e^{-2i\varphi})^{1/2}]^{-1/2} \tag{4.8}$$

and therefore $V(\varphi)R_0(\lambda,\varphi)$ is an integral kernel in $\mathscr{K}$.

For $g \in \mathscr{D}(H_0(\varphi))$, there exist constants a,b such that[14]

$$\|V(\varphi)g\| \leqslant a\|H_0(\varphi)g\| + b\,\|g\|; \tag{4.9}$$

for any $\epsilon > 0$, the constant a may be chosen such that $a<\epsilon$ {this follows from the fact that for any λ in $\rho[H_0(\varphi)]$, there is an f for which $g = R_0(\lambda,\varphi)f$, and from the inequality (4.8)}.

The operator $H_0(\varphi) + V(\varphi)$ with domain $\mathscr{D}(H_0(\varphi))$ is a closed operator in $L^2(\mathbb{R}_+,d^3k)$ (by a theorem on stability of closedness under relatively bounded perturbations[17]).

The proof[15] of Eq. (4.8) follows from the convergence and uniform boundedness (with respect to φ and k') of

$$\int_0^\infty W(|k-k'|e^{i\varphi})k^2\,dk$$

in $-\gamma \leqslant \varphi \leqslant \gamma$.

Hence the bound (4.8) is valid for $\varphi = \alpha$ or β, and it follows that the κ-norm [the left side of (2.8) in kk' representation]

$$\kappa(K) = \Big\{\int\int (|K(\vec{k}e^{i\alpha},\vec{k}'e^{-i\alpha})|^2 + |K(\vec{k}^{\,i\beta},\vec{k}'e^{-i\beta})|^2)d^3k\,d^3k'\Big\}^{1/2} \tag{4.10}$$

of $VR_0(\lambda)$ satisfies

$$\kappa(VR_0(\lambda)) \leqslant \text{const}\{[\text{Im}(\lambda e^{-2i\alpha})^{1/2}]^{-1} + [\text{Im}(\lambda e^{-2i\beta})^{1/2}]^{-1}\}^{1/2}. \tag{4.11}$$

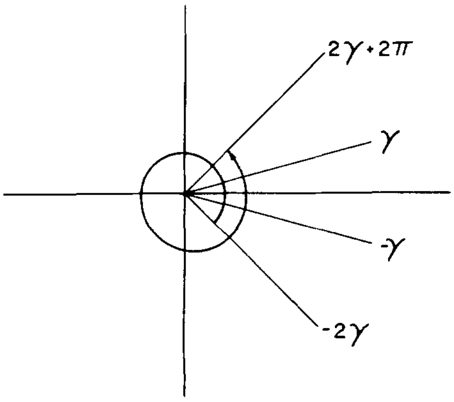


FIG. 1. Domain of analyticity for $\delta(\lambda)$ (example for $\gamma<\pi/2$).

Since H is a closed linear operator, we may define the resolvent

$$R(\lambda) = (H-\lambda)^{-1}, \tag{4.12}$$

which satisfies

$$R(\lambda) = R_0(\lambda) - R_0(\lambda)VR(\lambda). \tag{4.13}$$

Since $R_0(\lambda)V$ is in the $\mathscr{K}$ class, the Fredholm procedure may be applied to obtain[14]

$$R(\lambda) = R_0(\lambda) + \Delta(\lambda)R_0(\lambda)/\delta(\lambda), \tag{4.14}$$

where $\delta(\lambda)$ is a function independent of φ and analytic in the sector (see Fig. 1)

$$-2\min(\gamma,\tfrac{1}{2}\pi) < \arg\lambda < 2\pi + 2\min(\gamma,\tfrac{1}{2}\pi),$$

and $\Delta(\lambda)$ is an operator in $\mathscr{K}$ defined on $D(\alpha,\beta)$.

The equation in L^2 corresponding to (4.13) is

$$R(\lambda,\varphi) = R_0(\lambda,\varphi) - R_0(\lambda,\varphi)V(\varphi)R(\lambda,\varphi). \tag{4.15}$$

This equation is solved by the restriction of (4.14) to fixed φ, i.e.,

$$R(\lambda,\varphi) = R_0(\lambda,\varphi) + \Delta(\lambda,\varphi)R_0(\lambda,\varphi)/\delta(\lambda). \tag{4.16}$$

V. ISOLATED SINGULARITIES OF THE RESOLVENT

Let us denote the term in (4.15) which is not regular in the region $2\beta < \arg\lambda < 2\pi + 2\alpha$ as

$$F(\lambda,\varphi) = -R_0(\lambda,\varphi)V(\varphi)R(\lambda,\varphi). \tag{5.1}$$

For λ not too far from the negative real axis, according to (4.14) this corresponds to an operator $F(\lambda)$ in the class $\mathscr{K}$ on $D(\alpha,\beta)$, as well as an operator $F(\lambda,\varphi)$ which is in the Schmidt class in L^2. Either operator has the same kernel $F(\lambda,\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi})$ [as in Eq. (2.7)].

Let us first assume that $\alpha < 0 < \beta$, and let $\varphi = 0$. Since $H(0)$ is self-adjoint,

$$F^*(\bar{\lambda},0) = F(\lambda,0). \tag{5.2}$$

The kernel therefore has the property

$$\bar{F}(\bar{\lambda},\vec{k}',\vec{k}) = F(\lambda,\vec{k},\vec{k}'). \tag{5.3}$$

We now consider the Mellin transform [the three-dimensional form of Eq. (2.7)]

$$\mathbf{F}(\lambda,u,\omega,u',\omega',\varphi) = (1/2\pi)\int_0^\infty dk \int_0^\infty dk'\, e^{2i\varphi}(ke^{i\varphi})^{iu+1/2} \times (k'e^{-i\varphi})^{iu'+1/2} F(\lambda,\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}). \tag{5.4}$$

From (5.3), for $\varphi = 0$, it follows that

$$\bar{\mathbf{F}}(\bar{\lambda},u',\omega',u,\omega) = \mathbf{F}(\lambda,u,\omega,u',\omega'). \tag{5.5}$$

Now, using (2.7), we can reconstruct $\mathbf{F}(\lambda,u,\omega,u',\omega',\varphi)$ as

$$\begin{aligned}\mathbf{F}(\lambda,u,\omega,u',\omega',\varphi) &= e^{\varphi u}e^{-\varphi u' - i\varphi}\mathbf{F}(\lambda,u,\omega,u',\omega') \\ &= e^{\varphi u}e^{-\varphi u'}e^{-i\varphi}\bar{\mathbf{F}}(\bar{\lambda},u',\omega',u,\omega),\end{aligned} \tag{5.6}$$

i.e., we obtain the symmetry relation

$$\mathbf{F}(\lambda,u,\omega,u',\omega',\varphi) = \bar{\mathbf{F}}(\bar{\lambda},u',\omega',u,\omega,-\varphi). \tag{5.7}$$

Using the three-dimensional version of Eq. (2.9) for the inverse Mellin transform, we obtain

$$F(\lambda,\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}) = \bar{F}(\bar{\lambda},\vec{k}'e^{-i\varphi},\vec{k}e^{i\varphi}). \tag{5.8}$$

It follows from reality that, for $\varphi = 0$,

$$\bar{F}(\bar{\lambda},\vec{k},\vec{k}') = F(\lambda,\vec{k},\vec{k}'), \tag{5.9}$$

and, using (5.3), we obtain

$$F(\lambda,\vec{k},\vec{k}') = F(\lambda,\vec{k}',\vec{k}). \tag{5.10}$$

The Mellin transform for $\varphi = 0$ results in

$$\mathbf{F}(\lambda,u,\omega,u',\omega') = \mathbf{F}(\lambda,-u',\omega',-u,\omega), \tag{5.11}$$

and, using (2.7) again, we obtain

$$F(\lambda,\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}) = F(\lambda,\vec{k}'e^{i\varphi},\vec{k}e^{-i\varphi}). \tag{5.12}$$

For our purposes it will be useful to extend the validity of the relations obtained above to the case that $0 < \alpha < \beta$ or $\alpha < \beta < 0$, so that $\arg\lambda$ can exceed 2π or be negative (compare Figs. 2,3,4). To do this, suppose that we restrict φ to the interval $0 < \alpha' < \varphi < \beta$. The function $F(\lambda,\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi})$, considered as a kernel on $D(\alpha',\beta)$ satisfies the symmetry relations (5.7), (5.8), (5.11), and (5.12) for λ sufficiently close to the negative real axis; since it is analytic in λ, we can analytically continue the relation (5.12) to a region for which $\arg\lambda > 2\pi$. In this region, the function may have isolated singularities (marked ⊗ in Fig. 3), and, for this reason, although (5.11) remains valid, $\mathbf{F}(\lambda,u,\omega,u',\omega')$ may not be square-integrable. In the same way, we could choose to restrict φ to the interval $\alpha < \varphi < \beta' < 0$, and extend the validity of (5.12) to a region for which $\arg\lambda < 0$ (new singularities may occur in the region marked in Fig. 4). We now turn to the study of (5.8).

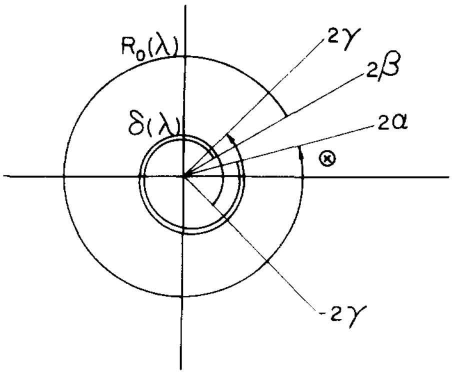


FIG. 3. Region of analyticity of $F(\lambda)$ for $0 < \alpha < \beta$.

To extend the applicability of Eq. (5.8), we first note that the left side can be continued to $\arg\lambda > 2\pi$ in the case $\varphi > 0$ and $F(\lambda,\vec{k}'e^{-i\varphi},\vec{k}e^{i\varphi})$ can be continued to $\arg\lambda < 0$. In order to maintain the condition $2\varphi < \arg\lambda < 2\varphi + 2\pi$, we can let $\bar{\lambda}$ denote the value of λ, for $\arg\lambda < 0$, to which we had continued the function F at $-\varphi$. Then the functions F at (φ,λ) and $(-\varphi,\bar{\lambda})$ are related by (5.8), even though they are not kernels on the same space $D(\alpha,\beta)$. In a later section, we shall define a space which includes both of the spaces on which these kernels act as subspaces.

As remarked by Van Winter,[15] the resolvent operator can be expanded in a Laurent series of the form

$$R(\lambda) = -(\lambda-\lambda_0)^{-1}P - \sum_{n=1}^{r}(\lambda-\lambda_0)^{-n-1}N^n + R_{\rm reg}(\lambda) \tag{5.13}$$

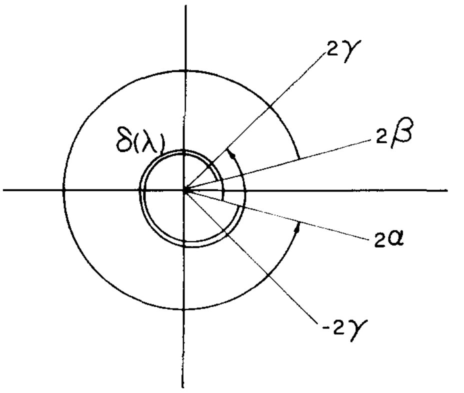


FIG. 2. Region of analyticity of $F(\lambda)$ for $\alpha < 0 < \beta$. Double line is overlap analyticity region of $\delta(\lambda)$, $R_0(\lambda)$.

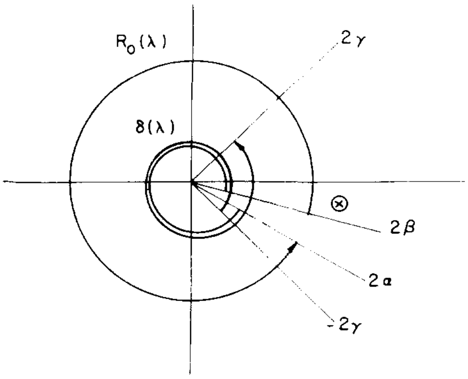


FIG. 4. Region of analyticity of $F(\lambda)$ for $\alpha < \beta < 0$.

around a pole at λ_0, where $R_{\text{reg}}(\lambda)$ is regular in an open set around λ_0. From (4.16) and (5.1), it follows that the operator (for Γ a contour enclosing only the singularity at λ_0)

$$P = -(1/2\pi i)\int_\Gamma R(\lambda)\,d\lambda$$

belongs to $\mathscr{K}$ on $D(\alpha,\beta)$ and has finite-dimensional range r and

$$N = (H-\lambda_0)P = -(1/2\pi i)\int_\Gamma (\lambda-\lambda_0)R(\lambda)\,d\lambda$$

is nilpotent such that $N^r = 0$. Note that $P^2 = P$ (but it is not necessarily self-adjoint), $N = PN = NP$, and that r is independent of φ (the proof is given by Van Winter,[14] where the fact that the multiplicity of a Schmidt class operator and its adjoint is the same is used in an essential way).

Suppose that the set $\{b_n(\vec{k}e^{i\varphi})\}$, $n = 1,\dots,r$, is a basis for the range of P. In the Mellin representation it follows that

$$\mathbf{b}_n(u,\omega,\varphi) = \int\!\!\int d\omega' du'\;\mathbf{P}(u,\omega,u',\omega')e^{-\varphi u'}b_n(u',\omega',\varphi). \tag{5.14}$$

According to (5.13) and the fact that the singularities in $R(\lambda)$ are due to singularities in $F(\lambda)$, the symmetry properties of the kernel $\mathbf{P}$ are the same as those for $\mathbf{F}$. It then follows from (5.11) that

$$\mathbf{b}_n(u,\omega,\varphi) = \int\!\!\int du' d\omega'\; e^{\varphi u} \times \mathbf{P}(-u',\omega'-u,\omega)e^{-\varphi u'}\mathbf{b}_n(u',\omega',\varphi)$$

and hence

$$\bar{\mathbf{b}}_n(-u,\omega,\varphi) = \int\!\!\int e^{-\varphi u} \times \bar{\mathbf{P}}(u',\omega',u,\omega)e^{\varphi u'}\,\bar{\mathbf{b}}_n(-u',\omega',\varphi)\,du'd\omega'. \tag{5.15}$$

The kernel in (5.15) is the kernel of the adjoint operator $\mathbf{P}^*(\varphi)$ in $\mathbf{L}^2$, the Mellin transform space of the L^2 defined in the discussion concerning Eq. (3.4); hence $\{\bar{\mathbf{b}}_n(-u,\omega,\varphi)\}$ is a basis for the range of $\mathbf{P}^*(\varphi)$. It therefore follows (since the $\{b_n\}$ are linearly independent functions) that

$$e^{\varphi u}\mathbf{P}(u,\omega,u',\omega')e^{-\varphi u'} = \sum_{nl} B_{nl}e^{\varphi u}\mathbf{b}_n(u,\omega)\mathbf{b}_l(-u',\omega')e^{-\varphi u'}. \tag{5.16}$$

In the following, we depart from the viewpoint taken by Van Winter.[15] If $e^{\varphi u - i\varphi/2}\mathbf{b}_n(u,\omega)$ is an element of $D(\alpha,\beta)$ for $\alpha<\varphi<\beta$, then $e^{-\varphi u' - i\varphi/2}\mathbf{b}_n(-u',\omega')$ is the complex conjugate of an element of $D(-\beta,-\alpha)$, with $-\beta<-\varphi<-\alpha$, i.e., the projection operator $\mathbf{P}$ will be interpreted as constructed in two different spaces and not, as in Ref. 15, constructed in a single space. Let us define

$$e^{-\varphi u + u\varphi/2}\,\mathbf{c}_n(u,\omega) = e^{-\varphi u + i\varphi/2}\bar{\mathbf{b}}_n(-u,\omega) \tag{5.17}$$

to obtain the form

$$e^{\varphi u}\mathbf{P}(u,\omega,u',\omega')e^{-\varphi u'} = \sum_{nl} B_{nl}e^{\varphi u}\mathbf{b}_n(u,\omega)\bar{\mathbf{c}}_l(u',\omega')e^{-\varphi u'}. \tag{5.18}$$

From the symmetry property (5.11) it follows that B_{nl} is symmetric and can therefore be diagonalized by a real orthogonal transformation. If

$$B_{nl} = \sum C_{nj}D_jC_{lj}$$

and we define

$$\mathbf{a}_j(u,\omega) = \sum_n \mathbf{b}_n(u,\omega)C_{nj},$$

$$\mathbf{d}_j(u,\omega) = \sum_n \mathbf{c}_n(u,\omega)C_{nj},$$

then (5.18) becomes (normalizing $\mathbf{a}_j$ and $\mathbf{d}_j$ so that all the D_j can be replaced by unity)

$$e^{\varphi u}\mathbf{P}(u,\omega,u',\omega')e^{-\varphi u'} = \sum_j e^{\varphi u}\mathbf{a}_j(u,\omega)\bar{\mathbf{d}}_j(u',\omega')e^{-\varphi u'}. \tag{5.19}$$

Since $\mathbf{Pa}_j = \mathbf{a}_j$, it follows that

$$\int\!\!\int \bar{\mathbf{d}}_j(u,\omega)\;\mathbf{a}_k(u,\omega)du\,d\omega = \delta_{jk}. \tag{5.20}$$

The inverse transformation (for the direction of $\vec{k}$ corresponding to ω), the generalization of (2.9) to three dimensions, is

$$P(\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}) = (1/2\pi)\int du\,du'(ke^{i\varphi})^{-iu-3/2}\mathbf{P}(u,\omega,u',\omega') \times \overline{(k'e^{-i\varphi})^{-iu'-3/2}}, \tag{5.21}$$

or with [note that $a_j(u,\omega)e^{u\varphi - i\varphi/2} = a_j(u,\omega,\varphi)$ is the Mellin transform of $a_j(\vec{k}e^{i\varphi})\cdot ke^{i\varphi}$ since we are working in three dimensions]

$$a_j(\vec{k}e^{i\varphi}) = [1/(2\pi)^{1/2}]\int du(ke^{i\varphi})^{-iu-3/2}\mathbf{a}_j(u,\omega), \tag{5.22}$$

and similarly for $\bar{d}_j(\vec{k}e^{-i\varphi})$ [corresponding to the conjugate of (5.22) with φ replaced by $-\varphi$], one obtains the representation

$$P(\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}) = \sum_j a_j(\vec{k}e^{i\varphi})\bar{d}_j(\vec{k}'e^{-i\varphi}), \tag{5.23}$$

where the element $a_j(\vec{k}e^{i\varphi})\in D(\alpha,\beta)$ corresponds to an isolated singularity of the resolvent. The basis functions satisfy

$$\int d^3k\, e^{3i\varphi}\bar{d}_j(\vec{k}e^{-i\varphi})a_k(\vec{k}e^{i\varphi}) = \delta_{jk}. \tag{5.24}$$

Note that this integral is independent of φ.

In case the two regions (α,β) and $(-\beta,-\alpha)$ overlap, they contain $\varphi = 0$. Let us consider this situation briefly. In this case, $R(\lambda,0)$ is the resolvent of an operator self-adjoint in L^2, and P must be of the form

$$P(\vec{k},\vec{k}') = \sum a_j(\vec{k})\bar{a}_j(\vec{k}') \tag{5.25}$$

so that $d_j(\vec{k}') = a_j(\vec{k}')$. Since, furthermore $a_j(\vec{k}')$ can be continued in the union of the two domains, (5.23) becomes

$$P(\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}) = \sum a_j(\vec{k}e^{i\varphi})\bar{a}_j(\vec{k}'e^{-i\varphi}) \tag{5.26}$$

and (5.24) becomes

$$\int d^3k\, e^{3i\varphi}\bar{a}_j(\vec{k}e^{-i\varphi})a_k(\vec{k}e^{i\varphi}) = \delta_{jk}. \tag{5.27}$$

Hence, for the case that an isolated pole is real, its corresponding analytic representation is of the form (5.23), but with $\bar{d}_j(\vec{k}'e^{-i\varphi})$ replaced by $\bar{a}_j(\vec{k}'e^{-i\varphi})$. The representation (5.23) is therefore a generalization of the structure of a self-adjoint projection operator. As we will show below, an operator of the type (5.23) can actually be defined in a partial inner product space,[4] in close analogy to the definition of a self-adjoint operator.

VI. Π-SELF-ADJOINTNESS

In this section we shall show that many of the notions available for operators in the usual Hilbert space are also available to operators constructed in two spaces. In particular, we shall show that the operator $P(\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi})$ for $\alpha<\varphi<\beta$ satisfies a symmetry relation analogous to self-adjointness in a generalized type of space. Let us define such a space by introducing the scalar product

$$(f,g)_\Pi = \int \bar{f}(\vec{k}e^{-i\varphi})g(\vec{k}e^{i\varphi})\, d^3k\, e^{3i\varphi}, \tag{6.1}$$

where $g(\vec{k}e^{i\varphi})\in D(\alpha,\beta)$ and $f(\vec{k}e^{-i\varphi})\in D(-\beta,-\alpha)$. In terms of the Mellin transformed functions,

$$(f,g)_\Pi = \int \bar{\mathbf{f}}(u,\omega)\mathbf{g}(u,\omega)du d\omega, \tag{6.2}$$

from which it is clear that $(f,g)_\Pi$ is independent of φ. We see from Eq. (5.24) that $a_k(\vec{k}e^{i\varphi})\in D(\alpha,\beta)$ is Π-orthogonal to $d_j(\vec{k}e^{-i\varphi})\in D(-\beta,-\alpha)$. In case $\mathbf{g}(u,\omega)\in D(-\beta,\beta)$ so that it is an L^2 function, $(g,g)_\Pi$ is the L^2-norm.

We remark that even though (6.2) is independent of φ, f and g are not, in general, L^2 functions. We may apply the Schwarz inequality, by providing the factors $e^{\varphi u - i\varphi/2}$ for $g(u,\omega)$ and $e^{-\varphi u + i\varphi/2}$ for $f(u,\omega)$ so that the integrand can be factored into L^2 functions, i.e.,

$$\begin{aligned}(f,g)_\Pi &= \int (\overline{e^{-\varphi u + i\varphi/2}\mathbf{f}(u,\omega)})(e^{\varphi u - i\varphi/2}\mathbf{g}(u,\omega))e^{i\varphi}\, du d\omega \\ &= \int \bar{\mathbf{f}}(u,\omega,-\varphi)\mathbf{g}(u,\omega,\varphi)e^{i\varphi}\, du d\omega,\end{aligned} \tag{6.3}$$

and hence

$$|(f,g)_\Pi|^2 \leqslant \left(\int |\mathbf{f}(u,\omega,-\varphi)|^2\, du d\omega\right)\left(\int |\mathbf{g}(u,\omega,\varphi)|^2\, du d\omega\right). \tag{6.4}$$

Since in (6.3), $f(-\varphi)$ or $g(\varphi)$ may vary through a dense set in L^2, $(f,g)_\Pi$ is nondegenerate.

We have defined the set $D(\alpha,\beta)$ such that $-\pi/2<\alpha,\beta<\pi/2$. With the inner product (6.2), the union of all spaces $D(\alpha,\beta)$ and $D(-\beta,-\alpha)$ forms a partial inner product space (Ref. 16, I), where $D(\pi/2,-\pi/2)$ is the set "compatible" (the scalar product exists) with all the space. It is also true that, for given α, β, $D(\alpha,\beta)\cup D(-\beta,-\alpha)$ is a partial inner product space, with $D(-\beta,\beta)$ as the compatible subset. This partial inner product space is a subspace of the larger union.

To compare the structure of this space with that of the Gel'fand triple,[9] we consider the four spaces, $D(\alpha,\beta)$, $D(-\beta,-\alpha)$, $D(-\beta,\beta)$, and $D(0,0)$ (L^2). Functions in $D(-\beta,\beta)$ form a dense subset of L^2. Furthermore, we may define the Π-scalar product of functions in $D(-\beta,\beta)$ with functions in $D(\alpha,\beta)$ and functions in $D(-\beta,-\alpha)$, in the following way. Let, for example, $\mathbf{g}(u,\omega)\in D(-\beta,\beta)$ and $\mathbf{f}(u,\omega)\in D(\alpha,\beta)$. Then, for $\alpha<\varphi<\beta$, (6.3) is valid, and hence

$$(f,g)_\Pi = \int \bar{\mathbf{f}}(u,\omega)\mathbf{g}(u,\omega)\, du d\omega \tag{6.5}$$

corresponds to a bounded [by (6.4)] linear functional of $g\in D(-\beta,\beta)$. This result has some similarity in structure with that of a Gel'fand triple, i.e.,

$$D(\alpha,\beta)\supset D(-\beta,\beta),$$

and $D(-\beta,\beta)$ is dense in L_2 in L_2-norm. It is, furthermore, true that in the $D(\alpha,\beta)$-norm there are sequences of elements of L_2 which converge to any element in $D(\alpha,\beta)$ but not all of L_2 is contained in $D(\alpha,\beta)$. This is a different situation from what one finds in the Gel'fand triple, in which the successively larger spaces contain the smaller, and is characteristic of the structure of a partial inner product space which looks like Fig. 5 (Muppet-like).

Antoine and Grossmann[16] have defined symmetric operators in partial inner product spaces. We shall apply the Antoine–Grossmann definition to operators mapping a partial inner product space into itself (See Ref. 16, II, Theorem 3.4):

In order to define the adjoint of an operator, we wish to prove the analog of the Reisz theorem for partial inner product spaces. Let $S=D(\alpha,\beta)\cup D(-\beta,-\alpha)$, $f\in D(\alpha,\beta)$, and $L{:}f\to\mathbb{C}$ such that

$$\begin{gathered}L(af_1+bf_2)=aL(f_1)+bL(f_2),\\ |L(f)|\leqslant K\min_\varphi\int |f(ke^{i\varphi})|^2\, d^3k \leqslant K\,\|f\|_{D(\alpha,\beta)};\end{gathered} \tag{6.6}$$

for f in $D(\alpha,\beta)$ we must also consider the mapping of $f(\varphi)$. We shall assume that

$$L(f(\varphi))=L(f) \tag{6.7}$$

independently of φ, for all $\alpha\leqslant\varphi\leqslant\beta$. As in (3.4), we may define a linear functional acting in the subset associated with φ as

$$L(\varphi)(f)=L(f(\varphi)); \tag{6.8}$$

the functional $L(\varphi)$, according to (6.7), is φ-independent.

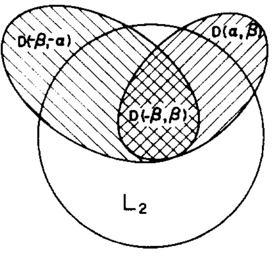


FIG. 5. Structure of the partial inner product space.

Suppose that the set W of elements of $D(\alpha,\beta)$ are mapped to zero, i.e.,

$$W = \{f \in D(\alpha,\beta) | L(f) = 0\} \tag{6.9}$$

and suppose W_{L^2} in the L^2 closure of $D(\alpha,\beta)$ such that ($\bar{L}$ is defined through the continuity of L)

$$W_{L^2} = \{g \in L^2 | \bar{L}(g) = 0\}. \tag{6.10}$$

We suppose that the complement of W_{L^2} in L^2 is not empty and denote it by $W^{\perp}_{L^2}$. For $g^* \in W^{\perp}_{L^2}$, suppose

$$\bar{L}(g^*) = 1. \tag{6.11}$$

It now follows that

$$g - \bar{L}(g) g^*$$

is in W_{L^2} and hence

$$(g^*, g - \bar{L}(g) g^*)_{L^2} = 0.$$

We thus obtain

$$\bar{L}(g) = (g^*, g)_{L^2} / \|g^*\|^2_{L^2}.$$

Taking f to be in $D(\alpha,\beta)$, φ independence of

$$L(f) = (g_L, f)_{L^2} \tag{6.12}$$

implies that $g_L = g^* / \|g^*\|^2_{L^2} \in D(-\beta, -\alpha)$, i.e., in Mellin transform representation, $\mathbf{g}_L(u,\omega,\varphi) = \mathbf{g}_L(u,\omega) e^{-u\varphi} e^{-i\varphi/2}$, so that

$$L(f) = \int \bar{\mathbf{g}}_L(u,\omega) \mathbf{f}(u,\omega) \, du d\omega. \tag{6.13}$$

This linear functional is of the form given in (6.5), i.e.,

$$L(f) = (g_L, f)_{\Pi}. \tag{6.14}$$

According to the usual Reisz theorem, applied to $D(\alpha,\beta)$ as a Hilbert space, it follows from the second inequality (weaker) in (6.6) that, for linear functions of this type, there exists an element $g_R \in D(\alpha,\beta)$ such that

$$L(f) = (g_R, f)_{D(\alpha,\beta)}. \tag{6.15}$$

In Mellin transform representation, it follows that

$$\mathbf{g}_R(u,\omega) = \mathbf{g}_L(-u,\omega). \tag{6.16}$$

With the help of the Reisz theorem, we may find the elements in $D(\alpha,\beta)$ which are orthogonal (i.e., the Π scalar product vanishes) to all of the elements in $D(-\beta, -\alpha)$ which correspond to isolated singularities of the resolvent. These elements in $D(\alpha,\beta)$ contain the generalized states associated with the continuous spectrum.

Let $g \in D(\alpha,\beta)$, $f \in D(-\beta, -\alpha)$, and A be a linear operator defined on some dense domain in $D(\alpha,\beta)$ [in the norm given by (1.4)] as $A_{\alpha,\beta}$, and on some dense domain in $D(-\beta, -\alpha)$ as $A_{\overline{\alpha\beta}}$. Then, for g in the domain of $A_{\alpha\beta}$ there exists a set of f's in $D(-\beta, -\alpha)$ (the domain of the adjoint) such that

$$b(f,g) = (f, A_{\alpha\beta} g)_{\Pi} \tag{6.17}$$

is a bilinear form L^2-continuous in g. It then follows that this bilinear form defines a linear operator $A^{*}_{\overline{\alpha\beta}}$ on $D(-\beta, -\alpha)$, satisfying

$$b(f,g) = (A^{*}_{\overline{\alpha\beta}} f, g)_{\Pi}. \tag{6.18}$$

Then, $A^{*}_{\overline{\alpha\beta}}$ is defined as the Π-adjoint of A in $D(-\beta, -\alpha)$ [in a similar way, $A^{*}_{\alpha\beta}$ is obtained by starting with the bilinear form $(g, A_{\overline{\alpha\beta}} f)_{\Pi}$, and is defined as the Π-adjoint of A in $D(\alpha,\beta)$].

We first remark that a set of vectors dense in $D(\alpha,\beta)$ is dense in the L^2 sense as well; i.e., given that for any $g \in D(\alpha,\beta)$ there exists an f in the dense set $\mathscr{D}(A)$ in $D(\alpha,\beta)$ such that

$$\|f - g\|_{D(\alpha,\beta)} < \epsilon, \tag{6.19}$$

then, for any $g' \in L^2$, there exists an $f \in \mathscr{D}(A)$ (for any φ in $\alpha < \varphi < \beta$) such that

$$\int |f(\vec{k}e^{i\varphi}) - g'(\vec{k})|^2 d^3k < \epsilon. \tag{6.20}$$

The result (6.20) follows from the fact that the $\{g(\vec{k}e^{i\varphi})\} \in D(\alpha,\beta)$ are dense in L^2 (Sec. III), so that in the expression

$$\int |f(\vec{k}e^{i\varphi}) - g(\vec{k}e^{i\varphi}) + g(\vec{k}e^{i\varphi}) - g'(\vec{k})|^2 d^3k$$
$$< \int |f(\vec{k}e^{i\varphi}) - g(\vec{k}e^{i\varphi})|^2 d^3k$$
$$+ \int |g(\vec{k}e^{i\varphi}) - g'(\vec{k})|^2 \, d^3k$$

there exists a $g(\vec{k}e^{i\varphi})$ such that

$$\int |g(\vec{k}e^{i\varphi}) - g'(\vec{k})|^2 \, d^3k < \epsilon/2$$

and it follows from (6.19) that there is an $f(\vec{k}e^{i\varphi}) \in \mathscr{D}(A)$ for which

$$\int |f(\vec{k}e^{i\varphi}) - g(\vec{k}e^{i\varphi})|^2 \, d^3k < \epsilon/2.$$

Consider now the bilinear form (6.17) in terms of the Mellin transform:

$$\begin{aligned} b(f,g) &= \int \bar{\mathbf{f}}(u,\omega) e^{-\varphi u} e^{-i\varphi/2} \{ e^{\varphi u - \varphi u'} \mathbf{A}_{\alpha\beta}(u,\omega,u',\omega') \\ &\qquad \times e^{\varphi u' - i\varphi/2} \mathbf{g}(u',\omega) e^{i\varphi} \, du du' d\omega d\omega' \} \\ &= \int \bar{\mathbf{f}}(u,\omega) e^{-\varphi u} (e^{\varphi u} \mathbf{A}_{\alpha\beta}(u,\omega,u',\omega') e^{-\varphi u'}) \\ &\qquad \times e^{\varphi u'} \mathbf{g}(u',\omega) du du' d\omega d\omega'. \end{aligned} \tag{6.21}$$

Since the domain of A in $D(\alpha,\beta)$ is dense, according to the result (6.20), the $\{e^{\varphi u'} \mathbf{g}(u',\omega)\}$ which can occur in (6.21) are also dense in L^2, and hence we can apply the Reisz theorem as proved above.

In case

$$A^{*}_{\alpha\beta} \supseteq A_{\overline{\alpha\beta}} \quad \text{and} \quad A^{*}_{\overline{\alpha\beta}} \supseteq \mathrm{A}_{\alpha\beta}, \tag{6.22}$$

we shall say that A is Π-symmetric (it can be closed) and if the equality holds in (6.22), A will be said to be Π-self-adjoint.

The operator P, as given in (5.23), explicitly satisfies the definition for a self-adjoint operator. To show that the operator H is Π-self-adjoint, we shall use the result of Sec. V. From the explicit form of $R_0(\lambda,\varphi)$ and Eq. (5.8), one finds that

$$R(\lambda, \vec{k}e^{i\varphi}, \vec{k}'e^{-i\varphi}) = \bar{R}(\bar{\lambda}, \vec{k}'e^{-i\varphi}, \vec{k}e^{i\varphi}). \tag{6.23}$$

By the definition of the resolvent,

$$\begin{aligned} &\int [\lambda \delta^3(\vec{k} - \vec{k}') e^{-3i\varphi} - H(\vec{k}e^{i\varphi}, \vec{k}'e^{-i\varphi})] R(\lambda, \vec{k}'e^{i\varphi}, \vec{k}''e^{-i\varphi}) \\ &\qquad \times e^{3i\varphi} \, d^3k' \\ &\quad = \delta^3(\vec{k} - \vec{k}'') e^{-3i\varphi}. \end{aligned} \tag{6.24}$$

By conjugating (6.24) and using (6.23), it follows that

$$H(\vec{k}e^{i\varphi},\vec{k}'e^{-i\varphi}) = \overline{H}(\vec{k}'e^{-i\varphi},\vec{k}e^{i\varphi}) \tag{6.25}$$

and hence H is Π-self-adjoint.

For completeness, we conclude this section with a brief and elementary illustration of the notions developed here for the resonance problem. We shall, in particular, describe the method of associating the decay of a state in time, due to the presence of a complex pole in the resolvent, with an element of $D(\alpha,\beta)$. The relation between decaying states and scattering resonance phenomena has been described, for example, by Horwitz and Marchand.[18]

We shall use $H(\vec{k}e^{-i\theta},\vec{k}'e^{i\theta})$ as the kernel of $H(\theta)$ to calculate the decay of a state ψ, where ψ and $H(\theta)$ have the following properties:

(1) $\psi \in D(\beta,-\beta)$, $\alpha<\beta$;

(2) $H(\alpha)$ has an eigenvalue at $E=E_0-i\Gamma/2$ of multiplicity 1 (see Ref. 12 for a more general situation), associated with eigenvector $\phi_0 \in D(\alpha,\beta)$.

(3) $\psi(\vec{k}e^{i\alpha})$, the analytic continuation of ψ from an L^2 function on the real line to a function defined along a line at angle α, is close to ϕ_0 in the Π norm.

The probability that ψ (the state at $t=0$) remains ψ at time t is

$$p(t) = |(\psi,e^{-iHt}\psi)|^2. \tag{6.26}$$

We shall use

$$a(t) \equiv (\psi,e^{-iHt}\psi) = (1/2\pi i)\int_C (\psi,(H-\lambda)^{-1}\psi)e^{-i\lambda t}\,d\lambda, \tag{6.27}$$

where C is a contour going around $\sigma(H)$, the spectrum of H (positive real half-line).

We now rotate C and replace $(\psi,(H-\lambda)^{-1}\psi)$ with $(\psi_\alpha,[H(\alpha)-\lambda]^{-1}\psi_\alpha)_\Pi$, where $\psi(ke^{-i\alpha}) \in D(-\beta,-\alpha)$ and $\psi(ke^{i\alpha}) \in D(\alpha,\beta)$ occur in the Π scalar product. Then,

$$a(t) = (1/2\pi i)\int_{C_1} (\psi_{\alpha_1},\phi_0)_\Pi(\phi_0,[H(\alpha)-\lambda]^{-1}\phi_0)_\Pi \times(\phi_0,\psi_\alpha)_\Pi e^{-i\lambda t}\,d\lambda + R(t), \tag{6.28}$$

where C_1 is around the eigenvalue E, and $R(t)$ represents contributions from other singularities. We now use the Laurent expansion (5.13) in the form

$$[H(\alpha)-\lambda]^{-1} = [1/(E-\lambda)]|\phi_0\rangle\,\langle\phi_0| + R_{\text{reg}}(\lambda) \tag{6.29}$$

to obtain

$$a(t) = e^{-i\epsilon t}(\psi_\alpha,\phi_0)_\Pi(\phi_0,\psi_\alpha)_\Pi + R(t),$$

where $R(t)$ is a relatively slowly varying function. This result shows the characteristic exponential decay law associated with a resonance pole.

VII. THE VARIATIONAL PRINCIPLE[19]

Let us consider the symmetric bilinear form (in Mellin transform representation)

$$(f|g) = \int \mathbf{f}(-u,\omega)\mathbf{g}(u,\omega)\,du\,d\omega = (g|f), \tag{7.1}$$

where $f, g \in D(\alpha,\beta)$. This form is equivalent to the Π scalar product (6.3), but in this type of representation Π-self-adjoint operators appear as symmetric. Although $(f|f)$ is not a norm in the usual sense, it is useful in providing a scale for comparison. It is also nondegenerate, since for $f(-u,\omega) = \bar{h}(u,\omega) \in D(-\beta,-\alpha)$, $(f|g) = (h,g)_\Pi$. For example, the quantity [for $(f|f)$,$(g|g)$ not zero]

$$F(f) = (f|g)/|(f|f)|^{1/2} \tag{7.2}$$

is stationary when $f(x) = zg(x)$, in which case it is equal to $(g|g)^{1/2}$. Consider the expansion of $F(f)$ in the neighborhood of $f=g$, i.e., for $f=g+\epsilon h$, where ϵ (real) is small, and h is arbitrary in $D(\alpha,\beta)$:

$$[F(g+\epsilon h)]^2 = |(g|g)|\{1-[\epsilon^2/|(g|g)|^2]\,\mathrm{Re}[C(h,g)] + O(\epsilon^3)\}, \tag{7.3}$$

where

$$C(h,g) = (h|h)(g|g)^* - [1/|(g|g)|^2(h|g)^2(g|g)^{*2}. \tag{7.4}$$

A similar calculation in the usual complex Hilbert space yields the expression $\|h\|^2\|g\|^2 - |(h,g)|^2 \geqslant 0$, where zero is achieved only if h is proportional to g, in place of $C(h,g)$. It is also true that $C(h,g)$ vanishes when h is proportional to g, i.e., for $h=zg$, any complex z. Since

$$C(e^{i\delta}h,g) = e^{2i\delta}C(h,g), \tag{7.5}$$

if $C(h,g)$ is not zero, we may pick the phase of h so that $F(g+\epsilon h)$ is a maximum (or minimum) as ϵ goes through zero.

To interpret this situation geometrically, let us examine the problem of constructing projections. Let g range over a closed linear manifold M in $D(\alpha,\beta)$, and suppose f is a given element of $D(\alpha,\beta)$ (which may or may not be in M). We now seek a g in M such that

$$G(g) = |(f-g|f-g)| \tag{7.6}$$

is *stationary*. Suppose that g_0 is an element of M which satisfies this requirement. Then,

$$\begin{aligned} G(g_0+\epsilon g) &= |(f-g_0+\epsilon g|f-g_0+\epsilon g)|^2 \\ \ldots &= |(h+\epsilon g|h+\epsilon g)|^2 \end{aligned} \tag{7.7}$$

is stationary as a function of ϵ, for arbitrary $g \in M$, at $\epsilon=0$. Expanding as a function of real ϵ, we obtain

$$\begin{aligned} G(g_0+\epsilon g) = {}& |(h|h)|^2 + 2\epsilon^2[\mathrm{Re}(h|h)(g|g)^* + 2|(h|g)|^2] \\ & + \epsilon^4|(g|g)|^2 + 4\epsilon^3\mathrm{Re}(g|g)(h|g)^* \\ & + 4\epsilon\mathrm{Re}(h|h)(h|g)^* \end{aligned} \tag{7.8}$$

and since the first derivative with respect to ϵ must vanish at $\epsilon=0$,

$$\mathrm{Re}(h|h)(h|g)^* = 0. \tag{7.9}$$

This condition must be valid for all $g \in M$ and hence, for g replaced by ig. It follows that if $(h|h) \neq 0$,

$$(h|g) = 0. \tag{7.10}$$

It is not true, however, that $G(g_0+\epsilon g)$ will be larger than $G(g_0)$ for any g in M. Since $(h|g)=0$, the positivity of the second derivative requires

$$\mathrm{Re}(h|h)(g|g)^* \geqslant 0. \tag{7.11}$$

This condition is sensitive to the phase of g in M, and, for any g, a change in phase can be introduced to satisfy (7.11) [for $(g|g) \neq 0$; in case $(g|g)=0$, $G(g_0+\epsilon g)$ does not depend on ϵ]. Although we shall not use this notion in an essential way

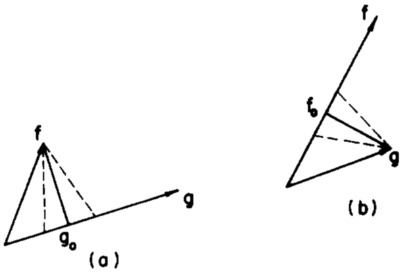


FIG. 6. Geometrical notion of orthogonality.

later, we could consider (7.11) to be part of the criterion for the "orthogonality" of two vectors, h and g, according to the form (7.1). To associate a geometrical intuition with the idea of orthogonality, each vector should be minimum under small perturbations parallel to the other. In particular, if the M utilized in (7.6) is one-dimensional, one has the picture of Fig. 6. Under small additions to g_0 [Fig. 6(a)], the "distance" of f to the corresponding point on the manifold should increase, and similarly for g. We now show that the phases of f and g can always be picked so that this picture is valid:

Let

$$h = f - ag, \quad h' = g - bf \tag{7.12}$$

so that $(h|g) = 0$ and $(h'|f) = 0$. It follows that (assume $(g|g) \neq 0$, $(f|f) \neq 0$)

$$a = (f|g)/(g|g), \quad b = (f|g)/(f|f).$$

Then,

$$(h|h) = (f|f) - (f|g)^2/(g|g), \quad (h'|h') = (g|g) - (g|f)^2/(f|f). \tag{7.13}$$

The condition (7.11) for each of these, implying "orthogonality," is [with the definition (7.4)]

$$\mathrm{Re}C(f,g) > 0, \quad \mathrm{Re}C(g,f) > 0. \tag{7.14}$$

The inequalities would imply that

$$|(h + \epsilon g|h + \epsilon g)| > |(h|h)| \tag{7.15}$$

for ϵ real.

If we let $f \rightarrow e^{i\delta} f$, then the requirement (7.14) becomes

$$\mathrm{Re}e^{2i\delta}C(f,g) > 0, \tag{7.16}$$

where we have used the fact that $C(g, e^{i\delta}f) = e^{-2i\delta}C(g,f)$. Such a rotation in the complex plane can turn any complex number to a position satisfying the required inequalities.

Returning to Eq. (7.3), let us consider a variation ϵh which is *orthogonal* to g in the sense of (7.10) and (7.11). The choice of phase in (7.5) is then such that the functional $F(f)$ defined in (7.2) is a minimum. The geometrical significance of the form $(f|g)$ is therefore similar to that of the usual complex Hilbert space scalar product, provided that the phases are chosen appropriately.

We now turn to the notion of projection operators. We say that a basis for a finite-dimensional manifold M, such that $\overline{\{f_i\}} = M$ in the $D(\alpha,\beta)$ topology, is nonnull if

$$\det(f_i|f_j) \neq 0. \tag{7.17}$$

This condition is necessary and sufficient for the existence of an orthonormal basis $\{\varphi_i\}$ such that $(\varphi_i|\varphi_j) = \delta_{ij}$. Since $(f_i|f_j)$ is symmetric, there is an orthogonal transformation A such that

$$\sum_{ij} A_{ki}(f_i|f_j)A_{lj} = \delta_{hl}\lambda_k,$$

where, by (7.17), $\lambda_k \neq 0$. Then, $\{\varphi_k = [1/(\lambda_k)^{1/2}]\sum_i A_{ki}f_i\}$ is the orthrnormal set. Suppose, on the other hand,

$$\det(f_i|f_j) = 0. \tag{7.18}$$

Then, there is no transformation to an orthonormal basis. Suppose

$$\varphi_i = \sum T_{ij}f_j$$

is an orthonormal basis. Then,

$$(\varphi_i|\varphi_j) = \sum_{kl} T_{ik}(f_k|f_l)T_{jl}$$

and

$$\det(\varphi_i|\varphi_j) = (\det T_{ij})^2 \det(f_i|f_j) = 0,$$

in contradiction with the assertion that the $\{\varphi_i\}$ form an orthonormal basis.

Not every manifold is nonnull. As we have remarked, $(f|g)$ is nondegenerate, since $(f|g) = 0$ for all $g \in D(\alpha,\beta)$ implies that $e^{-u\varphi}\bar{\mathbf{f}}(-u,\omega)$ is orthogonal in the L^2 sense to all of $\{e^{\varphi u}\mathbf{g}(u,\omega)\} = L^2$. For g restricted to M, a subspace of $D(\alpha,\beta)$, $(f|g) = 0$ implies only that $e^{-u\varphi}\bar{\mathbf{f}}(-u,\omega)$ is orthogonal, in the L^2 sense, to a subspace of L^2.

We have shown that every finite-dimensional manifold spanned by a nonnull basis (which we shall call a nonnull manifold) has an orthogonal basis. For such a nonnull manifold M, the decomposition

$$f = g_0 + h, \tag{7.19}$$

where $(h|g) = 0$ for all $g \in M$ and $g_0 \in M$ is unique. Let

$$g_0 = \sum a_i \varphi_i, \tag{7.20}$$

where $\{\varphi_i\}$ is the orthonormal basis in M. Then,

$$a_j = (\varphi_j|f)$$

and

$$h = f - \sum_j \varphi_j(\varphi_j|f). \tag{7.21}$$

Since

$$(h|h) = (f|f) - \sum_j (\varphi_j|f)^2,$$

the phase of f can always be chosen so that $\mathrm{Re}(h|h) > 0$, and hence $|(h + \Sigma\epsilon_i\varphi_i|h + \Sigma\epsilon_i\varphi_i)| > |(h|h)|$ ("orthogonality" of h to the orthonormal basis in M). We see that the mapping

$$f \rightarrow \sum_j \varphi_j(\varphi_j|f) \tag{7.22}$$

is precisely that of a projection operator of the type (5.26).

Let us now consider the "normalized" bilinear form (for $(g|g) \neq 0$)

$$E(g) = \frac{\int \mathbf{g}(-u,\omega)\mathbf{H}(u,\omega,u',\omega')\mathbf{g}(u',\omega')\, du du' d\omega d\omega'}{\int \mathbf{g}(-u,\omega)\mathbf{g}(u,\omega)\, du d\omega} \tag{7.23}$$

$$= \frac{(g|H|g)}{(g|g)}, \tag{7.24}$$

where, in terms of our notation of Eq. (6.2), we have taken $\bar{\mathbf{f}}(u,\omega)e^{-\varphi u} = \mathbf{g}(-u,\omega)e^{-\varphi u} \in D(-\beta, -\alpha)$. Then, for $\mathbf{h}(u,\omega)e^{\varphi u} \in D(\alpha,\beta)$,

$$\begin{aligned} E(g+\epsilon h) = E(g) &+ [2\epsilon/(g|g)][(h|H|g) - (h|g)E(g)] \\ &+ [\epsilon^2/(g|g)]\{(h|H|h) + E(g) \\ &\times [4(h|g)^2/(g|g) - (h|h)] \\ &- 4(h|H|g)(h|g)/(g|g)\} + O(\epsilon^3), \end{aligned} \tag{7.25}$$

using the symmetry property (5.11), i.e., $(h|H|g) = (g|H|h)$.

If g is a stationary point for the functional $E(g)$, then the linear term in ϵ must vanish:

$$(h|H|g) = E(g)(h|g) \tag{7.26}$$

for all $\mathbf{h}(u,\omega)e^{u\varphi} \in D(\alpha,\beta)$; since the bilinear form is nondegenerate, this implies

$$Hg = E(g)g, \tag{7.27}$$

i.e., $E(g)$ is an eigenvalue of H which is equal to the λ discussed in Sec. V, and $\mathbf{g}(u,\omega)e^{u\varphi} \in D(\alpha,\beta)$ is an eigenfunction. With (7.27), (7.25) becomes (if $(h|h) \neq 0$)

$$\begin{aligned} E(g+\epsilon h) = E(g) &+ \epsilon^2[(h|h)(g|g)^*/|(g|g)|^2] \\ &\times [E(h) - E(g)]. \end{aligned} \tag{7.28}$$

According to the symmetry (5.7), we have

$$e^{\varphi u}\mathbf{H}(u,\omega,u',\omega')e^{-\varphi u'} = e^{-\varphi u'}\overline{\mathbf{H}}(u',\omega',u,\omega)e^{\varphi u}. \tag{7.29}$$

Then, conjugating

$$\int e^{\varphi u}\mathbf{H}(u,\omega,u',\omega')e^{-\varphi u'}e^{\varphi u'}\mathbf{g}(u',\omega')\, du'd\omega'$$

$$= E(g)e^{\varphi u}\mathbf{g}(u,\omega), \tag{7.30}$$

we obtain

$$\int e^{-\varphi u}\mathbf{H}(u,\omega,u',\omega')e^{\varphi u'}e^{-\varphi u'}\bar{\mathbf{g}}(-u',\omega')\, du'd\omega'$$

$$= \bar{E}(g)e^{-\varphi u}\bar{g}(-u,\omega). \tag{7.31}$$

Each eigenvalue $E(g)$ corresponding to an eigenfunction $g(u,\omega)e^{u\varphi} \in D(\alpha,\beta)$ implies the existence of the complex conjugate eigenvalue for an eigenfunction $e^{-u\varphi}\bar{\mathbf{g}}(-u,\omega) \in D(-\beta,-\alpha)$ [as in (5.17)]. If $(g|g)$ is *maximal*, in the sense

$$\left| \int \mathbf{g}(-u,\omega)e^{-u\varphi}\mathbf{g}(u,\omega)e^{u\varphi}\, du \right|$$

$$\leqslant \int |\mathbf{g}(u,\omega)e^{u\varphi}|^2\, du,$$

and the equality is realized, then

$$\mathbf{g}(-u,\omega)e^{-u\varphi} = z\bar{\mathbf{g}}(u,\omega)e^{u\varphi} \tag{7.32}$$

and the variational principle (7.23), that E be stationary about g, reduces to the usual Hilbert space variational principal for a self-adjoint operator. In this case, $E(g)$ is *real*. Substituting (7.32) into (7.31), we obtain

$$\int e^{-\varphi u}\mathbf{H}(u,\omega,u',\omega')e^{\varphi u'}e^{\varphi u'}\mathbf{g}(u',\omega')\, du'd\omega'$$

$$= E(g)e^{\varphi u}\mathbf{g}(u,\omega); \tag{7.33}$$

hence, $e^{\varphi u}g(u,\omega)$ is an eigenfunction of $H(-\varphi)$ as well. By (7.32), $D(\alpha,\beta)$ and $D(-\beta,-\alpha)$ overlap. From this, it follows that $\mathbf{g}(u,\omega)$ is an L^2 function, and can be continued to both sides of the real axis (as already mentioned).

We now return to (7.28). In general,

$$\begin{aligned} |E(g+\epsilon h)|^2 = |E(g)|^2 &+ [2\epsilon^2/|(g|g)|^2] \\ &\times \mathrm{Re}[(E(g)^*E(h) - |E(g)|^2)(h|h)(g|g)^*] \\ &+ O(\epsilon^4) \end{aligned} \tag{7.34}$$

is not definitely greater or less than $|E(g)|^2$. The sign of the coefficient of ϵ^2 is given by

$$\sigma = \mathrm{sgn}\{|E(h)|\cos(\varphi_h - \varphi_g + \theta) - |E(g)|\cos\theta\}, \tag{7.35}$$

where

$$\varphi_n = \arg[E(h)], \ \varphi_g = \arg[E(g)], \ \theta = \arg[(h|h)(g|g)^*]. \tag{7.36}$$

Experimentally recognizable resonance poles generally lie close to the real axis (their imaginary parts are small compared to their real parts), in the lower half-plane, and hence it is of interest to consider the case when h also corresponds to an eigenfunction for eigenvalue $E(h)$, and φ_n, φ_g are *small*. If

$$|E(g)| < |E(h)|,$$

for example, if $E(g)$ is the eigenvalue of smallest magnitude, σ is determined by the sign of $\cos\theta$. It follows by the usual method that if $E(h) \neq E(g)$, $(g|h) = 0$. With our notion (7.10), (7.11) of "orthogonality," $\cos\theta \geqslant 0$, and $|E(g)|$ is therefore minimal with respect to the addition of a real multiple of another eigenvector [when the phase of this orthogonal eigenvector is chosen to satisfy (7.11)].

In case the angles φ_h and φ_g are small, the imaginary parts of $E(h)$ and $E(g)$ are also small, and the ordering in terms of the magnitudes is almost equivalent to an ordering in terms of the real parts of the complex eigenvalues. In fact, let us consider (for h, g eigenvectors)

$$\begin{aligned} \mathrm{Re}[E(g+\epsilon h)] = \mathrm{Re}[E(g)] &+ [\epsilon^2/|(g|g)|^2] \\ &\times \mathrm{Re}[(h|h)(g|g)^*(E(h) - E(g)]. \end{aligned} \tag{7.37}$$

The "orthogonality" condition between h and g required only that $\mathrm{Re}[(h|h)(g|g)^*] \geqslant 0$, and admits the possibility that

$$\mathrm{Im}[(h|h)(g|g)^*] = 0. \tag{7.38}$$

Assuming (7.38), it follows that the coefficient of ϵ^2 is positive if $\mathrm{Re}[E(h)] > \mathrm{Re}[E(g)]$, i.e., that $E(g)$ has the smaller real part. One can see that a similar conclusion is reached for the imaginary part. Again, assuming (7.38), we obtain

$$\begin{aligned} \mathrm{Im}[E(g+\epsilon h)] = \mathrm{Im}[E(g)] &+ [\epsilon^2/|(g|g)|^2] \\ &\times \mathrm{Re}[(h|h)(g|g)^*]\mathrm{Im}[E(h) - E(g)] \end{aligned} \tag{7.39}$$

so that the coefficient of ϵ^2 is positive if $\mathrm{Im}[E(h)] > \mathrm{Im}[E(g)]$. The minimal properties of the form $E(g)$ that we have obtained above refer to h as an eigenvector with eigenvalue $E(h)$ satisfying the "orthogonality" conditions (7.11) and possibly

(7.38) [$(h|g) = 0$ was not actually used]. Hence a variation $g \rightarrow g + \epsilon h$ for any h satisfying (7.11) and (7.38) will have these properties provided that $|E(h)| > |E(g)|$, $\mathrm{Re}[E(h)] > \mathrm{Re}[E(g)]$ or $\mathrm{Im}[E(h)] > \mathrm{Im}[E(g)]$. It is possible to order eigenvalues in this sense, but, for arbitrary h, these inequalities cannot be assured; hence the variational principle cannot be used directly to establish a bound. In particular, suppose $\{g_i\}$ to be eigenfunctions, and

$$h = \sum_{i=1}^{n} a_i g_i , \tag{7.40}$$

where $(g_i|g_0) = 0$ for $i \neq 0$, and $(g_i|g_j) = \delta_{ij}$. Then we can take $(h|h) = \Sigma_{i\neq 0} a_i^2 = 1$, and the coefficient of ϵ^2 in (7.28) would be

$$\sum_{i\neq 0} a_i^2 E(g_i) - E(g_0). \tag{7.41}$$

Since the a_i^2 are not real, the ordering of $\{E(g_i)\}$ does not determine the sign (for $n > 1$) of the real part or the imaginary part of (7.41), or of $\mathrm{Re}[\Sigma \alpha_i^2 E(g_i) E(g_0)^* - |E(g_0)|^2]$.

VIII. CONCLUSIONS

We have shown that complex poles of the resolvent of an operator which is self-adjoint in L^2 can be put into correspondence with complex eigenvalues of an analytic extension (if it exists) of the operator, with eigenfunctions in a space $D(\alpha,\beta)$. These eigenfunctions are not normalizable, and their scalar products with other elements of $D(\alpha,\beta)$ may not be defined. There is another, dual, space, $D(-\beta, -\alpha)$, however, with which one can define linear functionals on $D(\alpha,\beta)$ and construct bilinear forms with some of the properties of a scalar product in Hilbert space. In particular, the analog of the Riesz theorem is valid, and the geometrical interpretation of the bilinear form, with appropriate choices of phase, has some features in common with that of L^2 scalar products.

In this (Π-) product, the extended Hamiltonian, and the projection operators associated with its discrete complex eigenvalues, are Π-self-adjoint. The symmetry property discussed by Schieve and Bailey[10], though discussed in a more complicated way, corresponds to Π-self-adjointness. The space of eigenfunctions $D(\alpha,\beta)$ and its dual space, together with L^2, in which they are dense, form a partial inner product space of the type described by Antoine and Grossmann.[16]

The eigenfunctions corresponding to discrete complex eigenvalues satisfy a variational principle. Although the bilinear form satisfies ordering relations when perturbed in the direction of an eigenfunction, bounds of the type obtained for a real discrete spectrum are not directly obtainable.

ACKNOWLEDGMENTS

We wish to thank A. Grossmann, H. Baumgärtel, and W. C. Schieve for helpful discussions.

[1] A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decay in Nonrelativistic Quantum Mechanics* (Israel Program for Scientific Translations, Jerusalum, 1966) (Moscow, 1966).

[2] T. Berggren, Nucl. Phys. A **109**, 265 (1968).

[3] W. J. Romo, Nucl. Phys. A **116**, 617 (1968).

[4] R. C. Fuller, Phys. Rev. **188**, 1649 (1969).

[5] A. Grossmann, J. Math. Phys. **5**, 1025 (1964).

[6] J. Aguilar and J. M. Combes, Comm. Math. Phys. **22**, 269 (1971). E. Balslev and J. M. Combes, Comm. Math. Phys. **22**, 280 (1971). See also J. M. Combes, Proceedings of the International Congress of Mathematicians, Vancouver, 1974. These works will be referred to as ABC.

[7] J. S. Howland, J. Math. Anal. Appl. **50**, 415 (1975); W. Baumgärtel, Math. Nachr. **75**, 133 (1976), and references contained in these works.

[8] B. Simon, Ann. Math. **97**, 247 (1973).

[9] L. P. Horwitz and I. Sigal, Helv. Phys. Acta **51**, 685 (1979).

[10] E. C. G. Sudarshan, C. B. Chiu, and V. Gorini, Phys. Rev. D **18**, 2914 (1978); G. Parravicini, V. Gorini, and E. C. G. Sudarshan, J. Math. Phys. **21**, 2208 (1980); T. Bailey and W. C. Schieve, Nuovo Cimento A **47**, 231 (1978); W. C. Schieve and T. Bailey, in *Seventh International Colloquium on Group Theoretical Methods in Physics, Austin, Texas*, 1978 (Springer-Verlag, Berlin, 1979).

[11] A. Bohm, in *Quantum Mechanics* (Springer-Verlag, New York, 1979), Chap. XXI, and in *Seventh Colloquium on Group Theoretical Methods in Physics, Austin, Texas, 1978* (Spinger-Verlag, Berlin, 1979).

[12] E. Katznelson, J. Math. Phys. **21**, 1393 (1980). See also Ref. 10.

[13] A. Grossmann, in *Sixth International Colloquium on Group Theoretical Methods in Physics, Proceedings, Tübingen, 1977*, edited by P. Kramer and A. Rieckers (Springer-Verlag, Berlin, 1978), p. 162.

[14] C. Van Winter, Trans. Amer. Soc. **162**, 103 (1971).

[15] C. Van Winter, J. Math. Anal. Appl. **47**, 633 (1974).

[16] J.-P. Antoine and A. Grossmann, J. Funct. Anal. **23**, 369 (1976); 379 (1976).

[17] T. Kato, *Perturbation Theory for Linear Operators* (Springer-Verlag, Berlin, 1966), Chap. IV, Sec. 1.

[18] L. P. Horwitz and J. P. Marchand, Rocky M. J. Math. **1**, 225 (1971).

[19] See A. G. Ramm, J. Math. Phys. **21**, 2052 (1980), for a treatment of the Dirichlet problem for the Helmholtz equation, which, although it follows an approach quite different from ours, has some common features. For a discussion of variational techniques in Hilbert space, we refer the reader to F. J. Murray, *An Introduction to Linear Transformations in Hilbert Space* (Princeton U. P., Princeton, NJ, 1941).

The generalized anharmonic oscillator in three dimensions: Exact eigenvalues and eigenfunctions

Anita Rampal and K. Datta
Department of Physics and Astrophysics, University of Delhi, Delhi-110007, India

(Received 1 December 1981; accepted for publication 5 March 1982)

We study the generalized anharmonic oscillator in three dimensions described by the potentials of the form $\Sigma_{k=1}^{2m+1} b_k r^{2k}$. An asymptotic analysis of the Schrödinger equation yields the leading asymptotic behavior of the energy eigenfunctions in terms of the dominant $(m+1)$ coupling constants b_k, $m+1 \leqslant k \leqslant 2m+1$. Using an ansatz which incorporates this asymptotic behavior, we reduce the eigenvalue equation to an $(m+2)$-term difference equation. The corresponding Hill determinant may be made to factorize with a finite determinant as a factor if a set of constraints on the couplings is satisfied; an infinite sequence of such sets exists. The exact energy eigenvalues appear as the real roots of the finite factor of the Hill determinant; the corresponding wavefunctions are Gaussian weighted polynomials. We consider the potentials $\Sigma_1^3 b_k r^{2k}$ and $\Sigma_1^5 b_k r^{2k}$ explicitly; potentials of the form $\Sigma_1^{2m} b_j r^j$ and $\Sigma_1^{2m} b_j r^j + \delta/r$ containing both even and odd terms are also considered. Finally, we show that this method of constructing exact solutions fails for anharmonic potentials of the form $\Sigma_1^{2m} b_k r^{2k}$, of which the quartic anharmonic oscillator is the simplest example.

PACS numbers: 03.65.Ge

I. INTRODUCTION

The existence of exact energy eigenstates for certain anharmonic systems in one dimension has now been known for some time.[1-3] These states are characterized by exponentially weighted polynomial wavefunctions with eigenvalues given by analytic functions of the couplings; recently, a new class of such states with wavefunctions given by integral transforms has also been obtained.[4] Such eigenstates are not known for the simplest (i.e., quartic) anharmonic system; however, for the doubly anharmonic system, the freedom allowed by the appearance of two anharmonic couplings permits the construction of such states when certain constraints on the couplings are satisfied. It has therefore been conjectured[5] that it may be possible to construct such eigenstates for more complicated anharmonic systems described by potentials with several anharmonic terms. However, earlier treatments, based as they were on the continued fraction solution to contiguous three-term difference equations[1,3] or to particular solutions of second-order differential equations,[2,4,5] cannot easily be adapted to the general problem.

In this paper, we consider the construction of exact energy eigenstates of the generalized three-dimensional symmetric anharmonic oscillator, described by a potential of the form $\Sigma_1^{2m+1} b_k r^{2k}$. An asymptotic analysis of the wave equation shows that the controlling factor in the wavefunction for large r is determined by the $(m+1)$ coupling constants $b_{m+1},\ldots,b_{2m+1}$. Using an ansatz for the wavefunction with the correct asymptotic behavior, we reduce the wave equation to an $(m+2)$-term difference equation.[6] The corresponding Hill determinant whose zeroes give the energy eigenvalues is now almost triangular. This allows us to write a sequence of constraints on the couplings for which the Hill determinant factorizes into a finite determinant multiplying an infinite one. The exact energy eigenvalues appear as the real roots of the finite determinant whose elements are analytic functions of the couplings; for these energy values the difference equation terminates and the wavefunction, after the requirement of square integrability has been imposed by a suitable choice of an arbitrary constant, reduces to a Gaussian-weighted polynomial. The constraints appear in the form of m algebraic relations between the $(2m+1)$ couplings $b_1,\ldots,b_{2m+1}$; there is an infinite sequence of such relations. The constraints arrange themselves in a natural hierarchy; the degree of the algebraic equation to be solved to obtain the energy eigenvalues and the number of exact eigenstates increases as one progresses up the hierarchy. Thus, when the Nth set of constraints is satisfied, the eigenvalues appear as the real roots of an algebraic equation of degree N and the number of exact energy eigenstates obtained in this fashion is $S \leqslant N$. While an infinite number of such states can be found since the sequence of constraints is infinite, it must be emphasized that for each set of constraints being satisfied only a finite number of exact energy eigenstates can be found. The remaining infinity of eigenvalues must be found as the roots of the infinite determinant which is the remaining factor of the Hill determinant.

II. THE ASYMPTOTIC WAVEFUNCTION AND THE DIFFERENCE EQUATION

We write the radial Schrödinger equation for the reduced wavefunction as

$$\left[-\frac{d^2}{dr^2}+\sum_{k=1}^{2m+1} b_k r^{2k}+\frac{l(l+1)}{r^2}-E\right]\chi(r)=0 \tag{1}$$

in units $\hbar = 2m = 1$. $\chi(r)$ is the reduced radial wavefunction; the radial wavefunction $\psi(r)=(1/r)\chi(r)$. To extract the leading asymptotic behavior of $\chi(r)$, we note that, the equation being linear and of the second order, the controlling factor of the leading behavior may be expected to be exponential. Substituting

$$\chi(r)=e^{-f(r)}\phi(r), \tag{2}$$

0022-2488/83/040860-07$02.50

where $\phi(r)$ goes asymptotically as a power of r: ϕ'/ϕ, $\phi''/\phi \rightarrow 0$ as $r \rightarrow \infty$, we obtain from (1) the equation

$$(f')^2 - f'' - \frac{2\phi' f'}{\phi} + \frac{\phi''}{\phi} - \sum_{k=1}^{2m+1} b_k r^{2k} - \frac{l(l+1)}{r^2} + E = 0. \tag{3}$$

Asymptotically,

$$(f')^2 - f'' \sim \sum_{k=1}^{2m+1} b_k r^{2k}. \tag{4}$$

$f(r)$ thus may be expected to be a polynomial in r. Indeed, noting that $(f')^2$ must match the leading powers of the potential, we use the ansatz

$$f(r) = \sum_{i=1}^{m+1} \frac{d_i}{2i} r^{2i}. \tag{5}$$

Thus

$$(f')^2 = \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} d_i d_j r^{2(i+j-1)} \tag{6}$$

since $d_i = 0$, $i \geqslant m+2$, the double sum in (6) may be reordered in the form

$$(f')^2 = \sum_{k=1}^{2m+1} \left(\sum_{i=1}^{k} d_i d_{k+1-i} \right) r^{2k}. \tag{7}$$

Use of (7) and the expression for f'' in (4) results in the asymptotic relation

$$\sum_{k=1}^{2m+1} \left(\sum_{i=1}^{k} d_i d_{k+1-i} \right) r^{2k} - \sum_{i=1}^{m+1} (2i-1) d_i r^{2i-2} \sim \sum_{k=1}^{2m+1} b_k r^{2k}, \quad r \rightarrow \infty. \tag{8}$$

Equation (8) in turn implies the asymptotic relation

$$0 \sim \sum_{k=m+1}^{2m+1} \left[\left(\sum_{i=1}^{k} d_i d_{k+1-i} \right) - b_k \right] r^{2k}, \quad r \rightarrow \infty. \tag{9}$$

The $(m+1)$ coefficients d_i are thus uniquely determined in terms of the $(m+1)$ coupling constants $b_{m+1}, b_{m+2}, \ldots, b_{2m+1}$ through the $(m+1)$ algebraic relations

$$\sum_{i=1}^{k} d_i d_{k+1-i} = b_k, \quad m+1 \leqslant k \leqslant 2m+1 \tag{10}$$

subject to the constraint $d_i = 0$, $i \geqslant m+2$. Thus, when m is even, we have the relations

$$d_{m+1}^2 = b_{2m+1},$$

$$2d_m d_{m+1} = b_{2m},$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$2d_1 d_{m+1} + 2d_2 d_m + \cdots + d_{m/2+1}^2 = b_{m+1}.$$

The controlling factor in the asymptotic behavior of the wavefunction having thus been determined in terms of the $(m+1)$ couplings $b_{m+1}, b_{m+2}, \ldots, b_{2m+1}$, we choose for our wavefunction the Frobenius type ansatz

$$\chi(r) = e^{-f(r)} \sum_{n=0}^{\infty} a_n r^{n+p} \tag{11}$$

with $f(r)$ given by (5) and (10). Equation (1) now reduces to

$$\begin{aligned} \sum_{n=0}^{\infty} a_n r^{n+p} \Bigg[& \sum_{k=1}^{m} \left(\sum_{i=1}^{k} d_i d_{k+1-i} \right) r^{2k} \\ & - \sum_{k=0}^{m} (2k+1) d_{k+1} r^{2k} \\ & - 2(n+p) \sum_{k=0}^{m} d_{k+1} r^{2k} + \frac{(n+p)(n+p-1)}{r^2} \\ & + E - \frac{l(l+1)}{r^2} - \sum_{k=1}^{m} b_k r^{2k} \Bigg] = 0, \end{aligned} \tag{12}$$

where we have used (10) to remove the $(m+1)$ leading powers of the potential $b_k r^{2k}$, $m+1 \leqslant k \leqslant 2m+1$. We therefore reduce the differential equation to the $(m+2)$-term recurrence relation

$$a_{n+2} + A_{n,n} a_n + \sum_{k=1}^{m} A_{n,n-2k} a_{n-2k} = 0, \tag{13}$$

with

$$A_{n,n} \equiv \frac{E - d_1(2n+2p+1)}{(n+p+2)(n+p+1) - l(l+1)}, \tag{14}$$

$$A_{n,n-2k} \equiv \frac{(\Sigma_{i=1}^{k} d_i d_{k+1-i}) - (2n+2p-2k+1) d_{k+1} - b_k}{(n+p+2)(n+p+1) - l(l+1)}$$

$$(1 \leqslant k \leqslant m).$$

The indicial equation yields $p = (l+1)$ for a wavefunction regular at the origin. Thus

$$A_{n,n} = \frac{E - d_1(2n+2l+3)}{(n+2)(n+2l+3)}, \tag{15}$$

$$A_{n,n-2k} = \frac{(\Sigma_{i=1}^{k} d_i d_{k+1-i}) - (2n+2l-2k+3) d_{k+1} - b_k}{(n+2)(n+2l+3)} \quad (1 \leqslant k \leqslant m). \tag{16}$$

The eigenvalue parameter E appears only in the diagonal coefficient $A_{n,n}$ whereas the off-diagonal coefficients depend only on the couplings. We note that the recurrence relation (13) generates two sequences, one for the even coefficients and the other for the odd ones, in terms of arbitrary constants a_0 and a_1, respectively.

III. THE HILL DETERMINANT AND EXACT SOLUTIONS

The difference equation

$$a_{n+2} + A_{n,n} a_n + A_{n,n-2} a_{n-2} + \cdots + A_{n,n-2m} a_{n-2m} = 0$$

will have a nontrivial solution if the so-called Hill determinant for the problem vanishes:

$$D = \begin{vmatrix} A_{0,0} & 0 & 1 & 0 & 0 & \cdots & & & \\ 0 & A_{1,1} & 0 & 1 & 0 & 0 & \cdots & & \\ A_{2,0} & 0 & A_{2,2} & 0 & 1 & 0 & 0 & \cdots & \\ 0 & A_{3,1} & 0 & A_{3,3} & 0 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & & & & \ddots & & & \end{vmatrix} = 0. \tag{17}$$

The eigenvalues are roots of the Hill determinant $D(E)$. The recurrence relation (13) does not connect the even and odd coefficients; the infinite determinant can therefore be written as

$$D = \Delta\cdot\tilde{\Delta}, \tag{18}$$

where

$$\Delta \equiv \begin{vmatrix} A_{0,0} & 1 & 0 & 0 & \cdots & & \\ A_{2,0} & A_{2,2} & 1 & 0 & 0 & \cdots & \\ A_{4,0} & A_{4,2} & A_{4,4} & 1 & 0 & 0 & \cdots \\ \vdots & & & & \ddots & & \end{vmatrix}$$

and

$$\tilde{\Delta} \equiv \begin{vmatrix} A_{1,1} & 1 & 0 & 0 & \cdots & & \\ A_{3,1} & A_{3,3} & 1 & 0 & 0 & \cdots & \\ A_{5,1} & A_{5,3} & A_{5,5} & 1 & 0 & 0 & \cdots \\ \vdots & & & & \ddots & & \end{vmatrix} \tag{19}$$

so that an eigenvalue is a root of either Δ (E) or $\tilde{\Delta}$ (E) or a (possible) common root of both.

Neither Δ nor $\tilde{\Delta}$ is triangular; however, the matrix defined by omitting the first column of Δ or $\tilde{\Delta}$ is triangular with unit determinant, i.e.,

$$I \equiv \begin{vmatrix} 1 & 0 & 0 & \cdots & \\ A_{2,2} & 1 & 0 & 0 & \cdots \\ A_{4,2} & A_{4,4} & 1 & 0 & \cdots \\ \vdots & & \ddots & & \end{vmatrix} = 1 \tag{20}$$

and similarly

$$\tilde{I} \equiv \begin{vmatrix} 1 & 0 & 0 & \cdots & \\ A_{3,3} & 1 & 0 & \cdots & \\ A_{5,3} & A_{5,5} & 1 & 0 & \cdots \\ \vdots & & \ddots & & \end{vmatrix} = 1. \tag{21}$$

We solve the recursion (13) in terms of the lowest coefficient $a_0\,(a_1)$ and the truncations of Δ $(\tilde{\Delta}$ $)$ and I $(\tilde{I}$ $)$. Let Δ_j be the jth truncation of the infinite determinant Δ (i.e., the determinant consisting of the first j rows and columns). Then

$$\Delta_j \equiv \begin{vmatrix} A_{0,0} & 1 & 0 & \cdots & \\ A_{2,0} & A_{2,2} & 1 & 0 & \cdots \\ \vdots & & \ddots & & \\ A_{2j-2,0} & & \cdots & & A_{2j-2,2j-2} \end{vmatrix}. \tag{22}$$

Similarly, if I_j $(\tilde{I}_j)$ define truncations of I $(\tilde{I}$ $)$, then all the I_j $(\tilde{I}_j)$ equal unity:

$$I_j = \tilde{I}_j = 1 \quad \text{for } j = 1,2,\ldots . \tag{23}$$

The solution to the sequence of even coefficients generated from (13) may now be written as

$$a_{2j} = (-1)^j (\Delta_j / I_j) a_0 = (-1)^j \Delta_j a_0. \tag{24}$$

Substituting (24) in (13), we find that the truncated determinants Δ_j satisfy the $(m+2)$-term recurrence relation

$$\Delta_{j+1} = A_{2j,2j}\Delta_j + \sum_{k=1}^{m} (-1)^k A_{2j,2(j-k)}\Delta_{j-k} \quad (\Delta_0 = 1). \tag{25}$$

Similarly for the odd coefficients we obtain

$$a_{2j+1} = (-1)^j \tilde{\Delta}_j a_1, \tag{26}$$

where $\tilde{\Delta}_j$, the jth truncation of $\tilde{\Delta}$, satisfies the recurrence relation

$$\tilde{\Delta}_{j+1} = A_{2j+1,2j+1}\tilde{\Delta}_j + \sum_{k=1}^{m} (-1)^k A_{2j+1,2(j-k)+1}\tilde{\Delta}_{j-k} \quad (\tilde{\Delta}_0 = 1). \tag{27}$$

Consider, first, recursion (25) for the sequence of the truncations of Δ, the Hill determinant for the even coefficients. We find that Δ_N (where $N \geqslant 1$, is a fixed value of the variable j) becomes a factor of Δ_{N+1} provided

$$\sum_{k=1}^{m} (-1)^k A_{2N,2(N-k)}\Delta_{N-k} = 0. \tag{28}$$

For Δ_N to be a factor of both Δ_{N+1} and Δ_{N+2} a second condition has to be simultaneously satisfied, i.e., in addition to (28) we must also have

$$\sum_{k=2}^{m} (-1)^k A_{2(N+1),2(N+1-k)}\Delta_{N+1-k} = 0, \tag{29}$$

which follows from (25) written in the form

$$\Delta_{N+2} = A_{2(N+1),2(N+1)}\Delta_{N+1} - A_{2(N+1),2N}\Delta_N + \sum_{k=2}^{m} (-1)^k A_{2(N+1),2(N+1-k)}\Delta_{N+1-k}.$$

In general, for Δ_N to be a factor of $\Delta_{N+1}, \Delta_{N+2}, \cdots, \Delta_{N+m}$ we must require a set of m conditions to be simultaneously satisfied:

$$\sum_{k=1}^{m-S} (-1)^{k+S} A_{2(N+S),2(N-k)} \times \Delta_{N-k} = 0 \quad \text{for } S = 0,1,2,\ldots,(m-1). \tag{30}$$

These conditions ensure that Δ_N is a factor of Δ_{N+1}, ..., Δ_{N+m}; by virtue of recursion (25), Δ_N then becomes a factor of all subsequent truncations of Δ and hence factorizes out of Δ, the Hill determinant itself.

As we have noted earlier, the energy eigenvalue E appears only in the diagonal coefficient $A_{i,i}$ defined in (15); $\Delta_N(E)$ is therefore a polynomial of Nth degree in E. We note that the conditions (30) involve the off-diagonal coefficients $A_{i,j}$ $(i \neq j)$, defined in (16) and hence are constraints on the couplings b_k $(1 \leqslant k \leqslant 2m+1)$.

Thus when the couplings satisfy the constraints (30), Eq. (18) reduces to

$$D = \tilde{\Delta}\cdot\Delta_N\cdot \begin{vmatrix} A_{N,N} & 1 & 0 & \cdots & \\ A_{N+2,N} & A_{N+2,N+2} & 1 & 0 & \cdots \\ \vdots & & & \ddots & \end{vmatrix}. \tag{31}$$

The real roots (in E) of the equation $\Delta_N(E) = 0$ are thus exact energy eigenvalues of the problem subject to the couplings satisfying constraints (30); for these values of E equations (24), (25), and (30) ensure that all the $a_{2(N+k)}$, $k \geqslant 0$, vanish and the wavefunction reduces to

$$\chi(r) = e^{-f(r)} r^{l+1}\left(\sum_{n=0}^{N-1} a_{2n} r^{2n} + \sum_{n=0}^{\infty} a_{2n+1} r^{2n+1}\right). \tag{32}$$

We next show that the choice $a_1 = 0$ is necessary to render $\chi(r)$ a physically acceptable solution of (1). We can see that the odd series $\Sigma^\infty a_{2n+1} r^{2n+1}$ does not have the correct asymptotic behavior as $r \rightarrow \infty$; indeed, if written in the form

$a_1 r \Sigma_{n=0}^{\infty} g_n r^{2n}$, the sum $\Sigma^{\infty} g_n r^{2n}$ grows as $e^{2f(r)}$, as $r \to \infty$; consequently $\chi(r) \sim e^{f(r)}$, $r \to \infty$, and is not square integrable. To prove this, we note that

$$e^{2f(r)} = \sum_{n=0}^{\infty} C_n r^{2n}, \tag{33}$$

where $f(r)$ is given by (5) and the C_n's satisfy the $(m+2)$-term recurrence relation

$$nC_n - \sum_{i=1}^{m+1} d_i C_{n-i} = 0 \tag{34}$$

with the d_i given by (10).

From (13) we obtain the recursion satisfied by the g_n's:

$$g_n + A_{2n-1,2n-1}\, g_{n-1} + \sum_{k=1}^{m} A_{2n-1,2n-1-2k}\, g_{n-k-1} = 0. \tag{35}$$

For large n (35) takes the form

$$g_n \sim \frac{1}{n} \sum_{k=0}^{m} d_{k+1}\, g_{n-k-1}, \quad n \to \infty, \tag{36}$$

which becomes identical to (34). Since the asymptotic form of (35) determines the behavior of $\Sigma^{\infty} g_n r^{2n}$ for large r, the series $\Sigma^{\infty} g_n r^{2n}$ grows asymptotically as $e^{2f(r)}$.

We have therefore arrived at the following result: Whenever the couplings b_k $(1 \leqslant k \leqslant 2m+1)$ are such that they satisfy the constraints (30) for a specific value of N, say $N = N_1$, then p $(0 \leqslant p \leqslant N_1)$ exact eigenvalues [corresponding to the real roots of $\Delta_{N_1}(E) = 0$] can be found. Corresponding to each eigenvalue an exact exponentially weighted polynomial solution of the form

$$\chi(r) = e^{-f(r)} r^{l+1} \sum_{n=0}^{N_1-1} a_{2n} r^{2n} \tag{37}$$

exists, where the coefficients a_n satisfy a finite recursion relation (13) subject to (30). The remaining infinity of eigenvalues must be obtained as the roots of the two infinite determinants in (31).

A similar sequence of exact solutions and corresponding constraints follow from the Hill determinant for the odd coefficients. When the couplings are such that they satisfy the constraints

$$\sum_{k=1}^{m-S} (-1)^{k+S} A_{2(N+S)+1,2(N-k)+1} \times \widetilde{\Delta}_{N-k} = 0, \quad S = 0,1,\ldots,(m-1), \tag{38}$$

for a specific value of N, say $N = N_2$, then p $(0 \leqslant p \leqslant N_2)$ exact eigenvalues can be found as the real roots of $\widetilde{\Delta}_{N_2}(E) = 0$. The odd series terminates as all $a_{2(N_2+k)+1}$ vanish for $k \geqslant 0$; we choose $a_0 = 0$ to obtain physically acceptable wavefunctions. Corresponding to each eigenvalue, the exact Gaussian-weighted polynomial solution then is

$$\chi(r) = e^{-f(r)} r^{l+1} \sum_{n=0}^{N_2-1} a_{2n+1} r^{2n+1}, \tag{39}$$

where the odd coefficients satisfy the recursion (13), rendered finite by the constraints (38).

The exact energy eigenvalues and eigenfunctions of the generalized anharmonic oscillator in 1 dimension described by the potential $V(x) = \Sigma_{k=1}^{2m+1} b_k x^{2k}$, $-\infty < x < \infty$, may be obtained by identical methods. Since $x = 0$ is an ordinary point of the corresponding Schrödinger equation, one replaces Eqs. (13) and (14) with their counterparts in which $p = l = 0$.

IV. THE POTENTIALS $V(r) = \Sigma_{k=1}^{3} b_k r^{2k}$ AND $V(r) = \Sigma_{k=1}^{5} b_k r^{2k}$

To illustrate the power of the general method developed, we examine a few of the simpler anharmonic potentials which have been studied in the literature by different methods applicable to special cases.

Consider the potential $V(r) = b_1 r^2 + b_2 r^4 + b_3 r^6$. The choice of the ansatz

$$\chi(r) = e^{-f(r)} r^{l+1} \sum_{n=0}^{\infty} a_n r^n \tag{40}$$

with

$$f(r) = \tfrac{1}{2} d_1 r^2 + \tfrac{1}{4} d_2 r^4, \tag{41}$$

where

$$d_1 = b_2/2b_3^{1/2}, \quad d_2 = b_3^{1/2}, \tag{42}$$

corresponding to Eqs. (5) and (10), yields a three-term recurrence relation for the a_n's:

$$a_{n+2} + A_{n,n} a_n + A_{n,n-2} a_{n-2} = 0, \tag{43}$$

where

$$A_{n,n} \equiv [E - d_1(2n + 2l + 3)]/(n+2)(n + 2l + 3) \tag{44}$$

$$A_{n,n-2} \equiv [\gamma - (2n + 2l + 1)] d_2/(n+2)(n + 2l + 3) \tag{45}$$

and

$$\gamma \equiv (d_1^2 - b_1)/d_2. \tag{46}$$

From (30) it follows that the couplings have to satisfy only one constraint to ensure that Δ_N $(N = 1,2,\ldots)$ emerges as a factor in Δ. Thus Eq. (30) reduces to

$$A_{2N,2N-2} = 0 \quad (\text{when } \Delta_{N-1} \neq 0), \tag{47}$$

which is ensured by the couplings b_1, b_2, b_3, satisfying the equation

$$\gamma = 4N + 2l + 1. \tag{48}$$

The exact eigenvalues are the real roots of the $N \times N$ determinantal equation:

$$\Delta_N(E) = \begin{vmatrix} \dfrac{E - d_1(2l+3)}{2(2l+3)} & 1 & 0 & \cdots & & \\ \dfrac{(N-1)d_2}{(2l+5)} & \dfrac{E - d_1(2l+7)}{4(2l+5)} & 1 & \cdots & & \\ \vdots & \ddots & & & & \\ 0 & \cdots & 0 & & \dfrac{4d_2}{2N(2N+2l+1)} & \dfrac{E - d_1(4N+2l-1)}{2N(2N+2l+1)} \end{vmatrix} = 0. \tag{49}$$

The eigenfunction corresponding to each eigenvalue is given by

$$\chi(r) = e^{-f(r)} r^{l+1} \sum_{n=0}^{N-1} a_{2n} r^{2n}. \tag{50}$$

Here the a_n $(0 \leqslant n \leqslant N-1)$ satisfy the recursion (43) rendered finite by the constraint (48).

We note that the solutions for $l = 0$ are those obtained earlier for the $1-d$ doubly anharmonic oscillator corresponding to odd parity.[3] For low N the explicit solutions are easily obtained algebraically. For $N = 2$, the constraint on the couplings is given by $\gamma = 2l + 9$. The roots of $\Delta_2(E)$ are

$$E_{\pm} = d_1(2l+5) \pm 2[d_1^2 + 2d_2(2l+3)]^{1/2}. \tag{51}$$

The corresponding exact eigenfunctions are of the form

$$\chi_{\pm}(r) = a_0 e^{-f(r)} r^{l+1}[1 + (a_2^{\pm}/a_0) r^2], \tag{52}$$

where, from (43), (44), and (49), we have

$$\begin{aligned} \frac{a_2^{\pm}}{a_0} &= \frac{[d_1(2l+3) - E_{\pm}]}{2(2l+3)} \\ &= \frac{4d_2}{[E_{\pm} - d_1(2l+7)]}. \end{aligned} \tag{53}$$

A similar set of eigenvalues and eigenfunctions are generated by the odd sequence of coefficients in (43). Whenever the couplings satisfy the constraint

$$A_{2N+1,2N-1} = 0 \quad (\text{with } \tilde{\Delta}_{N-1} \neq 0),$$

i.e.,

$$\gamma = 4N + 2l + 3, \tag{54}$$

the real roots of $\tilde{\Delta}_N(E) = 0$ provide the eigenvalues; corresponding to each eigenvalue, the exact solution is

$$\chi(r) = e^{-f(r)} r^{l+1} \sum_{n=0}^{N-1} a_{2n+1} r^{2n+1}, \tag{55}$$

where a_n's satisfy the recursion (43) rendered finite by(54). For $N = 2$, the constraint on the couplings is given by $\gamma = 2l + 11$; the eigenvalues and eigenfunctions are

$$E_{\pm} = d_1(2l+7) \pm [d_1^2 + 6d_2(l+2)]^{1/2}, \tag{56}$$

$$\chi_{\pm}(r) = a_1 e^{-f(r)} r^{l+2}[1 + (a_3^{\pm}/a_1) r^2], \tag{57}$$

where

$$\begin{aligned} \frac{a_3^{\pm}}{a_1} &= \frac{[d_1(2l+5) - E_{\pm}]}{3(2l+4)} \\ &= \frac{4d_2}{[E_{\pm} - d_1(2l+9)]}. \end{aligned} \tag{58}$$

We next examine the potential

$$V(r) = \sum_{k=1}^{5} b_k r^{2k}. \tag{59}$$

Equations (5) and (10) now show that the function $f(r)$ is now a polynomial of third degree in r^2; the coefficients d_1, d_2, and d_3 in the ansatz are given in terms of the three leading couplings b_3, b_4, and b_5 by the relations

$$\begin{aligned} d_1 &= (1/2b_5^{1/2})(b_3 - b_4^2/4b_5), \\ d_2 &= b_4/2b_5^{1/2}, \\ d_3 &= b_5^{1/2}. \end{aligned} \tag{60}$$

The relation for the coefficients a_n is now a four-term recursion. Δ_N $(N = 1,2,...)$ becomes a factor of the Hill determinant Δ, whenever the couplings satisfy the following two constraints:

$$A_{2N,2N-2}\Delta_{N-1} - A_{2N,2N-4}\Delta_{N-2} = 0$$

and (61)

$$A_{2N+2,2N-2} = 0 \quad (\text{with } \Delta_{N-1} \neq 0).$$

Similarly for $\tilde{\Delta}_N$ $(N = 1,2,...)$ to emerge as a factor in $\tilde{\Delta}$, the couplings b_k $(1 \leqslant k \leqslant 5)$ must satisfy the following constraints:

$$A_{2N+1,2N-1}\tilde{\Delta}_{N-1} - A_{2N+1,2N+3}\tilde{\Delta}_{N-2} = 0$$

and (62)

$$A_{2N+3,2N-1} = 0 \quad (\text{with } \tilde{\Delta}_{N-1} \neq 0).$$

For $N = 1$, the relations (61) are

$$b_1 = (1/4b_5)(b_3 - b_4^2/4b_5)^2 - (2l+5)b_4/2b_5^{1/2}$$

and (63)

$$b_2 = (b_4/2b_5)(b_3 - b_4^2/4b_5) - (2l+7)b_5^{1/2};$$

when these constraints are satisfied, Δ_1 emerges as a factor in Δ. The eigenvalue is given by

$$E = d_1(2l+3) \tag{64}$$

and the corresponding eigenfunction is

$$\chi(r) = e^{-f(r)} r^{l+1} \quad (a_0 \text{ has been chosen to be unity}). \tag{65}$$

This solution, and (52), are the ones obtained by Flessas and Das.[5]

Similarly, whenever the couplings satisfy the constraints

$$b_1 = (1/4b_5)(b_3 - b_4^2/4b_5)^2 - (2l+7)b_4/2b_5^{1/2}$$

and (66)

$$b_2 = (b_4/2b_5)(b_3 - b_4^2/4b_5) - (2l+9)b_5^{1/2}.$$

$\tilde{\Delta}_1$ is a factor of the Hill determinant $\tilde{\Delta}$, and one obtains the exact solution given by

$$\chi(r) = e^{-f(r)} r^{l+2} \quad (a_1 \text{ has been chosen to be unity}) \tag{67}$$

with the corresponding eigenvalue

$$E = d_1(2l+5). \tag{68}$$

V. THE POTENTIALS $V(r) = \Sigma_{j=1}^{2m} b_j r^j$ AND $\Sigma_{j=1}^{2m} b_j r^j + \delta/r$

Recently, there has been some interest in anharmonic systems and confinement potentials with both odd and even terms; examples of such potentials with some exact solutions may be found in the literature.[7–9] Here we merely note that the methods developed in Sec. III apply to these cases also, though they have to be applied with some caution. The leading asymptotic behavior of the wavefunction is determined uniquely in terms of the leading couplings as before; however, the eigenvalue parameter now appears in the off-diagonal elements $A_{i,j}$ $(i \neq j)$ of the Hill determinant. Thus the m conditions of constraint which ensure that Δ_N $(N = 1,2,...)$ is a factor in the Hill determinant Δ now involve E; further, Δ_N is a polynomial of order p $(0 \leqslant p \leqslant N)$ in E. The conditions of constraint and the eigenvalue equation for exact solutions, $\Delta_N(E) = 0$, have now to be used in conjunction to obtain the exact energy eigenvalues and the corresponding constraints on the couplings. Thus for the potential

$$V(r) = b_1 r + b_2 r^2 + b_3 r^3 + b_4 r^4 \tag{69}$$

the coefficients determining the controlling factor $e^{-f(r)}$ of the leading asymptotic behavior of the wavefunction as given in (2) are

$$\sum_{i=1}^{k+1} d_i d_{k+2-i} = b_k, \quad 2 \leqslant k \leqslant 4, \tag{70}$$

where now

$$f(r) = \sum_{i=1}^{3} \frac{d_i r^i}{i}. \tag{71}$$

This yields

$$\begin{aligned} d_1 &= \frac{1}{2b_4{}^{1/2}}(b_2 - b_3^2/4b_4), \\ d_2 &= b_3/2b_4{}^{1/2}, \\ d_3 &= b_4{}^{1/2}. \end{aligned} \tag{72}$$

The four-term recursion relation satisfied by the coefficients a_n is

$$a_{n+1} + A_{n,n} a_n + A_{n,n-1} a_{n-1} + A_{n,n-2} a_{n-2} = 0, \tag{73}$$

where

$$\begin{aligned} A_{n,n} &\equiv -2(n+l+1)d_1/(n+1)(n+2l+2), \\ A_{n,n-1} &\equiv \frac{[E + d_1^2 - (2n+2l+1)d_2]}{(n+1)(n+2l+2)}, \\ A_{n,n-2} &\equiv \frac{[2d_1 d_2 - b_1 - 2(n+l)d_3]}{(n+1)(n+2l+2)}. \end{aligned} \tag{74}$$

The two constraints ensuring that Δ_1 is a factor in Δ are

$$E + d_1^2 - (2l+3)d_2 = 0$$

and

$$b_1 - 2d_1 d_2 + (2l+4)d_3 = 0 \tag{75}$$

with

$$\Delta_1 = -d_1. \tag{76}$$

Used in conjunction with $\Delta_1 = 0$, they yield the two constraints on the couplings,

$$b_2 = b_3^2/4b_4, \quad b_1 = -(2l+4)b_4{}^{1/2}, \tag{77}$$

and the energy eigenvalue

$$E = (2l+3)d_2, \tag{78}$$

with the wavefunction

$$\chi(r) = r^{l+1} \exp(-\tfrac{1}{2}d_2 r^2 - \tfrac{1}{3}d_3 r^3). \tag{79}$$

This is the solution written down by Khare.[7]

For Δ_2 to be a factor in Δ (with $\Delta_1 \neq 0$) the two conditions of constraint are

$$d_1[E + d_1^2 - (2l+5)d_2] = b_1 + (2l+4)d_3 - 2d_1 d_2$$

and (80)

$$b_1 = 2d_1 d_2 - (2l+6)d_3,$$

with

$$\Delta_2 = [(2l+3)(d_1^2 + d_2) - E]/2(2l+3); \tag{81}$$

used in conjunction with $\Delta_2 = 0$, Eqs. (80) yield the constraints on the couplings,

$$\begin{aligned} &b_1 + (2l+4)b_4{}^{1/2}[1 - (1/8b_4^2)(b_2 - b_3^2/4b_4)^3] = 0, \\ &(2l+6)b_4{}^{1/2} + b_1 - (b_3/2b_4)(b_2 - b_3^2/4b_4) = 0, \end{aligned} \tag{82}$$

and the energy eigenvalue

$$E = (2l+3)(d_1^2 + d_2), \tag{83}$$

with the wavefunction

$$\chi(r) = r^{l+1} \exp(-d_1 r - \tfrac{1}{2}d_2 r^2 - \tfrac{1}{3}d_3 r^3)(1 + d_1 r). \tag{84}$$

Confinement potentials of the form $\Sigma_{j=1}^{2m} b_j r^j + \delta/r$ present no new features: a sequence of exact eigenvalues and eigenfunctions may be constructed in an identical fashion.

VI. THE POTENTIALS $V(r) = \Sigma_{k=1}^{2m} b_k r^{2k}$

Finally, we note that the circumstance that the simplest anharmonic system, viz., the quartic oscillator, does not admit of such exact solutions is a particular instance of a wider phenomenon. The Schrödinger equation for even anharmonic potentials in which the highest power of the relevant coordinate variable is $4m$ ($m = 1,2,\ldots$) may, with a suitable ansatz, be reduced to a $(2m+2)$ recurrence relation:

$$\begin{aligned} a_{n+2} &+ \sum_{k=0}^{m} A_{n,n-2k+1} a_{n-2k+1} + A_{n,n} a_n \\ &+ \sum_{k=1}^{m-1} A_{n,n-2k} a_{n-2k} = 0, \end{aligned} \tag{85}$$

where

$$\begin{aligned} A_{n,n} &\equiv (E + d_1^2)/(n+2)(n+2l+3), \\ A_{n,n-2k+1} &\equiv \frac{-2(n+l+2-k)d_{k+1}}{(n+2)(n+2l+3)}, \\ A_{n,n-2k} &\equiv \frac{\left[\left(\sum_{i=1}^{k+1} d_i d_{k+2-i}\right) - b_k\right]}{(n+2)(n+2l+3)}. \end{aligned} \tag{86}$$

Here the d_i are given, as before, by the leading anharmonic couplings through the relations

$$\sum_{i=1}^{k+1} d_i d_{k+2-i} = b_k, \quad m \leqslant k \leqslant 2m, \tag{87}$$

subject to $d_i = 0$, $i \geqslant m+2$.

The structure of the coefficients (86) shows that it is now no longer possible to find a sequence of constraints which are consistent and ensure that the Hill determinant factorizes into a finite determinant times an infinite one.

VII. CONCLUSION

We have shown how an infinite sequence of exact energy eigenvalues and eigenfunctions for generalized three-dimensional anharmonic oscillators described by potentials of the form $\Sigma_{k=1}^{2m+1} b_k r^{2k}$ may be obtained. A set of constraints on the couplings ensures that the Hill determinant for the difference equation to which the Schrödinger equation may be reduced factorizes with a finite determinant as a factor. The exact energy eigenvalues are the real roots of the finite determinant; the corresponding eigenfunctions are Gaussian-weighted polynomials. For each set of constraints only a finite number of such energy eigenvalues may be obtained; however, an infinite sequence of such sets exists. Analogous results hold for the one-dimensional system and for potentials of the form $\Sigma^{2m} b_j r^j$ and $\Sigma^{2m} b_j r^j + \delta/r$ with odd and even terms; however, in the latter cases there arise a coupled set of constraints involving the energy eigenvalue and the couplings which have subsequently to be decoupled to obtain the energy eigenvalues. The method does not, however, yield a similar solution for potentials of the form $\Sigma^{2m} b_k r^{2k}$ of which the simplest case is the quartic anharmonic oscillator.

Note added in manuscript: After the submission of this note for publication, our attention has been drawn to the work of E. Magyari [Phys. Lett. A **81**, 116 (1981)], in which similar solutions have been obtained for the one-dimensional problem; however, in contrast to the results here obtained, he obtains only coupled equations for the energy eigenvalues and the constraints on the couplings.

ACKNOWLEDGMENT

One of us (A.R.) acknowledges financial assistance from the Department of Atomic Energy, Government of India.

---

[1] V. Singh, S. N. Biswas, and K. Datta, Phys. Rev. D **18**, 1901 (1978).

[2] G. P. Flessas, Phys. Lett. A **72**, 289 (1979).

[3] V. Singh, Anita Rampal, S. N. Biswas, and K. Datta, Lett. Math. Phys. **4**, 131 (1980).

[4] G. P. Flessas, Phys. Lett. A **81**, 17 (1981) and J. Phys. A **14**, L209 (1981); see also Anita Rampal and K. Datta, Univ. of Delhi, preprint (1981).

[5] G. P. Flessas and K. P. Das, Phys. Lett. A **78**, 19 (1980).

[6] An alternative approach, used earlier in the study of the doubly anharmonic oscillator, employs the recursion relation for the coefficients of the complete asymptotic series for the wavefunction. In the present instance, the use of the asymptotic series yields no additional information. See K. Datta and Anita Rampal, Phys. Rev. D **23**, 2875 (1981).

[7] A. Khare, Phys. Lett. A **83**, 237 (1981).

[8] V. Singh, S. N. Biswas, and K. Datta, Lett. Math. Phys. **3**, 73 (1979).

[9] J. Killingbeck, J. Phys. A **13**, L393 (1980).

The accidental degeneracy problem in nonrelativistic quantum mechanics

Avinash Khare

Department of Theoretical Physics, The University, Manchester M13 9PL, United Kingdom and Institute of Physics, Bhubaneswar-751005, India[a]

(Received 8 October 1981; accepted for publication 12 February 1982)

I show that the class of potentials given by $V(r) = Ar^{2d-2} - Br^{d-2}$ possesses partial accidental degeneracy given by $E_{n_1,l_2} = E_{n_2,l_1}$ in case $d = (l_2 - l_1)/(n_2 - n_1)$, $B = [l + dn + (1-d)/2] \times (2A/\mu)^{1/2}$. It is further shown that, for a given potential, as the number of dimensions change, the accidental degeneracy pattern also changes except when $d = 1$ and 2. Using these results, it is then shown that for the bottom quark-antiquark ($b\bar{b}$) bound system, most likely $E_{3S} < E_{1F} < E_{2D}$. Finally I also make some conjectures about the ordering of levels for a wide class of potentials.

PACS numbers: 03.65.Ge, 03.65.Fd

I. INTRODUCTION

It is well known that the Coulomb and the oscillator potentials are the only examples possessing accidental degeneracy in nonrelativistic quantum mechanics.[1] In particular, whereas the energy eigenvalues of the Coulomb potential

$$V(r) = -\alpha/r \tag{1}$$

satisfy

$$E_{n+1,l} = E_{n,l+1} = \cdots, \tag{2}$$

those of the harmonic oscillator potential

$$V(r) = \tfrac{1}{2}Kr^2 \tag{3}$$

satisfy

$$E_{n+1,l} = E_{n,l+2} = \cdots. \tag{4}$$

The remarkable thing about Eqs. (2) and (4) is that they are valid for any value of n and l. Here $E_{n,l}$ denotes the energy eigenvalue corresponding to a state of angular momentum l ($l = 0,1,2,\cdots$) and the number of nodes of the reduced radial wavefunction including at $r = 0$ being n ($n = 1,2,3,\cdots$).

Even thought it is known that there are no other potentials possessing such accidental degeneracy, there are a number of questions which, to my best knowledge, have remained unanswered in the literature. Some of these are:

(i) Are there potentials which possess at least a partial accidental degeneracy? For example, we know that for the Coulomb potential $E_{2S} = E_{1P}$ while for the oscillator potential $E_{2S} = E_{1D}$. It is then natural to inquire if there exists a potential for which $E_{2S} = E_{1F}$ or $E_{2S} = E_{1G}$ or in general $E_{2S} = E_{1l}$? Clearly for large l this would (if at all) only be possible for a potential which is highly singular as $r \to \infty$.

(ii) A somewhat related question is if there exists a potential for which $E_{1P} = E_{3S}$ or E_{4S} or in general E_{nS}? Obviously such potentials (if they at all exist) have to be more singular then $-1/r$ and less singular than $-1/r^2$ as $r \to \infty$.

(iii) For the Coulomb potential we also know that

$$E_{n,S} = E_{n-1,P} = \cdots = E_{1,l=n-1}, \tag{5}$$

while for the oscillator potential we have

$$E_{n,S} = E_{n-1,D} = \cdots = E_{1,l=2n-2}, \tag{6a}$$

$$E_{n,P} = E_{n-1,F} = \cdots = E_{1,l=2n-1}. \tag{6b}$$

Thus it is natural to inquire if one can generalize these statements and find a potential which at least for a given n_1, n_2, l_1, and l_2 satisfy

$$\begin{aligned}\cdots = E_{2n_1-n_2,2l_2-l_1} &= E_{n_1,l_2} = E_{n_2,l_1}\\ &= E_{2n_2-n_1,2l_1-l_2\cdots}.\end{aligned} \tag{7}$$

The number of accidentally degenerate levels are of course restricted since $n \geqslant 1$ and $l \geqslant 0$.

The purpose of this paper is to provide answers to the questions raised above. In particular, I prove the following theorem[2]:

Theorem 1: The class of potentials given by

$$V(r) = Ar^{2d-2} - Br^{d-2}, \quad A, B > 0, \tag{8}$$

where d is any positive rational number, exhibits partial accidental degeneracy as given by Eq. (7) provided that

$$d = (l_2 - l_1)/(n_2 - n_1), \tag{9}$$

$$B = (2A/\mu)^{1/2}\left[l + dn + \frac{(1-d)}{2}\right]. \tag{10}$$

Here μ is the reduced mass of the system. It must be emphasized here that the above theorem is not merely of academic interest. Some possible applications are:

(a) Suppose in a certain spectrum one observes that $E_{n_1,l_2} = E_{n_2,l_1}$, and further let us assume that the dynamics of the system can be understood in terms of nonrelativistic quantum mechanics. Clearly, the knowledge of the potential possessing such an accidental degeneracy would be of utmost importance in understanding the dynamics of the system under consideration.

(b) Even if E_{n_1,l_2} is close to E_{n_2,l_1}, though not exactly equal, the knowledge of the potential for which they are exactly equal could be quite useful as the perturbation theory around it is likely to be pretty accurate.

(c) In fact, even if $E_{n_1,l_2} > (<) E_{n_2,l_1}$ in a given spectrum the knowledge of the potential for which they are equal would provide considerable restrictions on the form of the potential. For example, in the charmonium spectrum, one experimentally observed that

$$E_{1D} > E_{2S} > E_{1P}. \tag{11}$$

Since for the Coulomb potential we know that $E_{1D} = E_{3S} > E_{2S} = E_{1P}$ while for the oscillator potential $E_{1D} = E_{2S} >$

---

[a] Present address.

0022-2488/83/040867-07$02.50

E_{1P}, one concludes that the charmonium potential should be something in between the two, i.e., the confining part of the potential is at best like r^2 as $r \rightarrow \infty$.[3] As a further illustration of these ideas, I propose to answer the question if for the bottomonium system (made out of b quark and antiquark)

$$E_{1F} \gtrless E_{3S}. \tag{12}$$

This question is quite relevant because if it turns out that $E_{1F} < E_{3S}$, then one could hopefully produce and detect the $1\,^3F_3$ level via the transitions

$$\Upsilon'' \rightarrow 1\,^3F_2 + \gamma \rightarrow \Upsilon' + 2\gamma. \tag{13}$$

(d) Finally the knowledge of the potential having $E_{n_1,l_2} = E_{n_2,l_1}$ could help *à la* Grosse and Martin[3] in deriving theorems about the ordering of levels for a wide class of potentials.

The plane of the paper is as follows: In Sec. II, I solve the Schrödinger equation for the potential (8) and obtain conditions under which it possesses partial accidental degeneracy. The expression for total degeneracy is also given here. In Sec. III, I consider potential (8) in p space dimensions and show that if p changes, the accidental degeneracy pattern also changes except when $d = 1$ or 2. In Sec. IV, I consider the potential in classical particle mechanics and show that closed particle trajectories exist for $E = 0$. I also show here that for $E = 0$ the Bohr–Sommerfeld quantization condition reproduces the accidental degeneracy pattern obtained in Sec. II. Finaly, in Sec. V, I consider some possible applications of the results obtained in this paper and make several conjectures about the ordering of levels for a wide class of potentials.

II. DEGENERACY IN THREE SPACE DIMENSIONS

The solution of the Schrödinger equation for the potential (8) can be written as

$$\psi(r, \theta, \phi) = y_{lm}(\theta,\phi) R_l(r)/r, \tag{14}$$

where $R_l(r)$ is a solution of the equation ($\hbar = 1$)

$$R_l''(r) + [2E\mu + 2B\mu r^{d-2} - 2A\mu r^{2d-2} - l(l+1)/r^2] R_l(r) = 0. \tag{15}$$

On substituting

$$R_l(r) = r^{l+1} f(r) \exp\{[-(2A\mu)^{1/2}/d]r^d\}, \tag{16}$$

it can be easily shown that if $E = 0$, then f satisfies the confluent hypergeometric equation

$$xf''(x) + \left(\frac{d+2l+1}{d} - x\right) f'(x) + \left[\frac{B}{d}\left(\frac{\mu}{2A}\right)^{1/2} - \frac{2l+d+1}{2d}\right] f(x) = 0, \tag{17}$$

where $x = 2(2A\mu)^{1/2} r^d/d$. Thus, if the bound state condition

$$\frac{B}{d}\left(\frac{\mu}{2A}\right)^{1/2} - \frac{2l+d+1}{2d} = n - 1, \quad n = 1,2,3,\cdots, \tag{18}$$

is satisfied, then the eignvalues and eigenfunctions of the Schrödinger equation for the potential (8) are given by

$$E_{n,l} = 0, \tag{19a}$$

$$\psi(r,\theta,\phi) = N_{nl} r^l y_{lm}(\theta,\phi) \exp\{[-(2A\mu)^{1/2}/d]r^d\} \times {}_1F_1(-n+1,(d+2l+1)/d;(2/d)(2A\mu)^{1/2} r^d). \tag{19b}$$

Let us now turn to the question of accidental degeneracy. To that purpose, let us write $d = d_1/d_2$, where d_1 and d_2 are both integers which can take any value 1, 2, 3,$\cdots$, and, to avoid duplication, let us demand that, for a given d, d_1 and d_2 do not have any common factor. In that case condition (18) [which is identical to Eq. (10)] can also be written as

$$B = \frac{1}{d_2}\left(\frac{2A}{\mu}\right)^{1/2}\left(N + \frac{d_1+d_2}{2}\right), \tag{20}$$

where

$$N = (n-1)d_1 + ld_2 \tag{21}$$

is an integer. Hence for a given N one has the accidental degeneracy as given by Eq. (7) provided that

$$(n_1 - 1)d_1 + l_2 d_2 = (n_2 - 1)d_1 + l_1 d_2, \tag{22}$$

which is equivalent to Eq. (9). Thus we have shown that, as long as Eqs. (9) and (10) are satisfied for a given N, the potential given by Eq. (8) possesses accidental degeneracy as given by Eq. (7). Since $n \geqslant 1$ and $l \geqslant 0$, the number of degenerate levels are obviously limited.

It must be emphasized here that the relation (20) between the coupling constants A and B depends on N. Further even for a given value of d, different values of N imply *different* potentials. Hence, the class of potentials as given by Eq. (8) possesses only *partial* accidental degeneracy. The only two exceptions being when $d = 1$ and $d = 2$.

(*i*) $d = 1$: In this case the potential (8) reduces to the Coulomb potential, i.e.,

$$V(r) = A - B/r. \tag{23}$$

Thus A is essentially $-E$ so that the bound state condition (20) gives us the energy eigenvalues of the Coulomb potential. In other words, different values of N do not imply different potentials but just different eigenvalues of the same Coulomb potential, and hence one has full accidental degeneracy for any value of N as given by Eq. (2).

(*ii*) $d = 2$: In this case the potential (8) reduces to the oscillator potential, i.e.,

$$V(r) = Ar^2 - B \tag{24}$$

so that B is essentially E, and hence, as above, different values of N just correspond to different energy eigenvalues of the same oscillator potential. Needless to say that, for $d = 1$ and 2, the eigenfunctions as given by Eq. (19b) reduce to those of the Coulomb and the oscillator potentials, respectively.

To understand further the partial accidental degeneracy question for the class of potentials (8), let us concentrate on Eqs. (20) and (21). From these we note that if $N < d_1 d_2$, then there is no accidental degeneracy. In fact, for a given d_1 and d_2, and $N = 1, 2, 3, \cdots, \min(d_1 - 1, d_2 - 1)$, etc. are not even allowed as Eq. (21) cannot be satisfied in those cases. It is not difficult to convince oneself that for a given d_1 and d_2 the number of N values not allowed are

$$\tfrac{1}{2}(d_1 - 1)(d_2 - 1). \tag{25}$$

$N = 0$ is, however, always allowed, and it corresponds to $n = 1, l = 0$, i.e., $1S$ state, and is, as expected, nondegenerate. Thus, for the class of potentials (8), $E_{1S} = 0$ provided that

$$B = \tfrac{1}{2}(2A/\mu)^{1/2}(1+d). \qquad (26)$$

The accidental degeneracy starts occurring when $N \geqslant d_1 d_2$. In particular, if

$$d_1 d_2 \leqslant N \leqslant 2d_1 d_2 - 1, \qquad (27)$$

then two levels are accidentally degenerate except for those $\frac{1}{2}(d_1 - 1)(d_2 - 1)$ values of N given by

$$N = N_{\text{forbidden}} + d_1 d_2, \qquad (28)$$

for which there is no accidental degeneracy. Generalizing, it is not difficult to convince oneself that if

$$(m-1)d_1 d_2 \leqslant N \leqslant m d_1 d_2 - 1, \qquad (29)$$

then m levels will be accidentally degenerate. The only exceptions are those $\frac{1}{2}(d_1 - 1)(d_2 - 1)$ values of N given by

$$N = N_{\text{forbidden}} + (m-1)d_1 d_2, \qquad (30)$$

for which only $(m - 1)$ levels are accidentally degenerate. Since the number of accidentally degenerate levels depend only on the product $d_1 d_2$, it is clear that for a given N the accidental degeneracy is same for d and d^{-1}.

Total degeneracy: Since a level with angular momentum l is $(2l + 1)$-fold degenerate, it may be worthwhile to calculate the total degeneracy for a given value of N and see as to how it varies with d_1, d_2, and N. Clearly, if

$$N = 0, d_1, 2d_1, \ldots, (d_2 - 1)d_1, \qquad (31)$$

then from Eq. (21) it follows that $l = 0$ so that the total degeneracy $D = 1$. On the other hand, if

$$N = d_2, d_2 + d_1, \ldots, d_2 + md_1, md_1 < (d_1 - 1)d_2, \qquad (32)$$

then from Eq. (21) it is clear that $l = 1$ and hence $D = 3$. Of course, for the $\frac{1}{2}(d_1 - 1)(d_2 - 1)$ forbidden values of N, $D = 0$.

Now, as N increases by $d_1 d_2$ units from those given by Eq. (31), it is clear that both $l = 1$ and $l = d_1$ are allowed and hence $D = 2d_1 + 2$, while, if N increases by $d_1 d_2$ units from those given by Eq. (32), then $l = 1$ and $d_1 + 1$ so that $D = 2d_1 + 6$. Thus it is not very difficult to convince oneself that in general if

$$\text{(i)} \quad N = md_2 + kd_1, \quad md_2 + kd_1 + d_1 d_2, \cdots \qquad (33)$$

$(m, k = 0, 1, 2, \cdots)$, then

$$D = (1/d_1 d_2^2)(N + d_1 d_2 - kd_1 - md_2) \times (N + md_2 - kd_1 + d_2). \qquad (34)$$

Note that here either $md_2 + kd_1 < d_1 d_2$ or

$$md_2 + kd_1 = N_{\text{forbidden}} + d_1 d_2. \qquad (35)$$

$$\text{(ii)} \quad N = N_{\text{forbidden}}, \quad D = 0. \qquad (36)$$

In the special case of the Coulomb potential, $d = 1$ so that the total degeneracy formula given by Eq. (34) takes the well-known form

$$D = (N+1)^2, \quad N = 0, 1, 2, \cdots. \qquad (37)$$

On the other hand, for the oscillator potential, $d = 2$ ($d_1 = 2$, $d_2 = 1$), and Eqs. (33) and (34) give, as expected,

$$D = \tfrac{1}{2}(N+1)\times(N+2), \quad N = 0, 1, 2, \cdots. \qquad (38)$$

At this stage it may be worthwhile to discuss a few specific examples which would also help in answering some of the questions raised in the Introduction. From Eqs. (9) and (10) it is clear that

$$E_{2S} = E_{1F} = 0 \qquad (39)$$

would be true for the potential

$$V(r) = Ar^4 - 5(2A/\mu)^{1/2} r. \qquad (40)$$

The corresponding eigenfunctions are given by Eq. (19b) with $d = 3$. Clearly the total degeneracy D is 8. On the other hand,

$$E_{2P} = E_{1G} = 0 \qquad (41)$$

is true for the potential

$$V(r) = Ar^4 - 6(2A/\mu)^{1/2} r, \qquad (42)$$

and hence $D = 12$ while

$$E_{1,l=N} = E_{2,N-3} = \cdots E_{n+1,N-3n} = \cdots = 0 \qquad (43)$$

would be valid for the potential

$$V(r) = Ar^4 - (N+2)(2A/\mu)^{1/2} r. \qquad (44)$$

On the other hand, $E_{2S} = E_{1,l}$ would be valid for the potential (8) with $d = l$, i.e.,

$$V(r) = Ar^{2l-2} - \tfrac{1}{2}(3l+1)(2A/\mu)^{1/2} r^{l-2}, \qquad (45)$$

and in this case $D = 2(l + 1)$, which also follows from the formula (34) by using $d_1 = l$, $d_2 = 1$, $N = l$, and $m = k = 0$.

Proceeding in the same way, it follows that

$$E_{1P} = E_{3S} = 0 \qquad (46)$$

would be valid for the potential

$$V(r) = \frac{A}{r} - \tfrac{7}{4}(2A/\mu)^{1/2} r^{-3/2}, \qquad (47)$$

while

$$E_{1P} = E_{n+1,S} = 0 \qquad (48)$$

would be true for the potential (8) with $d = 1/n$, i.e.,

$$V(r) = \frac{A}{r^{2-2/n}} - \frac{(3n+1)}{2n}\left(\frac{2A}{\mu}\right)^{1/2} \frac{1}{r^{2-1/n}}. \qquad (49)$$

Finally let us inquire about the degeneracy structure for a somewhat nontrivial case say $d = 8/5$. From Eqs. (8)–(10) it follows that the bound state spectrum with $E = 0$ is possible provided

$$B = \tfrac{1}{5}(2A/\mu)^{1/2}(N + \tfrac{13}{2}), \qquad (50)$$

where

$$N = 8(n-1) + 5l. \qquad (51)$$

From the formula (25) it is clear that there are 14 missing values of N, and they are given by

$$N_{\text{missing}} = 1, 2, 3, 4, 6, 7, 9, 11, 12, 14, 17, 19, 22, 27. \qquad (52)$$

The accidental degeneracy can occur only if $N \geqslant 40$. For example, if

$$V(r) = Ar^{1.2} - \tfrac{93}{10}(2A/\mu)^{1/2} r^{-0.4}, \qquad (53)$$

then

$$E_{6,S} = E_{1,l=8} = 0. \qquad (54)$$

TABLE I. Variation of accidental degeneracy pattern with d_1 and d_2.

d_1 d_2	1	2	3	$n+1$
1 $N=l+1$	$E_{1,l+1}=E_{2,l}$	$E_{1,l+1}=E_{2,l-1}$	$E_{1,l+1}=E_{2,l-2}$	$E_{1,l+1}=E_{2,l-n}$
2 $N=2(l+1)$	$E_{1,l+1}=E_{3,l}$		$E_{1,l+1}=E_{3,l-2}$	$E_{1,l+1}=E_{3,l-n}$
3 $N=3(l+1)$	$E_{1,l+1}=E_{4,l}$	$E_{1,l+1}=E_{4,l-1}$		$E_{1,l+1}=E_{4,l-n}$
n $N=n(l+1)$	$E_{1,l+1}=E_{n+1,l}$	$E_{1,l+1}=E_{n+1,l-1}$	$E_{1,l+1}=E_{n+1,l-2}$	$E_{1,l+1}=E_{n+1,l-n}$

The variation of accidental degeneracy with d_1 and d_2 is given in Table I [of course, B and A have to be appropriately related in each case as given by Eq. (10)].

III. VARIATION OF DEGENERACY WITH DIMENSIONS

In the last section we have shown that the potential (8) possesses partial accidental degeneracy in three space dimensions provided Eqs. (9) and (10) are satisfied. It may, therefore, be worthwhile to inquire if the accidental degeneracy continues to be there in p space dimensions or not, and, if yes, then how does the accidental degeneracy pattern change with p for a given potential. In this context it may be noted that in the exceptional cases of the Coulomb and the oscillator potentials the accidental degeneracy pattern *does not* change with p, i.e,. Eqs. (2) and (4) are valid in p space dimensions; only the magnitude of energy eigenvalues changes with p.

If I write the solution of the Schrödinger equation in p-space dimensions for the potential (8) as

$$\psi(r,\theta,\phi_1,\phi_2,\dots,\phi_{p-2}) = y_{l(\theta,\phi_1,\phi_2,\dots,\phi_{p-2})}^{m_1,m_2,\dots,m_{p-2}} R_l(r)/r, \qquad (55)$$

then it is not difficult to show that $R_l(r)$ satisfies Eq. (15) with l being replaced everywhere by a where $a = l + (p-3)/2$. Thus the whole discussion of Sec. II up to Eq. (22) goes through with l being replaced everywhere by a. In other words, the potential (8) possess partial accidental degeneracy as given by Eq. (7) provided Eq. (9) is satisfied, and, in addition, B and A are related by

$$B = \frac{1}{d_2}\left[N + \frac{pd_2}{2} + \frac{d_1 - 2d_2}{2}\right]\left(\frac{2A}{\mu}\right)^{1/2}, \qquad (56)$$

with N being given as before by Eq. (21). Since the relation (56) between coupling constants A and B depends on p, it is clear that for a given potential the accidental degeneracy pattern is going to be different in different dimensions. In other words, what potential will exhibit a given degeneracy pattern, e.g., $E_{n_1,l_2} = E_{n_2,l_1}$ will also depend on the value of p. However, since the relation (9) is p-independent, it is clear that the class of potentials exhibiting $E_{n_1,l_2} = E_{n_2,l_1}$ will, nevertheless, be restricted to having same d.

Let me now discuss a few specific examples. Consider, for example, the potential (42), i.e.,

$$V(r) = Ar^4 - 6(2A/\mu)^{1/2}r, \qquad (42)$$

which possesses the following accidental degeneracy in three space dimensions

$$E_{1G} = E_{2P} = 0, \quad p = 3. \qquad (41)$$

Let us now see as to how the degeneracy pattern changes with p. Since the relation (9) is still valid in p dimensions, it is clear that if there are degenerate levels, then they will satisfy

$$(l_2 - l_1) = 3(n_2 - n_1). \qquad (57)$$

Using Eqs. (8) and (42) it is clear that in this case $B = 6$ $(2A/\mu)^{1/2}$, which when substituted in Eq. (56) gives (note $d_1 = 3$, $d_2 = 1$)

$$6(n-1) + 2l + p = 11. \qquad (58)$$

Clearly, the relation (58) cannot be satisfied if p is even. In fact, even when p is odd, it can only be satisfied if $p \leqslant 11$ as $n \geqslant 1$, $l \geqslant 0$. Further, in addition to $p = 3$, the accidental degeneracy can only occur in five space dimensions for the potential (42) in which case it exhibits

$$E_{1F} = E_{2S} = 0, \ p = 5. \qquad (59)$$

Let us now look at this problem from another angle, i.e., instead of concentrating on the same potential (42), let us concentrate on the degeneracy $E_{1G} = E_{2P}$ and inquire as to what class of potentials will exhibit it in various space dimensions. Clearly, this class must have $d = 3$ and $N = (n-1)d_1 + ld_2 = 4$ so that Eq. (56) will simplify to

$$B = [4 + (p+1)/2](2A/\mu)^{1/2}. \qquad (60)$$

Hence the class of potentials given by

$$V(r) = Ar^4 - [4 + (p+1)/2](2A/\mu)^{1/2}r \qquad (61)$$

will exhibit the degeneracy $E_{1G} = E_{2P} = 0$ in p space dimensions.

From the structure of Eq. (56) it is clear that, for integer d, it can only be satisfied either for even or for odd p. In fact, even for fractional d the same is still true as long as d_2 is an odd integer. On the other hand, if d_2 is even, then it is possible to satisfy Eq. (56) in both even and odd space dimensions.

From Eq. (56) it is also clear that even in p dimensions, the potential (8) will possess at least one exact solution provided $N \geqslant 0$, and hence

$$B \geqslant \tfrac{1}{2}(2A/\mu)^{1/2}(d + p - 2), \qquad (62)$$

while accidental degeneracy can only occur in p dimensions if $N \geqslant d_1 d_2$, which implies

$$B \geqslant \tfrac{1}{2}(2A/\mu)^{1/2}(2d_1 + d + p - 2). \qquad (63)$$

In the special case of one space dimension, $l = 0$, and obviously there is no degeneracy, and we merely obtain a class of exact solutions for the potential (8) given by

$$E_n = 0, \tag{64a}$$

$$\psi_n(x) = N_n \exp\left[-(2A\mu)^{1/2}x^d/d \right] \times {}_1F_1(-n+1,(d-1)/d;(2/d)(2A\mu)^{1/2}x^d), \tag{64b}$$

provided that

$$B = (2A/\mu)^{1/2}[nd - (1+d)/2], \quad n = 1, 2, \cdots. \tag{65}$$

In order to ensure that $\psi_n(x) \to 0$ as $x \to \pm\infty$, one should perhaps restrict oneself to only even vaues of d. In that case, one has only an even class of solutions, i.e., with an even number of nodes. It is, of course, quite straightforward to obtain the odd class of solutions, i.e., one can show that the exact odd solutions for the potential (8) are

$$E_n = 0, \tag{66a}$$

$$\psi_n(x) = N_n x \exp\{[-(2A\mu)^{1/2}/d]x^d\} \times {}_1F_1(-n+1,(d+1)/d;(2/d)(2A\mu)^{1/2}x^d), \tag{66b}$$

provided that

$$B = (2A/\mu)^{1/2}[nd - (d-1)/2], \quad n = 1, 2, \cdots. \tag{67}$$

IV. THE ORIGIN OF THE PARTIAL ACCIDENTAL DEGENERACY

In the last section we have shown that the potential (8) exhibits partial accidental degeneracy in p dimensions in case Eqs. (9) and (56) are satisfied. Now it is well known that the accidental degeneracy for the Coulomb and the oscillator potentials is in a sense related to the fact that, in classical particle mechanics, closed particle trajectories exist in these fields.[4] It may therefore be worthwhile to inquire if, at least for $E = 0$, the potentials given by Eq. (8) have closed particle trajectories.

The equation for the path of the particle in the central potential (8) is given by[4]

$$(2\mu)^{1/2}\phi = \int \frac{L\,dr/r^2}{[E - Ar^{2d-2} + Br^{d-2} - L^2/2\mu r^2]^{1/2}} + \text{const}, \tag{68}$$

where L is the angular momentum of the particle. In order to show that for $E = 0$ the path is closed, one has to demonstrate that

$$\Delta\phi = \frac{2L}{(2\mu A)^{1/2}} \int_{r_{\min}}^{r_{\max}} \frac{dr/r^2}{[(B/A)r^d - r^{2d} - L^2/2\mu A]^{1/2}} \tag{69}$$

is a rational function of 2π, i.e., $\Delta\phi = 2\pi n_1/n_2$, where n_1 and n_2 are integers. On substituting $r^d = t$ in Eq. (69), it is not very difficult to integrate Eq. (69) and show that

$$\Delta\phi = (2/d)[\sin^{-1}(1) - \sin^{-1}(1)], \tag{70}$$

i.e., $\Delta\phi$ is indeed a rational function of 2π and hence for $E = 0$ closed trajectories exist for the class of potentials (8).

In the case of the Coulomb and the oscillator potentials, it is also well known that the *exact* eigenvalue spectrum can be obtained from the Bohr–Sommerfeld quantization condition [with, of course, the usual replacement of $l(l+1)$ by $(l+1/2)^2$ and that the higher order WKB corrections are all zero.[5] It may therefore be worthwhile to inquire if, at least for $E = 0$, the Bohr–Sommerfeld-quantization condition reproduces the bound state constraint (10) and further if the higher-order WKB corrections are indeed zero or not.

In the case of the potential (8) the Bohr–Sommerfeld quantization condition leads us to

$$\int_{r_{\min}}^{r_{\max}} [E - Ar^{2d-2} + Br^{d-2} - (l+\tfrac{1}{2})^2/2\mu r^2]^{1/2} dr = (n - \tfrac{1}{2})h/2(2\mu)^{1/2}. \tag{71}$$

For $E = 0$ (and $\hbar = 1$) this reduces to

$$\int_{r_{\min}}^{r_{\max}} \left[\frac{B}{A}r^d - r^{2d} - \frac{(l+\frac{1}{2})^2}{2\mu A} \right]^{1/2} \frac{dr}{r} = (n - \tfrac{1}{2}) \frac{\pi}{(2\mu A)^{1/2}}. \tag{72}$$

On using $r^d = t$, it is not difficult to integrate the 1hs of Eq. (72) and obtain

$$\frac{B\pi}{2Ad} - \frac{(l+\frac{1}{2})\pi}{(2\mu A)^{1/2}d} = \frac{(n-\frac{1}{2})\pi}{(2\mu A)^{1/2}}, \tag{73}$$

which is equivalent to the bound state condition (10). Thus we have shown that for the class of potentials (8) the Bohr–Sommerfeld quantization condition reproduces the bound state constraint (10) in case $E_{n,l} = 0$. Following Ref. 5, one can also calculate the higher-order WKB corrections and show that they are all zero.

In the case of the Coulomb and the oscillator potentials we also know that $\Delta r_{1l}/\langle r \rangle_{1l} \to 0$ as $l \to \infty$, i.e., in the semiclassical limit the particle is practically localized in the vicinity of a sphere. Let us see if it is also true for the class of potentials (8). To that purpose, we have to first calculate the normalization constant N_{1l} appearing in Eq. (19b). On using the formula[6] ($\operatorname{Re}\nu > 0$; n is an integer)

$$\int_0^\infty e^{-kz} z^{\nu-1} [{}_1F_1(-n,\gamma,kz)]^2\, dz = \frac{\Gamma(\nu) n!}{k^\nu \gamma(\gamma+1)\cdots(\gamma+n-1)} \times \left[1 + \sum_{s=0}^{n-1} \frac{n(n-1)\cdots(n-s)}{[(s+1)!]^2} \times \frac{(\gamma-\nu-s-1)(\gamma-\nu-s)\cdots(\gamma-\nu+s)}{\gamma(\gamma+1)\cdots(\gamma+s)} \right], \tag{74}$$

it easily follows that

$$N_{1l}^2 = \frac{2(2A\mu)^{1/2}}{\Gamma((2l+3)/d)} \left(\frac{2(2A\mu)^{1/2}}{d} \right)^{(2l+3)/d-1}. \tag{75}$$

Using formula (74) and Eq. (19b), one can easily calculate the expectation value of r and r^2

$$\langle r \rangle_{1l} = \frac{\Gamma((2l+4)/d)}{\Gamma((2l+3)/d)} \left(\frac{d}{2(2A\mu)^{1/2}} \right)^{1/d}, \tag{76a}$$

$$\langle r^2 \rangle_{1l} = \frac{\Gamma((2l+5)/d)}{\Gamma((2l+3)/d)} \left(\frac{d}{2(2A\mu)^{1/2}} \right)^{1/d}. \tag{76b}$$

Hence it easily follows that

$$\frac{\Delta r_{1l}}{\langle r \rangle_{1l}} = \left[\frac{\Gamma((2l+3)/d)\Gamma((2l+5)/d))}{\Gamma^2((2l+4)/d))} - 1 \right]^{1/2} \tag{77}$$

so that, as $l \rightarrow \infty$, $\Delta r^d_{1l} \langle r \rangle_{1l} \rightarrow 0$, i.e., the particle is practically localized in the vicinity of a sphere. Thus from various angles we see that the class of potentials (8) exhibits similar features to those of the Coulomb and the oscillator potentials.

What is the origin of the partial accidental degeneracy exhibited by the class of potentials (8)? For example, in the case of the Coulomb potential we know that, in addition to $\mathbf{L}^2$ and $\mathbf{L}_z$, there is another object, called the Runge–Lenz vector, which commutes with the Hamiltonian for this potential.[7] Similarly in the case of the oscillator potential the Hamiltonian is invariant under SU(3) group,[7] which is wider than the three-dimensional rotation group O(3). It is therefore natural to inquire about the extra symmetry possessed by the Hamiltonian for the class of potentials (8). Unfortunately, so far we have not succeeded in our endeavor. A somewhat related question is to inquire if, at least for $E = 0$, the Schrödinger equation for the class of potentials (8) can be solved in any other coordinate system. Note that the Schrödinger equation for both the Coulomb and the oscillator potentials can be solved in more than one coordinate system. Again, we have not been successful in answering this question.

V. APPLICATIONS

In this section we shall discuss some of the applications of the result that the class of potentials (8) exhibit accidental degeneracy as given by Eq. (7) in case Eqs. (9) and (10) are satisfied.

1F level of bottomonium: The first application which we have in mind is to the bottom quark–antiquark ($b\bar{b}$) bound system which, at least to zeroeth approximation, can be understood in terms of nonrelativistic quantum mechanics along with v^2/c^2 corrections.[8] The exact form of the $b\bar{b}$ potential is not known. However, from the asymptotic freedom argument, we expect that at short distances the potential should behave like $-\alpha_s/r$ (with logarithmic corrections) while at long distances the quark confinement plus flavor independence of the potential and the fact that $E^{c\bar{c}}_{ID} > E^{c\bar{c}}_{2S}$ indicates that the confining part of the potential $V_c(r)$ could at most behave like[3] r^2 as $r \rightarrow \infty$. Finally, the fact that $\Gamma(J/\psi \rightarrow e^+e^-) > \Gamma(\psi' \rightarrow e^+e^-)$ indicates that most likely $V_c(r)$ is concave in nature,[9] i.e., $V_c(r) = \int_0^1 \rho(\alpha) r^\alpha d\alpha, \rho(\alpha) \geqslant 0$. The question which we would like to raise here is if $E_{1F} >$ or $< E_{3S}$ for the $b\bar{b}$ system. As has been noted in the Introduction, if $E_{1F} < E_{3S}$, then at least the $1\,{}^3F_2$ state could be produced and detected via the transitions

$$\Upsilon''(3\,{}^3S_1) \rightarrow 1\,{}^3F_2 + \gamma \rightarrow \Upsilon'(2\,{}^3S_1) + 2\gamma.$$

To begin with let us note that the charmonium data and the flavor independence of the potential tell us that $E^{b\bar{b}}_{1F} > E^{b\bar{b}}_{1D} > E^{b\bar{b}}_{2S}$. Since for the Coulomb potential $E_{1F} = E_{4S}$ while for the oscillator potential $E_{2D} = E_{3S} > E_{1F} = E_{2P} > E_{2S}$, it is not clear from here if $E_{1F} >$ or $< E_{3S}$ in the case of the $b\bar{b}$ system. However, from Eqs. (7)–(9) we notice that

$$E_{1F} = E_{3S} \tag{78}$$

is true for the potential

$$V(r) = Ar - \tfrac{17}{4}(2A/\mu)^{1/2} r^{-1/2}. \tag{79}$$

Since for the Coulomb potential $E_{1F} = E_{4S}$ while for the oscillator potential $E_{1F} < E_{3S}$ and further since the $b\bar{b}$ potential is most likely of the form (79) but with the $-r^{-0.5}$ term being replaced by $-[r \log(r/r_0)]^{-1}$ and Ar replaced by $\int_0^1 \rho(\alpha) r^\alpha \, d\alpha, \rho(\alpha) \geqslant 0$, it appears that for the $b\bar{b}$ system

$$E_{2D} > E_{1F} > E_{3S}. \tag{80}$$

Since $\Upsilon'''(4\,{}^3S_1)$ is above Zweig threshold, it will predominantly decay to $b\bar{b}$ mesons. Of course, since $E(1\,{}^3F_2) < F_{1F}$ while $E(3\,{}^3S_1) > E(3S)$, it is quite possible that $E(3\,{}^3S_1) > E(1\,{}^3F_2)$. However, the splitting would be quite small so that there is no realistic chance of detecting the $1\,{}^3F_2$ level of the $b\bar{b}$ system.

Ordering of levels: Finally I wish to make few conjectures about the ordering of levels for a wide class of potentials.

Conjecture 1: For the class of potentials

$$V(r) = \int_{-1}^{2} \rho(\alpha) r^\alpha \epsilon(\alpha) \, d\alpha, \quad \rho(\alpha) \geqslant 0, \tag{81}$$

I speculate that

$$E_{n+1,l} \leqslant E_{n,l+2}. \tag{82}$$

The following arguments provide some support to this conjecture: (a) For the special case of $n = 1$ and $l = 0$ this has already been rigorously proved[3]; (b) for $d = 2$, i.e., for the oscillator potential, Eq. (82) is known to be an equality; (c) for $d = 1$, i.e., the Coulomb potential, the inequality (82) is indeed satisfied.

Conjecture 2: For the class of potentials

$$V(r) = \int_{2}^{\infty} \rho(\alpha) r^\alpha \, d\alpha, \quad \rho(\alpha) \geqslant 0, \tag{83}$$

I propose the opposite inequality

$$E_{n+1,l} \geqslant E_{n,l+2}. \tag{84}$$

Support to this conjecture comes from the fact that: (a) For the special case of $n = 1$, $l = 0$, it has been rigorously proved[3]; (b) for $d = 2$, Eq. (84) is known to be an equality; (c) for $d = 3, 4, 5, \cdots$, etc., it can be explicitly seen from Eqs. (7)–(10) that the inequality is indeed satisfied; (d) for some specific anharmonic oscillator models the inequality (84) with $n = 1$ has been explicitly proved.[10]

Conjecture 3: For the class of potentials given by

$$V(r) = \int_{4}^{\infty} \rho(\alpha) r^\alpha \, d\alpha, \quad \rho(\alpha) \geqslant 0, \tag{85}$$

I conjecture that

$$E_{n+1,l} \geqslant E_{n,l+3}. \tag{86}$$

Support for this conjecture comes from the fact that: (a) From Eqs. (7)–(10) it is known that equality $E_{n+1,l} = E_{n,l+3}$ is true for the class of potentials given by $V(r) = Ar^4 - Br$ (i.e., $d = 3$); (b) for $d = 4, 5, \cdots$, etc., Eqs. (7)–(10) offer numerous examples which indeed satisfy $E_{n+1,l} > E_{n,l+3}$.

Conjecture 4: Generalizing, for the class of potentials given by

$$V(r) = \int_{2m-2}^{\infty} \rho(\alpha) r^\alpha \, d\alpha, \quad \rho(\alpha) \geqslant 0, \ m > 1, \tag{87}$$

I conjecture that

$$E_{n+1,l} > E_{n,l+m}. \qquad (88)$$

In support of this let us note that Eq. (88) is an equality in the case of the potential $V(r) = Ar^{2m-2} - Br^{m-2}$ and that Eqs. (7)–(10) offer explicit examples in which, for $d = m + 1$, $m + 2, \cdots$ etc., one has indeed $E_{n+1,l} > E_{n,l+m}$. Generalizing, I also conjecture that, for the class of potentials

$$V(r) = \int_{2(l_2 - l_1)/(n_2 - n_1) - 2}^{\infty} \rho(\alpha) r^{\alpha}\, d\alpha,$$

$$\rho(\alpha) > 0, \ l_2 - l_1 > n_2 - n_1, \qquad (89)$$

one has the inequality

$$E_{n_2,l_1} > E_{n_1,l_2}. \qquad (90)$$

Can one say something about the form of the potentials satisfying $E_{n_2,l_1} < E_{n_1,l_2}$? Apart from Conjecture 1, I have no definite answer to this question. Is it that if for given n_1, n_2, and l_1, and l_2, that equality $E_{n_2,l_1} = E_{n_1,l_2}$ is satisfied for say $d = d_1$, then $E_{n_2,l_1} < E_{n_1,l_2}$ will be satisfied by all these potentials with $d < d_1$?

ACKNOWLEDGMENT

Most of the work was done while I was at the Manchester University. It is a pleasure to thank Professor A. Donnachie for the warm and coordial hospitality that I have enjoyed over there.

[1] See, for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), 3rd ed.

[2] A short account of this work has appeared in A. Khare, Lett. Math. Phys. **5**, 539 (1981).

[3] A. Martin, Phys. Lett. B **67**, 330 (1977); H. Grosse, Phys. Lett. B **68**, 343 (1977); H. Grosse and A.Martin, Nucl. Phys. **B 132**, 125 (1978).

[4] L. D. Landau and E. M. Lifshitz, *Mechanics* (Pergamon, New York, (1959).

[5] J. B. Krieger and C. Rosenzweig, Phys. Rev. **164**, 171 (1967).

[6] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*; *Nonrelativistic Theory* (Pergamon, New York, 1976), 3rd ed., p. 663.

[7] L. I. Schiff, Ref. 1, and L. D. Landau and E. M. Lifshitz, Ref. 6.

[8] For a recent review see H. Grosse and A. Martin, Phys. Rep. **60**, 341 (1980); V. A. Novikov *et al.*, Phys. Rep. **41**, 1 (1978).

[9] A. Martin, Phys. Lett. B **70**, 192 (1977).

[10] A. Khare, Phys. Lett. A **83**, 237 (1981).

Remarks on canonical transformations in phase-space path integrals

Christopher C. Gerry
Division of Science and Mathematics, University of Minnesota at Morris, Morris, Minnesota 56267

(Received 28 June 1982; accepted for publication 12 November 1982)

We study canonical transformations in phase-space path integrals in the Schrödinger representation. Using the example of a contact transformation implemented in each short-time propagator via the midpoint method, we show that the "measure" $(dp_N/2\pi\hbar)\,\Pi_{j=1}^{N}\,[dp_j dq_j/(2\pi\hbar)]$, apart from the unpaired dp_N, cannot be considered as a product of Liouville measures.

PACS numbers: 03.65.Fd

I. INTRODUCTION

It is well known that path integration in phase-space is encumbered with a number of ambiguities. These are principally concerned with operator ordering[1] and the implementation of canonical transformations. It is the latter problem which is to be discussed in this paper.

It has recently been claimed that the path integral has direct meaning only in Cartesian coordinates.[2] Nevertheless it has been shown that transforming to polar coordinates in the Lagrangian path integral, for example, actually enlarges the class of problems solvable by direct path integration.[3] Even when the exact propagator is not calculable, point canonical transformations of a nonlinear type, when correctly implemented, do yield useful information such as the correct semiclassical limit.[4]

In the discussion above, the emphasis has been on the Lagrangian form of the path integral which in one dimension is symbolically written as

$$K(q'',q';\tau) = \mathcal{N}\int \mathcal{D}\, q(t)\exp\left\{\frac{i}{\hbar}\int_0^{\tau} L(q,\dot{q};t)\,dt\right\} \quad (1.1)$$

($\mathcal{N}$ being a normalization factor), where it appears that only contact transformations are relevant. On the other hand the form of the phase-space path integral

$$K(q'',q';\tau) = \int \mathcal{D}\, p(t)\,\mathcal{D}q(t)\exp\left\{\frac{i}{\hbar}\int_0^{\tau}[p\dot{q} - H(p,q)]\,dt\right\}, \quad (1.2)$$

where $H(p,q) = \frac{1}{2}p^2 + V(q)$, has led many authors to believe that more general canonical transformations are possible. For instance, Clutton-Brock[5] and more recently Duru and Keyman[6] have attempted to use the machinery of Hamilton–Jacobi theory to obtain path integral solutions for some simple potentials. In doing so they have tacitly assumed that formal manipulations of the path integral using the classical formalism are possible. As an illustration, consider a particular time lattice version of Eq. (1.2), e.g.,

$$K(q'',q';\tau) = \lim_{n\to\infty}\int \prod_{j=1}^{N}\frac{d\,p_j}{(2\pi\hbar)}\prod_{j=1}^{N-1}dq_j \times\exp\left\{\frac{i}{\hbar}\sum_{j=1}^{N}\left[p_j(q_j - q_{j-1}) - \epsilon H\left(p_j, \frac{q_j+q_{j-1}}{2}\right)\right]\right]\Bigg\}. \quad (1.3)$$

The measure in the above could be written

$$\prod_{j=1}^{N}\frac{dp_j}{(2\pi\hbar)}\prod_{j=1}^{N-1}dq_j = \frac{dp_N}{(2\pi\hbar)}\prod_{j=1}^{N-1}\frac{dp_j dq_j}{(2\pi\hbar)}, \quad (1.4)$$

which gives the impression that, except for the unpaired dp_N, Eq. (1.4) consists of a product of Liouville measures $d\mu_i = dp_i \wedge dq_i$ and that under a canonical transformation $(q,p)\mapsto(Q,P)$ one has

$$\frac{dp_N}{(2\pi\hbar)}\prod_{j=1}^{N-1}\frac{dp_j dq_j}{(2\pi\hbar)} = \frac{dp_N}{(2\pi\hbar)}\prod_{j=1}^{N-1}\frac{dP_j dQ_j}{(2\pi\hbar)}, \quad (1.5)$$

where it is assumed that

$$\partial(P_j, Q_j)/\partial(p_j, q_j) = 1$$

for $1 \leqslant j \leqslant N-1$. However, it should be remembered that despite appearances, p_j is not really considered to be canonically conjugate to q_j in this lattice version of (1.2), rather it is taken as fixed over the interval $(j, j-1)$. This is a manifestation of the uncertainty principle. Equation (1.5) makes sense only if q and p are canonically conjugate and therefore cast suspicion on the possibility of implementing general canonical transformations. In fact the idea was shown to be questionable by Garrod[7] who noted that with the propagator for the harmonic oscillator written in terms of action-angle variables, the energy spectrum is obtained without the zero point energy and furthermore, the propagator does not satisfy the unitarity condition. In any case it is not clear how a direct transformation would be performed because of the mismatch in the number of p and q integrals (this is not true for the coherent-state integrals[8]). We thus expect that for Eq. (1.3) only contact transformations are relevant.

Restricting attention to contact transformations such that $Q = Q(q)$, $P = p(\partial q/\partial Q)$, Klauder writes the transformed propagator as

$$\lim_{N\to\infty}\int\left(\frac{\partial Q''}{\partial q''}\right)\prod_{j=1}^{N}\frac{dP_j}{(2\pi\hbar)}\prod_{j=1}^{N-1}dQ_j \times\exp\left[\frac{i}{\hbar}\sum_{j=1}^{N}\left\{P_j\frac{\partial Q_j}{\partial q_j}(Q_j)[q_j(Q_j) - q_{j-1}(Q_{j-1})] - \epsilon H'(P_j,Q_j)\right\}\right], \quad (1.6)$$

where

$$H'(P_j,Q_j) = H\left(P_j\frac{\partial Q_j}{\partial q_j}(Q_j), q_j(Q_j)\right).$$

In the continuous limit we have

$$\frac{\partial Q''}{\partial q''}\int \mathscr{D}\, P(t)\mathscr{D}Q(t)\exp\left\{\frac{i}{\hbar}\int_0^T [P\dot{Q}-H'(P,Q)]dt\right\}. \tag{1.7}$$

Here again it has been assumed that (1.5) is valid and that the factor $(\partial Q''/\partial q'')$ arises from the unpaired dp''. However, using a harmonic oscillator coupled to external sources $H(p,q)=\frac{1}{2}(p^2+\omega^2q^2)-Jq$, Gervais and Jevicki[9] used the Feynman diagram technique to show that (1.6) leads to erroneous results. They went on to show that starting with the *Lagrangian* form of the path integral, that the correct propagator is obtained by making the transformation $q=f(Q)$ in each short-time propagator expanding $f(Q_j)$ and $f(Q_{j-1})$ about the midpoint $\overline{Q}_j=(Q_j+Q_{j-1})/2$ and[10] retaining terms up to order $(\Delta Q_j)^4/\epsilon\sim\epsilon$. The extra terms can be cast into a correction term to the potential of the form

$$\Delta V(\overline{Q}_j)=\hbar^2\left\{\frac{1}{8}\frac{f'''(\overline{Q}_j)}{[f'(\overline{Q}_j)]^3}\right.$$
$$\left.-\frac{1}{8}\left[\left(\frac{f'''(\overline{Q}_j)}{f'(\overline{Q}_j)}\right)^2-\frac{f'''(\overline{Q}_j)}{f'(\overline{Q}_j)}\right][f'(\overline{Q}_j)]^2\right\}, \tag{1.8}$$

where the primes refer to derivatives with respect to Q.

We felt it is instructive to examine contact transformations starting from the phase-space path integral rather than the Lagrangian form. In the next section we discuss a version of the midpoint method suitable for transforming Eq. (1.3). In the foregoing discussion it will become clear that Eq. (1.5) is false for all contact transformations but a scaling of coordinates. Finally it should be pointed out that we begin with the classical variables rather than unitary transformations in quantum mechanics as in the work of Fanelli.[11]

II. CONTACT TRANSFORMATION IN THE PHASE-SPACE PATH INTEGRAL

From classical mechanics,[12] canonical transformations are obtained from a generating function F such that

$$p\dot{q}-H(q,p,t)=P\dot{Q}-K(P,Q,T)+\frac{dF}{dt}, \tag{2.1}$$

where

$$\delta\int_{t_1}^{t_2}\frac{dF}{dt}dt=0.$$

With $F=F_1(q,Q,t)$ we obtain $p=\partial F_1/\partial q$, $P=-\partial F_1/\partial Q$, and $K=H+\partial F_1/\partial t$. To obtain contact transformations we write $F_1=F_2(q,P,t)-QP$ with $F_2=Q(q)P$ so that $K\equiv H$ and

$$p=\frac{\partial F_2}{\partial q}=\frac{\partial Q(q)}{\partial q}P, \tag{2.2}$$

$$Q(q)=\frac{\partial F_2}{\partial P}. \tag{2.3}$$

Assuming that Eq. (2.3) can be inverted to yield $q=f(Q)$, we follow Gervais and Jevicki[9] and expand $f(Q_j)$ and $f(Q_{j-1})$ about the midpoint $\overline{Q}_j=(Q_j+Q_{j-1})/2$, retaining terms to order ϵ. We obtain

$$\begin{aligned}q_j&=f(\overline{Q}_j)+\tfrac{1}{2}f'(\overline{Q}_j)\Delta Q_j+\tfrac{1}{8}f''(\overline{Q}_j)(\Delta Q_j)^2\\&\quad+\tfrac{1}{48}f'''(\overline{Q}_j)(\Delta Q_j)^3+\cdots,\\q_{j-1}&=f(\overline{Q}_j)-\tfrac{1}{2}f'(\overline{Q}_j)\Delta Q_j+\tfrac{1}{8}f''(\overline{Q}_j)(\Delta Q_j)^2\\&\quad-\tfrac{1}{48}f'''(\overline{Q}_j)(\Delta Q_j)^3+\cdots.\end{aligned} \tag{2.4}$$

Now instead of using (2.2) directly in (1.3) as was done by Klauder[8] in Eq. (1.6), we write it as

$$p_j=P_j(\Delta Q_j/\Delta q_j), \tag{2.5}$$

as the momenta are defined over the interval rather than the endpoints. We have Δq_j from Eqs. (2.4) so that Eq. (2.5) becomes

$$p_j=P_j(f')^{-1}\left[1-\frac{1}{24}\frac{f'''}{f'}(\Delta Q_j)^2\right], \tag{2.6}$$

where it is to be understood that f' and f''', etc. are evaluated at $\overline{Q}_j$. Now from Eq. (1.3) the short action

$$S(t_j-t_{j-1})=p_j\Delta q_j-\epsilon[p_j^2/2m+V(q_j)], \tag{2.7}$$

using (2.5) and (2.6), becomes

$$S(t_j-t_{j-1})=P_j\Delta Q_j-\epsilon\left[\frac{P_j^2}{2m}(f')^{-2}-\frac{P_j^2f'''}{24m(f')^3}(\Delta Q_j)^2\right.$$
$$\left.+V[f(\overline{Q}_j)]\right]. \tag{2.8}$$

Upon exponentiation we have

$$\begin{aligned}&\exp\left\{\frac{i}{\hbar}S(t_j-t_{j-1})\right\}\\&\quad=\exp\left\{\frac{i}{\hbar}P_j\Delta Q_j-\frac{i\epsilon}{\hbar}\left[\frac{1}{2m}P_j^2(f')^{-2}+V[f(\overline{Q}_j)]\right]\right\}\\&\quad\times\left\{1+\frac{i\epsilon}{\hbar}\frac{P_j^2}{2m}\frac{f'''}{(f')^3}(\Delta Q_j)^2\right\},\end{aligned} \tag{2.9}$$

where we have retained the term $\epsilon P_i^2(\Delta Q_j)^2$ as it is of order ϵ.

We now consider the transformation of the measure. For the reasons previously discussed, we treat dq_j and dp_j separately. To begin we symmetrize dq_j about Q_j and Q_{j-1} to obtain

$$\prod_{j=1}^{N-1}dq_j=[f'(Q'')f'(Q')]^{-1/2}\prod_{j=1}^{N}[f'(Q_j)f'(Q_{j-1})]^{1/2}$$
$$\times\prod_{j=1}^{N-1}dQ_j. \tag{2.10}$$

We expand $f'(Q_j)$ and $f'(Q_{j-1})$ about $\overline{Q}_j$ to get

$$\prod_{j=1}^{N-1}dq_j=[f'(Q'')f'(Q')]^{-1/2}$$
$$\times\prod_{j=1}^{N}f'(\overline{Q}_j)\left\{1-\frac{1}{8}\lambda(\overline{Q}_j)(\Delta Q_j)^2\right\}\prod_{j=1}^{N-1}dQ_j, \tag{2.11}$$

where

$$\lambda(\overline{Q}_j)=\left(\frac{f''(\overline{Q}_j)}{f'(\overline{Q}_j)}\right)^2-\frac{f'''(\overline{Q}_j)}{f'(\overline{Q}_j)}. \tag{2.12}$$

For the momentum measure we simply use Eq. (2.6) to write

$$dp_j=dP_j(f')^{-1}\left[1-\frac{1}{24}\frac{f'''}{f'}(\Delta Q_j)^2\right]. \tag{2.13}$$

Thus, combining Eqs. (2.11) and (2.13) we have, to order ϵ,

$$\prod_{j=1}^{N} \frac{dp_j}{(2\pi\hbar)} \prod_{j=1}^{N-1} dq_j = [f'(Q'')f'(Q')]^{-1/2} \prod_{j=1}^{N} \frac{dP_j}{2\pi\hbar} \prod_{j=1}^{N-1} dQ_j \times \prod_{j=1}^{N} \left[1 - \left(\frac{1}{24}\frac{f'''}{f'} + \frac{1}{8}\lambda\right)(\Delta Q_j)^2\right]. \tag{2.14}$$

Now, combining the results of Eqs. (2.9) and (2.14) we have the transformed propagator

$$\begin{aligned} &K[f(Q''),f(Q'),\tau] \\ &= \lim_{N\to\infty} [f'(Q'')f'(Q')]^{-1/2} \int \prod_{j=1}^{N} \frac{dP_j}{(2\pi\hbar)} \prod_{j=1}^{N-1} dQ_j \\ &\quad \times \exp\left\{\frac{i}{\hbar} \sum_{j=1}^{N} \left[P_j \Delta Q_j - \frac{\epsilon P_j^2}{2m}(f')^{-2} - \epsilon V[f(\overline{Q}_j)]\right]\right\} \\ &\quad \times \prod_{j=1}^{N} \left\{1 - \left(\frac{1}{24}\frac{f'''}{f'} + \frac{\lambda}{8}\right)(\Delta Q_j)^2 + \frac{i\epsilon}{\hbar}\frac{P_j^2}{24m}\frac{f'''}{(f')^3}(\Delta Q_j)^2\right\}. \end{aligned} \tag{2.15}$$

To obtain a more useful form we do the following. We note that if we replace $\overline{Q}_j$ in f', f'', etc., by Q_{j-1} the difference in Eq. (2.15) will be on order $\epsilon^{3/2}$ or higher. Next we use the identities

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx\, dy\, e^{-ay^2 + bxy} x^2 = \frac{\pi}{2\sqrt{a}}\left(-\frac{b^2}{4a}\right)^{-3/2}, \tag{2.16a}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dx\, dy\, e^{-ay^2 + bxy} x^2 y^2 = \frac{\pi}{2}\left(-\frac{b^2}{4a}\right)^{-3/2}\left[1 + \frac{3}{16}\left(-\frac{b^2}{4a}\right)^{-1}\right] \tag{2.16b}$$

to see that (2.15) differs from

$$\begin{aligned} &K[f(Q''),f(Q'),\tau] \\ &\quad = \lim_{N\to\infty} [f'(Q'')f'(Q')]^{-1/2} \int \prod_{j=1}^{N} \frac{dP_j}{(2\pi\hbar)} \prod_{j=1}^{N-1} dQ_j \\ &\quad \times \exp\left\{\frac{i}{\hbar} \sum_{j=1}^{N} \left[P_j \Delta Q_j - \epsilon\frac{P_j^2}{2m}(f'(\overline{Q}_j))^{-2} \right.\right. \\ &\quad \left.\left. - \epsilon V[f(\overline{Q}_j)] - \epsilon \Delta V(\overline{Q}_j)\right]\right\}, \end{aligned} \tag{2.17}$$

where $\Delta V(\overline{Q}_j)$ is given by Eq. (1.8) by terms of order $\epsilon^{3/2}$ or higher. One can now perform the momentum integrations to obtain the Lagrangian path integral

$$\begin{aligned} K[f(Q''),f(Q');\tau] = &\lim_{N\to\infty} [f'(Q'')f'(Q')]^{-1/2} \\ &\times \left(\frac{m}{2\pi i\hbar T}\right)^{N/2} \int \prod_{j=1}^{N} f'(\overline{Q}_j) \prod_{j=1}^{N-1} dQ_j \\ &\times \exp\left\{\frac{i}{\hbar} \sum_{j=1}^{N} \left[\frac{m}{2\epsilon}[f'(\overline{Q}_j)]^2(\Delta Q_j)^2 \right.\right. \\ &\left.\left. - \epsilon V[f(\overline{Q}_j)] - \epsilon \Delta V(\overline{Q}_j)\right]\right\}, \end{aligned} \tag{2.18}$$

which is the one-dimensional version of the results of Gervais and Jevicki.[9] Assuming that $K(q'', q';\tau)$ propagates the wave function $\psi(q,t)$ via

$$\psi(q'',\tau) = \int K(q'', q';\tau)\psi(q',0)dq', \tag{2.19}$$

then the propagator

$$\widetilde{K}(Q'',Q';\tau) = [f'(Q'')f'(Q')]^{1/2} K[f(Q''),f(Q');\tau]$$

propagates the wave functions $\widetilde{\psi}(Q,t) = [f'(Q)]^{-1/2}\psi(f(Q),t)$, as

$$\widetilde{\psi}(Q'',\tau) = \int \widetilde{K}(Q'',Q';\tau)\widetilde{\psi}(Q',0)dQ'. \tag{2.20}$$

We have previously used the results above to calculate radial path integrals semiclassically.[4] In order to perform the semiclassical path integration it is necessary to transform from the radial variable r whose range if $[0,\infty)$ to one whose range is $(-\infty,\infty)$. In the radial path integral[3] the effective potential is

$$U(\sqrt{r_j r_{j-1}}) = \frac{\hbar^2 l(l+1)}{2m r_j r_{j-1}} + V(\sqrt{r_j r_{j-1}}). \tag{2.21}$$

A transformation which maps the semi-infinite range of r onto the infinite range of x is $r = e^x$. Applying this in Eq. (1.8) yields

$$\Delta V(\bar{x}_j) = (\hbar^2/8m)e^{-2\bar{x}_j} \tag{2.22}$$

so that the effective potential gets modified to

$$U(e^{\bar{x}_j}) + \Delta V(\bar{x}_j) = \frac{\hbar^2(l+\frac{1}{2})^2 e^{-2\bar{x}_j}}{2m} + V(e^{\bar{x}_j}). \tag{2.23}$$

The transformation $r = e^x$ has effectively modified the angular momentum term: $l(l+1) \to (l+\frac{1}{2})^2$. This is the Langer modification which is necessary to obtain energy levels with the correct degeneracy for the Coulomb and harmonic oscillator potentials. It should be pointed out that the transformation $r = e^x$ as opposed to any other which maps $[0,\infty)$ to $(-\infty,\infty)$, is used as only it yields phase shifts in semiclassical scattering which vanish as $V(r) \to 0$.

III. DISCUSSION

We have used here a particular procedure, the midpoint method, to implement a contact transformation in the phase-space integral. It is not clear how other canonical transformation could be implemented for path integrals in the Schrödinger representation since they are not symmetrical in the p and q integrals. Furthermore, the procedure used here may not be unique as, indeed, a particular lattice space version of path integrals is not unique.[8] Nevertheless, the results here are known to be correct as they have been tested via the diagrammatic technique in perturbative field theory[9] and in nonrelativistic semiclassical quantum mechanics.[4] The main point of the work is that formal manipulations of phase space paths using the classical formalism are unjustified. In particular, Eq. (1.4) cannot be treated as a Liuoville measure except, as can be seen from (2.14), a scaling of the coordinates. It seems likely that the correct results obtained in Refs. 5 and 6 are due to the fact that they have treated simple potentials such as the harmonic oscillator, where exact and semiclassical results coincide.[13]

It would be interesting to see how the results here are related to other definitions of the phase path integral which do not rely on a lattice space formulation followed by a limit.[14] One should also consider the coherent-state path integral which is symmetrical in p and q and presumably exhibits covariance under general canonical transformations.[8] It is likely, however, that one might expect there to arise corrections to the Hamiltonian as happens in the Schrödinger representation. These matters are being investigated and will be reported elsewhere.

[1] L. Cohen, J. Math. **11**, 3296 (1970).

[2] M. S. Marinov, Phys. Rep. **60**, 1 (1980).

[3] D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969); A. Inomata and V. A. Singh, J. Math. Phys. **19**, 2318 (1978); C. C. Gerry and V. A. Singh, Phys. Rev. D **20**, 2550 (1979); **21**, 2979 (1980).

[4] C. C. Gerry and A. Inomata, Phys. Lett. A **84**, 172 (1981); J. Math. Phys. **23**, 2402 (1982).

[5] M. Clutton-Brock, Proc. Cambridge Philos. Soc. **61**, 201 (1965).

[6] L. H. Duru and E. Keyman, International Center for Theoretical Physics Preprint IC1801129 (1980).

[7] C. Garrod, Rev. Mod. Phys. **38**, 483 (1966).

[8] J. R. Klauder, Acta. Phys. Austriaca Suppl. XXII, 3 (1980).

[9] J. L. Gervais and A. Jevicki, Nucl. Phys. B **110**, 53 (1976).

[10] S. F. Edwards and Y. V. Gulyaev, Proc. R. Soc. London Ser. A **279**, 229 (1964).

[11] R. Fanelli, J. Math. Phys. **17**, 490 (1976).

[12] H.Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1950).

[13] L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981), Chap. 6.

[14] M. N. Mizrahi, J. Math. Phys. **19**, 298 (1978); C. DeWitt-Morette, Comm. Math. Phys. **28**, 47 (1972).

A recurrence relation for the phase shifts of exponential and Yukawa potentials

B. G. Sidharth[a)]
Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy
International Centre for Theoretical Physics, Trieste, Italy

(Received 23 July 1981; accepted for publication 25 January 1982)

Using the Born approximation for high energies, we deduce a recurrence relation for the higher order phase shifts of Yukawa and exponential potentials and their superpositions. Thence a formula for the total cross section is deduced. It is shown that these formulae are particularly useful when the energy is very high. Next, using the low energy technique of expansion in powers of the energy, we deduce an asymptotic formula when l is large, for δ_{l+1}/δ_l. The remarkable thing is that the recurrence relation is identical to the asymptotic formula. It is verified that the formulae for the phase shifts give satisfactory results, which are better than the Born approximation, from $l = 1$ onwards, for scattering by helium and hydrogen atoms, while the formula for the total cross section gives better results than an improved variational technique or the Born approximation.

PACS numbers: 03.65.Nk

I. INTRODUCTION

In this paper, we first obtain a recurrence relation for the higher-order phase shifts of Yukawa and exponential potentials and also a superposition of the two, for high energies and/or weak potentials, for which the first Born approximation is known to be good.

Next, for the same class of attractive potentials, we obtain an asymptotic formula for large l at low energies, in which case, the phase shifts can be expanded as a power series in K.

The remarkable feature is that, though these are well-separated energy domains, the recurrence relation and the asymptotic formula are equivalent.

II. THE RECURRENCE RELATION

We consider high energies and/or weak potentials. Our starting point is the radial Schrödinger equation,

$$u_l'' + [K^2 - l(l+1)/r^2 - \lambda U(r)]u_l = 0, \quad u_l(0) = 0, \tag{1}$$

for which it is known that[1] the phase shifts δ_l are given by

$$\sin\delta_l = -K\lambda \int_0^\infty rj_l(Kr)u_l(r)U(r)\,dr, \tag{2}$$

while Eq. (1) can be written in integral form as

$$u_l = rj_l(Kr)\cos\delta_l + \lambda Krn_l(Kr)\int_0^r r'j_l(Kr')U(r')u_l(r')\,dr' + \lambda Krj_l(Kr)\int_r^\infty r'n_l(Kr')U(r')u_l(r')\,dr'. \tag{3}$$

Introducing (3) in (2), we get,

$$\tan\delta_l(\lambda) = \lambda KB_l + O(\lambda^2), \tag{4}$$

where

$$B_l = -\int_0^\infty [rj_l(Kr)]^2U(r)\,dr. \tag{5}$$

So, for small λ, (4) becomes

$$\tan\delta_l = \lambda KB_l, \tag{6}$$

which is the first Born approximation for δ_l. As is well known, (6) can be obtained directly from (1) and (2): If $|\lambda U(r)| \ll K^2$ except near the origin, where $-l(l+1)/r^2$ dominates anyway, then u_l approximately satisfies the equation

$$u_l'' + [K^2 - l(l+1)/r^2]u_l = 0,$$

with the only admissible solution

$$u_l = rj_l(Kr).$$

When this is substituted in (2), we get

$$\sin\delta_l = \lambda KB_l. \tag{7}$$

This is an equivalent form of (6), because, at the high energies or low potential strengths at which the Born approximation (6) or (7) is valid, the phase shift δ_l is small, so that $\tan\delta_l \approx \sin\delta_l \approx \delta_l$. So (6) or (7) is valid for high energies and/or weak potentials, for any form of the potential.

We will now deduce a recurrence relation for the B_l for the exponential potential, $U(r) = \lambda\exp(-br)$, and the Yukawa potential, $U(r) = (\lambda/r)\exp(-br)$.

Following (5), let us define

$$B_l^{(m)} \equiv \int_0^\infty r^m\exp(-br)[rj_l(Kr)]^2\,dr, \quad m = -1,0.$$

Next, we use the formula[2]

$$\int_0^\infty \exp(-br)\ J_\nu^2(Kr)dr = (1/\pi K)Q_{\nu-1/2}((b^2 + 2K^2)/2K^2),$$

where Q_l is the Legendre function of the second kind.

Remembering that $J_{l+1/2}(r) = [(2r/\pi)\ j_l(r)]^{1/2}$, we get,

---

[a)] Permanent address: Birla Planetarium, 96, Chowringhee Road, Calcutta 700071, India.

0022-2488/83/040878-05$02.50

on putting $\nu = l + \frac{1}{2}$,

$$B_l^{(-1)} = (1/2K^2)\cdot Q_l(b^2/2K^2 + 1). \tag{8}$$

Now applying the formula[3]

$$(l+1)Q_{l+1}(z) - (2l+1)zQ_l(z) + lQ_{l-1}(z) = 0, \tag{9}$$

we get finally

$$2(l+1)B_{l+1}^{(-1)} + 2lB_{l-1}^{(-1)} = (2l+1)\cdot(b^2/K^2 + 2)B_l^{(-1)}, \tag{10}$$

which is a recurrence relation for the $B_l^{(-1)}$.

To obtain a recurrence relation for the $B_l^{(0)}$, we observe that it can be easily proved that the integrals for $B_l^{(-1)}$ are uniformly convergent with respect to b. So differentiating (8) with respect to b, within the integral, we get

$$B_l^{(0)} = -\frac{1}{2K^2}\frac{d}{db}\left[Q_l\left(\frac{b^2}{2K^2} + 1\right)\right]. \tag{11}$$

Next, using the formula (cf. Ref. 3),

$$Q'_{l+1}(z) - Q'_{l-1}(z) = (2l+1)Q_l(z)$$

in conjunction with (9), further manipulation yields

$$(2l+1)(b^2/K^2 + 2)B_l^{(0)} = 2lB_{l+1}^{(0)} + 2(l+1)B_{l-1}^{(0)}, \tag{12}$$

which is a recurrence relation for the $B_l^{(0)}$.

Owing to the smooth behavior of the spherical Bessel functions with respect to l, we can write $(B_{l+1}/B_l) \rightarrow \beta$ as $l \rightarrow \infty$, in (10) and (12).

For large l, (10) and (12) yield

$$(2 + b^2/K^2)\beta = 1 + \beta^2. \tag{13}$$

We choose

$$\beta = \tfrac{1}{2}\left[(2 + b^2/K^2) - \left(\frac{b}{K}\right)(4 + b^2/K^2)^{1/2}\right]. \tag{14}$$

(For the other root, which is > 1, the series for the scattering amplitude diverges.)

In view of (6) or (7), we can therefore write, owing to the fact that $\delta_l \rightarrow 0$, as $l \rightarrow \infty$,

$$\frac{\tan\delta_{l+1}}{\tan\delta_l} \approx \frac{\sin\delta_{l+1}}{\sin\delta_l} \approx \frac{\delta_{l+1}}{\delta_l} \approx \beta, \quad \text{for large } l, \tag{15}$$

for high energies ($K^2 \gg 1$) and/or weak potentials ($\lambda \ll 1$).

We now show that, for large l, (15) is valid for potentials which are a superposition of an exponential and a Yukawa potential, viz.,

$$U(r) = [A + (B/r)]\exp(-br).$$

For such potentials, we define

$$B_l = \int_0^\infty (A + B/r)\exp(-br)[rj_l(Kr)]^2\,dr.$$

Also, (10) and (12) take, for large l, the asymptotic forms

$$B_{l+1}^{(-1)} + B_{l-1}^{(-1)} = \left(\frac{b^2}{K^2} + 2\right)B_l^{(-1)},$$

$$B_{l+1}^{(0)} + B_{l-1}^{(0)} = \left(\frac{b^2}{K^2} + 2\right)B_l^{(0)},$$

respectively.

Multiplying the first of these by B and the second by A and adding, we get

$$B_{l+1} + B_{l-1} = (b^2/K^2 + 2)B_l.$$

TABLE I. Phase shifts for scattering by helium atoms.

K	δ_2 (exact)[a]	δ_2 (formula)[b]	δ_2 (Born)
3	0.0946	0.0920	0.0769
4	0.1304	0.13088	0.1130
5	0.1524	0.1569	0.1378

[a] The exact values of δ_2.[4]

[b] The values of δ_2 as computed from formula (16), using exact values of δ_1 as obtained from Ref. 4.

So, if $B_{l+1}/B_l \rightarrow \beta$ as $l \rightarrow \infty$, we get back Eq. (13) for β and then (15).

Equation (15) can be used to calculate the cross sections. For example, if the equation

$$\sin\delta_{l+1}/\sin\delta_l = \beta \tag{16}$$

is a good approximation for $l \geqslant L$, then

$$\sigma = \frac{4\pi}{K^2}\left[\sum_{m=0}^{L-1}(2m+1)\sin^2\delta_m\right] + \frac{4\pi}{K^2}\left[\sum_{m=L}^{\infty}(2m+1)\sin^2\delta_m\right].$$

Using (16), the second summation above can be written as

$$\frac{\sin^2\delta_L}{\beta^{2L}}[(2L+1)\beta^{2L} + (2L+3)\beta^{2L+2} + \dots]$$
$$= \frac{\sin^2\delta_L}{\beta^{2L}}\frac{d}{d\beta}\left[\sum_{m=L}^{\infty}\beta^{2m+1}\right]$$
$$= \frac{(2L+1) - (2L-1)\beta^2}{(1-\beta^2)^2}\sin^2\delta_L,$$

whence

$$\sigma = \frac{4\pi}{K^2}\left[\sum_{m=0}^{L-1}(2m+1)\sin^2\delta_m\right] + \frac{4\pi}{K^2}\frac{(2L+1) - (2L-1)\beta^2}{(1-\beta^2)^2}\sin^2\delta_L. \tag{17}$$

In particular, if $L = 0$, (17) becomes

$$\sigma = \frac{4\pi}{K^2}\cdot\left[\frac{1+\beta^2}{(1-\beta^2)^2}\right]\sin^2\delta_0. \tag{18}$$

We now make two remarks:

In the high energy case, $b/k \ll 1$, and so, from (14), we get $\beta \approx 1 - b/K \approx 1$. In this case, the relation (15) or (16) is par-

TABLE II. Phase shifts for scattering by hydrogen atoms.

K	δ_2 (exact)[a]	δ_2 (formula)[b]
1	0.0178	0.01909
0.5	0.0056	0.0059
0.25	0.0014	0.0014

[a] The exact values of δ_2.[6]

[b] The values of δ_2 as computed from formula (16), using exact values of δ_1 given in Ref. 6. The agreement is good considering that we are not in the high energy case. We also have that, for energies K^2 equal to 1.75 and 1.5, $\sin\delta_1/\sin\delta_0$, as given by (16), is, respectively, 0.247 6431 and 0.225 1483 while its actual values are, respectively, 0.215 4928 and 0.197 5663.[7]

TABLE III. Total cross section σ for scattering by hydrogen atoms.

K^2	σ (exact) is greater than[a]	σ (formula)[b]	σ (variational)[c]	σ (Born)[d]
1	2.62	2.70	2.34	1.54
1.75	1.40	1.48	1.34	1.04
2.25	1.09	1.136	1.05	0.854
4	0.60	0.64	0.587	0.522

[a] The cross section calculated using accurate values of the first few phase shifts.[8,9]

[b] The cross section using formula (18).

[c] The total cross section using a sophisticated variational technique.

[d] The total cross section using the Born approximation.[8]

ticularly useful. For, as is usually done, one could calculate the first L phase shifts and then compute the cross section, neglecting the contribution of the δ_l for $l > L$. However, even though these δ_l may be individually small, (17) shows that, owing to the factor $[(2L+1)-(2L-1)\beta^2]/(1-\beta^2)^2$, their collective contribution to the scattering cross section is appreciable.

Next we note that (15) holds for small values of $|\delta_l|$. In fact, in the absence of resonances, if $|\delta_l|$ is small, then so are the subsequent phase shifts, because they fall off monotonically.[4,5] A large value of $|\delta_l|$, like $2\pi + |\theta|$ is associated with a resonance. In such a case, neither the Born approximation nor (15) hold for that particular l.

Formulae (15) or (16) and (18) are numerically illustrated in Tables I, II, and III for two cases of practical interest: for scattering by the static field of the hydrogen atom, in which case

$$U(r) = -2(1/r+1)\exp(-2r);$$

for scattering by an approximate static field of the helium atom, in which case

$$U(r) = -2(1/r+z)\exp(-2zr), \quad z = 27/16.$$

It is seen that formula (16) for the phase shifts gives good results from $l = 1$ itself, even though the energies are not high and even though, for the potentials considered, the large l version of (10) and (12) are required in the derivation of (16), so that when l is not large, (16) is only approximate. Moreover, for the total cross section, formula (18), which uses (16) from $l = 0$ itself, already gives better results than a refined variational technique or the Born approximation.

III. THE ASYMPTOTIC FORMULA

First we quote the following results[10]:

At low energies, the phase shifts can be computed from the function $t_l(r) = \tan\delta_l(r)$, where $t_l(\infty) = \tan\delta_l$, δ_l being the usual phase shifts, and where

$$t_l(r) = K^{2l+1}[a_l(r) + O(K^2)].$$

If we write

$$a_l(r) \equiv \alpha_l(r)/\{[(2l+1)!!][(2l-1)!!]\},$$

the $\alpha_l(r)$ are known to satisfy the Riccati equation

$$\alpha_l'(r) = \frac{-\lambda U(r)}{(2l+1)}[r^{l+1} + r^{-l}\alpha_l(r)]^2 \tag{19}$$

with the following behavior near the origin,

$$\alpha_l(r) \sim r^{2l+3-m} \quad \text{as } r \to 0, \tag{20}$$

where

$$U(r) \sim r^{-m} \quad \text{as } r \to 0.$$

In the subsequent discussion we will be interested only in the case $m < 2$, and attractive potentials, $\lambda U(r) < 0$ everywhere, and also sufficiently large l for which there are no bound states.

In the absence of bound states it is known that $\delta_l(r)$ is a continuous function of r. From (19) and (20) it now follows that, as $\alpha_l(0) = 0$ and $\alpha_l'(r) > 0$, so

$$\alpha_l(r) > 0 \quad \text{for } r > 0. \tag{21}$$

Also, it is seen from (19) that $\alpha_l'(r)$ itself is a continuous function of r. Substituting (21) in (19), we get the lower bound

$$\alpha_l \equiv \alpha_l(\infty) > -\frac{\lambda}{(2l+1)}\int_0^\infty U(r) r^{2l+2}\, dr. \tag{22}$$

Next we will deduce an upper bound for α_l. From the known behavior of $\alpha_l(r)$ for small r, given in (20) and from (21), it follows that

$$\alpha_l(r)/(pr^{2l+1}) \sim r^{2-m} \to 0+ \quad \text{as } r \to 0,$$

where p is an arbitrary positive number. So,

$$\alpha_l(r) \leqslant pr^{2l+1}, \quad \text{for small } r. \tag{23}$$

We shall prove that, given an arbitrarily small $p > 0$, there exists an L such that (23) is true for all r whenever $l > L$.

First, we observe that as $\alpha_l(r)$ and pr^{2l+1} are both continuous functions, and, in view of (23), there can be only two possibilities:

(A) There is at least one number $r_1 > 0$ such that

$$\begin{aligned} &\alpha_l(r) < pr^{2l+1} \quad \text{for } r < r_1,\\ &\alpha_l(r_1) = pr_1^{2l+1},\\ &\alpha_l(R) > pR^{2l+1}, \quad \text{for some } R > r_1, \end{aligned} \tag{24}$$

(B) $\alpha_l(r) \leqslant pr^{2l+1}$ for all $r > 0$. $\quad$ (25)

We first consider case (A). We divide both sides of (19) by $\{\alpha_l(r)\}^2$, and then substitute the third relation of (24) and integrate between r_1 and R. This leads to

$$\frac{1}{\alpha_l(r_1)} - \frac{1}{\alpha_l(R)} < \frac{-\lambda(1/p+1)^2}{(2l+1)}\int_{r_1}^R U(r) r^{-2l}\, dr.$$

Our choice of the potential ensures that

$$-\lambda U(r) = |\lambda (U(r)| < A/r^2, \quad \text{for all } r > 0,$$

where A is some (*a priori*) fixed number, whence we get, after using the second relation in (24),

$$\frac{1}{\alpha_l(R)} > \frac{1}{pr_1^{2l+1}} - \frac{1}{r_1^{2l+1}} \cdot \frac{(1+1/p)^2}{(2l+1)^2} A \left[1 - \left(\frac{r_1}{R}\right)^{2l+1}\right].$$

Now $(r_1/R)^{2l+1} \to 0$ as $l \to \infty$, because $(r_1/R) < 1$. So we can write

$$\frac{pr_1^{2l+1}}{\alpha_l(R)} > \left[1 - \frac{A}{(2l+1)^2}\left(p + \frac{1}{p} + 2\right)(1+\epsilon_1)\right], \tag{26}$$

$$\text{for } l > L_1,$$

where L_1 is some large positive number.

Let us choose a number B, such that $A/B = \epsilon_2$ is as small as desired. Also p is arbitrary and as yet unspecified. Let us write

$$p = B/(2l+1)^2 \tag{27}$$

and consider values of $l > L_2$, say, for which $p = \epsilon_3$ is as small as we please. Specifically,

$$L_2 = \tfrac{1}{2}(B/\epsilon_3)^{1/2} = \tfrac{1}{2}(A/\epsilon_2\epsilon_3)^{1/2}.$$

We note that $r_1 > 0$ depends on the choice of p and l, but we need not bother as long as this choice leaves us in case (A). If this choice is possible only under (B), this will be discussed separately.

To sum up, we *choose* p according to (27), where B is any l-independent number, which merely satisfies the requirement $(A/B) \ll 1$, A being some given (positive) number. Further, we *choose* L_2 such that if $l > L_2$, then $p \ll 1$.

Inequality (26) now becomes

$$pr_1^{2l+1}/\alpha_l(R) > \{1 - (A/B)(1+\epsilon_1) + O(\epsilon_2\epsilon_3)\}$$

$$= [1 - \epsilon_2(1+\epsilon_1)]$$

for $l > \max(L_1, L_2) \equiv L$.

Finally, on using the last inequality of (24), this becomes

$$(r_1/R)^{2l+1} > pr_1^{2l+1}/\alpha_l(R) > (1-\epsilon_2) \quad \text{for } l > L. \tag{28}$$

As $(r_1/R) < 1$, $(r_1/R)^{2l+1}$ can be made arbitrarily small by choosing l suitably large. Hence (28) is impossible.

For completeness it may be added that inequality (28) may hold if the zeroes of the function $\alpha_l(r) - pr^{2l+1}$ are unbounded and further if the distance between successive zeroes becomes arbitrarily small as $l \to \infty$ where $p = B/(2l+1)^2$. In that case, if $r_1 > 0$ and $r_2 > 0$ are successive zeroes and $r_1 < R < r_2$, then $(r_1/R) \to 1$ as $l \to \infty$. But this possibility is ruled out owing to the smooth behavior of $\alpha_l(r)$ and pr^{2l+1}.

Thus we are left only with case (B). That is

$$\alpha_l(r) < pr^{2l+1} \quad \text{for } r > 0 \text{ and } l > L,$$

where p is given by (27).

Substituting this in (19) and integrating, we get the desired upper bound

$$\alpha_l(\infty) < [-\lambda(1+p)^2/(2l+1)] \int_0^\infty U(r) r^{2l+2}\, dr,$$

where $p \to 0$ as $l \to \infty$.

TABLE IV. Phase shifts for scattering by hydrogen atoms (low energy).

K^2	δ_2 (exact)[a]	δ_2 (formula)[b]
0.25	0.0014	0.0016
0.10	0.0002	0.0002
0.010	0.0000	0.0000

[a] The exact values of δ_2.[6]

[b] The values of δ_2 as computed from formula (32), using exact values of δ_1 as given in Ref. 6.

Comparing this with inequality (22), we deduce that

$$\alpha_l \approx [-\lambda/(2l+1)] \int_0^\infty U(r) r^{2l+2}\, dr. \tag{29}$$

We remark that no specific form of the potential has been assumed in deducing (29).

Using the definition of δ_l in terms of $\alpha_l(\infty) \equiv \alpha_l$, we can easily deduce that

$$\frac{\delta_{l+1}}{\delta_l} = \frac{K^2}{(2l+3)^2} \frac{\int_0^\infty U(r) r^{2l+4}\, dr}{\int_0^\infty U(r) r^{2l+2}\, dr}. \tag{30}$$

We first consider (30) for the attractive potential,

$$U(r) = -r^m \exp(-br^2).$$

Then we get

$$\frac{\delta_{l+1}}{\delta_l} = \frac{K^2}{(2l+3)^2} \cdot \frac{(2l+3+m)}{2b} \to \frac{K^2}{4bl} \tag{31}$$

for large l. This result is identical to the known relation for any energy, for this class of potentials.[11] This provides a check on the validity of (29).

We next use (30) for $U(r) = -\lambda r^m \exp(-br)$. This time we get, for large l,

$$\delta_{l+1}/\delta_l = (K/b)^2. \tag{32}$$

It can be verified that, for large l, (32) also holds for $U(r) = -(\Sigma_{i=1}^n A_i r^{m_i}) \exp(-br)$, where A_i are all > 0 and the m_i, which need not be integers, are all > -2.

It is remarkable that (32) can be deduced from (15) with β given by (14), in the limit $K \to 0$. [If b is so small that $(K/b) > 1$, it means that the potential is so strong that our assumption that there are no bound states or zero-energy resonances is no longer credible, and our low-energy approximation is not valid. However, for fixed b, (32) or (29) is correct for sufficiently low energies.]

Formula (32) is illustrated in the case of scattering by the static field of the hydrogen atom, in Table IV. The formula is seen to give good results from $l = 1$ itself.

IV. CONCLUSION

We have deduced a recurrence relation for the higher-order phase shifts of exponential and Yukawa potentials and their super-positions, for high energies, and an asymptotic formula for δ_{l+1}/δ_l for low energies. The recurrence relation is identical with the asymptotic formula: (32) and (15) [or (16)] can both be written as

$$\delta_{l+1}/\delta_l = \beta, \tag{33}$$

where β, which is independent of l, satisfies (14).

Thence we have deduced an expression for the total cross section, involving only a few phase shifts, viz., (17) and (18).

Further, if the energy is high ($K \gg 1$), then $\beta \approx 1 - b/K \approx 1$, and we have seen that the relation (33) is indeed useful, particularly in computing the cross sections.

In any case, (33) shows that for the energy ranges considered, the phase shifts do not fall off very rapidly, as rapidly, for instance, as Gaussian potentials, which, according to (31), fall off like the inverse factorial of l, or finite range potentials, which fall off even faster.[11,12]

We have also verified that formula (33) [or (16) or (32)] gives satisfactory results from $l = 1$ onwards, for scattering by atomic hydrogen and helium, while formula (18), for the total cross section, gives better results than an improved variational technique or the Born approximation.

ACKNOWLEDGMENTS

I would like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste. I would also like to thank Professor P. Budinich and Professor L. Fonda for inviting me to the International School for Advanced Studies, Trieste. I am very grateful to Professor G. C. Ghirardi for useful discussions in connection with this paper as also to the referee, in the light of whose constructive comments the paper has been rewritten and the practical aspects of the work have been incorporated.

[1] P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, Mass., 1965), pp. 175–9.
[2] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1962).
[3] E. T. Whittaker and G. N.Watson, *A Course of Modern Anaylsis* (Cambridge U.P., Cambridge, 1969), p. 318.
[4] N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford U.P., Oxford, 1965), pp. 571, 465.
[5] M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964), p. 258.
[6] T. L. John, Proc. Phys. Soc. **76**, 532 (1960).
[7] S. Chandrashekhar and Breen, Astrophys. J. **103**, 41 (1946).
[8] S. Altshuler, Phys. Rev. **87**, 992 (1952).
[9] Smith, Miller and Mumford, Proc. Phys. Soc. **76**, 559 (1960).
[10] F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967), pp. 67–70.
[11] B. G. Sidharth, Nuov. Cimento **46**A(3), 419 (1978).
[12] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), p. 107.

A coordinate-free derivation of a generalized geodesic deviation equation

N. S. Swaminarayan
Department of Mathematics, Chelsea College, University of London, Manresa Road, London SW3 6LX, United Kingdom
J. L. Safko
Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208

(Received 27 July 1982; accepted for publication 22 October 1982)

A simple, coordinate-free exact derivation of the geodesic deviation equation is given. This result includes the possibility of nonvanishing torsion. We then show that this form of the geodesic deviation equation can be specialized to various results given previously in the literature.

PACS numbers: 04.20.Me

I. INTRODUCTION

The motion of a test particle along a geodesic of the space-time plays the same role in the general theory of relativity as the uniform linear motion of a test particle plays in the Newtonian mechanics. The Riemann curvature tensor, which is the indicator for the presence of the gravitational field does not enter the geodesic equation; however, it does show up in the geodesic deviation equation. This equation was first formulated by Levi-Civita[1] and independently proposed by Synge[2] and Schild,[3] while considering a two-parameter family of curves. The geodesic deviation equation gives the relative acceleration between two nearby particles moving along nearly identical geodesic paths.

Pirani[4] used a parallelly propagated tetrad and put the geodesic deviation equation into a form comparable to the "Newtonian" case. In his work the tetrad components of the Riemann tensor were seen to correspond to the tidal forces due to a "Newtownian" gravitational potential in the observer's rest frame.

Manoff[5] has discussed the relation between the geodesic deviation equation and the Lie derivative of the defining vector fields. He has given a summary of various formalisms, methods and/or modifications of the geodesic equation used by several authors to investigate different physical situations.

Dolan *et al.*[6,7] have shown that the magnitude of the deviation vector satisfies a scalar equation, which can be solved to obtain formal solutions. Novello *et al.*[8] have generalized the geodesic deviation equation to the idea of a Jacobi field. Further generalizations have been given by Weber,[9] Bazanski,[10,11] and Hodgkinson,[12]

In this paper we give in a coordinate-free way a simple derivation of the geodesic deviation equation allowing for the possibility of nonvanishing torsion. The plan of the paper is as follows:

The basic equation is derived in Sec. II. In Sec. III, we show how this equation reduces to the results of several other authors. The relevance of the Lie transfer for the deviation vector is considered in Sec. IV along with the work of several other authors on the higher-order geodesic equation. Finally, in Sec. V we summarize our conclusions.

II. THE BASIC EQUATION

Let $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ be vector fields on a four-dimensional Riemannian manifold M_4 that is locally Lorentzian, and is endowed with a connection and a metric g. This connection defines the operation of covariant differentiation, ∇, of vectors and tensors on M_4. $[\mathbf{u},\mathbf{v}]$ denotes the Lie bracket of the vector fields $\mathbf{u}$ and $\mathbf{v}$. The basic definitions of the torsion vector field T and the curvature operator R are

$$T(\mathbf{u},\mathbf{v}) = \nabla_{\mathbf{u}}\mathbf{v} - \nabla_{\mathbf{v}}\mathbf{u} - [\mathbf{u},\mathbf{v}], \tag{2.1}$$

$$R(\mathbf{u},\mathbf{v})\mathbf{w} = \nabla_{\mathbf{u}}\nabla_{\mathbf{v}}\mathbf{w} - \nabla_{\mathbf{v}}\nabla_{\mathbf{u}}\mathbf{w} - \nabla_{[\mathbf{u},\mathbf{v}]}\mathbf{w}. \tag{2.2}$$

The Lie derivative of $\mathbf{v}$ via the vector field $\mathbf{u}$ is given by

$$\mathscr{L}_{\mathbf{u}}\mathbf{v} = [\mathbf{u},\mathbf{v}]. \tag{2.3}$$

Let $\mathbf{u}$ and $\boldsymbol{\eta}$ be two vector fields on M_4; then Eq. (2.2) becomes $R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = \nabla_{\boldsymbol{\eta}}\nabla_{\mathbf{u}}\mathbf{u} - \nabla_{\mathbf{u}}\nabla_{\boldsymbol{\eta}}\mathbf{u} - \nabla_{[\boldsymbol{\eta},\mathbf{u}]}\mathbf{u}$, which upon substitution of Eq. (2.1) gives

$$R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = \nabla_{\boldsymbol{\eta}}\nabla_{\mathbf{u}}\mathbf{u} - \nabla_{\mathbf{u}}\{\nabla_{\mathbf{u}}\boldsymbol{\eta} + T(\boldsymbol{\eta},\mathbf{u}) + [\boldsymbol{\eta},\mathbf{u}]\} - \nabla_{[\boldsymbol{\eta},\mathbf{u}]}\mathbf{u}.$$

The last equation can be written as

$$\nabla_{\mathbf{u}}^2\boldsymbol{\eta} + R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = \nabla_{\boldsymbol{\eta}}(\nabla_{\mathbf{u}}\mathbf{u}) + \nabla_{\mathbf{u}}T(\mathbf{u},\boldsymbol{\eta}) + \nabla_{\mathbf{u}}[\mathbf{u},\boldsymbol{\eta}] + \nabla_{[\mathbf{u},\boldsymbol{\eta}]}\mathbf{u}. \tag{2.4}$$

Equation (2.4) is a direct consequence of the definitions of the curvature and the torsion. No other ideas or calculations are needed to obtain Eq. (2.4). We will refer to it as the basic equation.

Consider a congruence of curves, which are parametrized by the same parameter s, at least on a small compact domain.[13] Each curve of this congruence is a possible worldline for an observer or a test particle. The tangent vector field to the congruence is denoted by $\mathbf{u}$, and $\boldsymbol{\eta}$ denotes the deviation vector field, which connects points on two infinitesimally neighboring curves of the congruence with equal values of the parameter s. The parameter along an integral curve for $\boldsymbol{\eta}$ will be denoted by v. With this specification Eq. (2.4) is a generalized deviation equation. It includes the possibility of nonvanishing torsion.

0022-2488/83/040883-03$02.50

III. SPECIAL CASES

An observer is following a curve Γ, a member of the congruence in M_4. The velocity vector of the curve is denoted by the unit tangent vector **u**, to the curve Γ. The acceleration of the curve $\nabla_{\mathbf{u}}\mathbf{u}$ is denoted by **a**.

$$\mathbf{u}\cdot\mathbf{u} = 1 \tag{3.1a}$$

and

$$\mathbf{a} = \nabla_{\mathbf{u}}\mathbf{u}. \tag{3.1b}$$

Except when otherwise stated, it will be assumed that the vector $\boldsymbol{\eta}$ is carried along Γ by Lie transport, i.e.,

$$\mathscr{L}_{\mathbf{u}}\boldsymbol{\eta} = [\mathbf{u},\boldsymbol{\eta}] = 0. \tag{3.2}$$

We shall return to the discussion of Lie transport in Sec. IV. If we assume that torsion vanishes, the basic equation (2.4) now reads[9]

$$\nabla^2_{\mathbf{u}}\boldsymbol{\eta} + R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = \nabla_{\boldsymbol{\eta}}\mathbf{a}. \tag{3.3}$$

Case (i): $\mathbf{a} = 0$. In this case Γ is a geodesic, and the reference frame of the observer is carried along Γ by parallel transport. Equation (3.3) reduces to

$$\nabla^2_{\mathbf{u}}\boldsymbol{\eta} + R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = 0. \tag{3.4}$$

Following Matte,[14] let us set

$$E_{ab} = -R_{apbq}u^pu^q, \text{ i.e., } \underline{E}(\ ) = -R(\ ,u)u. \tag{3.5}$$

Notice that E_{ab} is trace-free in empty space, where $R_{ab} = 0$. We have

$$\frac{D^2\eta^a}{\partial s^2} = E^{a}_{\ b}\eta^b, \text{ i.e., } \nabla^2_{\mathbf{u}}\boldsymbol{\eta} = \mathbf{E}\cdot\boldsymbol{\eta}, \tag{3.6}$$

where $\mathbf{E}\cdot\boldsymbol{\eta} = \mathbf{E}(\boldsymbol{\eta})$.

Any $\boldsymbol{\eta}$ is called a Jacobi vector field along Γ, if it satisfies an equation of the type Eq. (3.6).

Case (ii): $\mathbf{a}\neq 0$. Then Γ is not a geodesic, and the Eq. (3.6) is modified to read

$$\frac{D^2\eta^a}{\partial s^2} = E^{a}_{\ b}\eta^b + \eta^p\nabla_p a^a, \text{ i.e., } \nabla^2_{\mathbf{u}}\boldsymbol{\eta} = \mathbf{E}\cdot\boldsymbol{\eta} + \nabla_{\boldsymbol{\eta}}\mathbf{a}. \tag{3.7}$$

If the nongravitational forces present are such that we may set

$$\nabla_b a^a = N^{a}_{\ b} - E^{a}_{\ b}, \tag{3.8}$$

then we get

$$\frac{D^2\eta^a}{\partial s^2} = N^{a}_{\ b}\eta^b. \tag{3.9}$$

Novello *et al.*[8] call the solutions of Eq. (3.9) a Generalized Jacob Field (GJF). They discuss possible forms of $N^{a}_{\ b}$ and consider several particular cases, where $N^{a}_{\ b}$ is a polynomial function of the curvature tensor.

Case (iii): $a\neq 0$. In this case it is more appropriate to use Fermi–Walker transport (FWT) along Γ for the reference frame of the observer. FWT assures a nonrotating set of axes along Γ and reduces to parallel transport when Γ is a geodesic. The Fermi–Walker derivative $D_F/\partial s$ along Γ for a vector field **X** is defined by

$$\frac{D_F\mathbf{X}}{\partial s} = \frac{D\mathbf{X}}{\partial s} + (\mathbf{X}\cdot\mathbf{a})\mathbf{u} - (\mathbf{X}\cdot\mathbf{u})\mathbf{a}. \tag{3.10}$$

Notice that

$$\frac{D_F\mathbf{u}}{\partial s} = 0, \tag{3.10a}$$

and

$$\frac{D_F X_\perp}{\partial s} = \left(\frac{DX_\perp}{\partial s}\right)_\perp, \tag{3.10b}$$

where $\perp$ denotes the part of the vector perpendicular to u. Any vector **X** can be split up into two parts,

$$\mathbf{X} = \mathbf{X}_\perp + (\mathbf{X}\cdot\mathbf{u})\mathbf{u}, \tag{3.11}$$

where $\mathbf{X}_\perp\cdot\mathbf{u} = 0$. The perpendicular component can be calculated by the use of a projection operator;
$h_{ab} = g_{ab} - u_a u_b$.
A straightforward calculation shows that Eq. (3.3) becomes

$$\frac{D^2_F\eta^a_\perp}{\partial s^2} = -R^{\ a}_{bcd}u^bu^d\eta^c_\perp + h^{\ a}_{b}\eta^c_\perp\nabla_c a^b - a_b\eta^b_\perp a^a. \tag{3.12}$$

The last equation may be expressed in either of the following ways:

$$\frac{D^2_F\boldsymbol{\eta}_\perp}{\partial s^2} = -R(\boldsymbol{\eta}_\perp,\mathbf{u})\mathbf{u} + (\nabla_{\boldsymbol{\eta}_\perp}\mathbf{a})_\perp - (\mathbf{a}\cdot\boldsymbol{\eta}_\perp)\mathbf{a} \tag{3.12a}$$

$$= -R(\boldsymbol{\eta}_\perp,\mathbf{u})\mathbf{u} + \left(\frac{D\mathbf{a}}{\partial v}\right)_\perp - (\mathbf{u}\cdot\boldsymbol{\eta})\frac{D_F\mathbf{a}}{\partial s} - (\mathbf{a}\cdot\boldsymbol{\eta}_\perp)\mathbf{a}, \tag{3.12b}$$

where $D/\partial v = \eta^a\nabla_a$.

$\boldsymbol{\eta}$ can be interpreted as the separation between two neighboring test particles, as measured in the rest frame of one of them, and the second FW derivative of $\boldsymbol{\eta}$ gives the relative acceleration between the two test particles. Thus use of FWT is fully justified and the equations (3.12) give an appropriate generalization of the geodesic deviation equation for accelerated curves. Since Eq. (3.10b) implies that

$$\frac{D^2_F\boldsymbol{\eta}_\perp}{\partial s^2} = \left(\frac{D}{\partial s}\left(\frac{D}{\partial s}\boldsymbol{\eta}_\perp\right)_\perp\right)_\perp,$$

these results are equivalent to those proposed by Hawking and Ellis.[15]

Case (iv): $\mathbf{a}\neq 0$. $\mathscr{L}_{\mathbf{u}}\boldsymbol{\eta}\neq 0$. In this case we give up the assumption that $\boldsymbol{\eta}$ is Lie transported along Γ, and use the notation of Lie derivative. The vanishing of the torsion implies that

$$\nabla_{\boldsymbol{\eta}}\mathbf{a} - \nabla_{\mathbf{a}}\boldsymbol{\eta} = [\boldsymbol{\eta},\mathbf{a}] = \mathscr{L}_{\boldsymbol{\eta}}\mathbf{a},$$

i.e.,

$$\nabla_{\boldsymbol{\eta}}\mathbf{a} = \mathscr{L}_{\boldsymbol{\eta}}\mathbf{a} + \nabla_{\mathbf{a}}\boldsymbol{\eta}.$$

The basic equation can be written as

$$\nabla^2_{\mathbf{u}}\boldsymbol{\eta} + R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = \nabla_{\boldsymbol{\eta}}\mathbf{a} - \nabla_{\mathbf{u}}(\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u}) - \nabla_{\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u}}\mathbf{u}, \tag{3.13}$$

which becomes

$$\nabla^2_{\mathbf{u}}\boldsymbol{\eta} + R(\boldsymbol{\eta},\mathbf{u})\mathbf{u} = \nabla_{\mathbf{a}}\boldsymbol{\eta} + \mathscr{L}_{\boldsymbol{\eta}}\mathbf{a} - \nabla_{\mathbf{u}}(\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u}) - \nabla_{\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u}}\mathbf{u}. \tag{3.14}$$

This is the coordinate-free representation of "the generalized deviation equation" of Manoff[5] (his Eq. 2.6). He has considered several cases by choosing $\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u}$ equal to suitable (different) vectors in some detail.

IV. LIE TRANSPORT

The statement that $\boldsymbol{\eta}$ is Lie transported along Γ is obviously equivalent to saying that $\mathbf{u}$ is Lie transported along the integral curve of $\boldsymbol{\eta}$. The derivation of

$$\mathscr{L}_{\mathbf{u}}\boldsymbol{\eta} = [\mathbf{u},\boldsymbol{\eta}] = 0, \tag{4.1}$$

given by Kilmister,[16] clearly shows that (4.1) is a covariant generalization of a frame-dependent equation, obtained by assuming $\boldsymbol{\eta}$ to be small and neglecting terms involving the products of $\boldsymbol{\eta}$ with itself. This together with the assumption about the parametrization of the curves of the congruence by the same parameter s restricts the use of the standard geodesic equation to the case where $\boldsymbol{\eta}$ and its rate of change are small.

This suggests that to generalize the geodesic deviation equation to the case where $\boldsymbol{\eta}$ and/or its rate of change are not necessarily small, we should consider the cases where $\mathscr{L}_{\mathbf{u}}\boldsymbol{\eta}$ is nonzero. This also implies an investigation into the relationship of parameters along two curves of the congruence. This has been done by Hodgkinson[12] and Bazanski.[10,11] A coordinate-independent formulation remains to be discussed.

Another way to consider this is that $\boldsymbol{\eta}$ and $\mathbf{u}$ are usually considered as belonging to the tangent space at each point. Physically we are interested in particles which have finite separation in the Riemannian space. Thus, when the lowest order is not sufficient it will be necessary to look at the relations among the successive tangent spaces to the manifold.

We must indicate a possible error in Manoff's[5] paper, where he considers the reduction of his generalized equation to that due to Hawking and Ellis[15] (EH equation). Since the EH equation can be obtained from Eq. (3.3), $\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u}$ necessarily vanishes. However, Manoff (his case 6 in Sec. 3.1) obtains the EH equation under the assumption $\mathscr{L}_{\boldsymbol{\eta}}\mathbf{u} = (\nabla_{\mathbf{u}}(\mathbf{u}\cdot\boldsymbol{\eta}))\mathbf{u}$. The supplementary condition [Manoff's Eq. (3.19)], viz., $\mathbf{a}\cdot\boldsymbol{\eta} = 0$, is also difficult to understand in the context of this case.

V. SUMMARY

With Eq. (2.4), we have given a simple coordinate-free derivation for a generalized geodesic deviation equation, which would allow the possibility of nonvanishing torsion. Equation (3.3) provides the most common specialization to either geodesic motion or motion with external (nongravitational) forces. The most appropriate generalization of the geodesic equation to nongeodesic motion is given by either Eq. (3.12) or Eq. (3.14).

Clearly further work is needed since the coordinate-free formulation of the deviation equation has not been done for $D\mathbf{u}/\partial v$ or $D\boldsymbol{\eta}/\partial s$ nonnegligible. This will be the subject of future work.

ACKNOWLEDGMENTS

One of us (N.S.S.) wishes to thank the Department of Physics and Astronomy, University of South Carolina for hospitality, facilities, and support afforded during his stay in Columbia. Part of the work was done at the Auburn University, Alabama, where he spent the academic year 1981–82 as a visiting Associate Professor.

[1] T. Levi-Civita, Math. Ann. **97**, 129 (1929).
[2] J. L. Synge, *Relativity: The General Theory* (North-Holland, Amsterdam, 1960), p. 20.
[3] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto, Toronto, 1952).
[4] F. A. E. Pirani, Acta Phys. Polon. **15**, 389 (1956).
[5] S. Manoff, Gen. Relativ. Gravit. **11**, 189–204 (1979).
[6] P. Dolan, P. Choudhury, and J. L. Safko, J. Aust. Math. Soc. Ser. B **22**, 28–33 (1980).
[7] P. Dolan, P. Choudhury, and N. S. Swaminarayan (unpublished).
[8] M. Novello, I. D. Soares, and J. M. Salim, Gen. Relativ. Gravit. **8**, 95–102 (1977).
[9] J. Weber, *General Relativity and Gravitational Waves* (Interscience, New York, 1961), Chap. 8.
[10] S. L. Bazanski, Ann. Inst. Henri Poincaré **27**, 115–144 (1977).
[11] S. L. Bazanski, Ann. Inst. Henri Poincaré **27**, 145–166 (1977).
[12] D. E. Hodgkinson, Gen. Relativ. Gravit. **3**, 351–375 (1972).
[13] G. F. R. Ellis, "Relativistic Cosmology" in *General Relativity and Cosmology: Redocti S. J. F.*, edited by R. Sachs (Academic, New York, 1971), Vol. XLVIII.
[14] A. Matte, Can. J. Math. **5**, 1–16 (1953).
[15] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U. P., Cambridge, 1974), p. 81.
[16] C. W. Kilmister, *General Theory of Relativity* (Pergamon, New York, 1973), p. 21.

Geometrical interpretation of a generalized theory of gravitation

G. Kunstatter and J. W. Moffat
Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada
J. Malzan
Department of Mathematics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

(Received 27 July 1982; accepted for publication 29 October 1982)

The geometrical structure is developed for a theory of gravitation, based on a nonsymmetric metric in a four-dimensional real manifold. The local fiber bundle gauge group is GL(4,*R*), which contains the (local) homogeneous Lorentz gauge group SO(3,1) of general relativity.

PACS numbers: 04.50. + h, 02.40.Ky

I. INTRODUCTION

The general theory of relativity is based on the view that the geometrical structure of space-time be considered a dynamical quantity. The metric evolves according to Einstein's equations. This metric is taken to be a real bilinear form which acts on the tangent space at each point of the space-time manifold. In this paper, we will examine the consequences of enlarging the tangent bundle in a particularly simple way to introduce additional geometrical structure. The procedure is completely analogous to the complexification of the tangent bundle.[1] Instead of allowing the functions to take their values in the field of complex numbers, which leads to complexification, we shall assume that they take their values in the ring of hyperbolic complex or "double" numbers. Such numbers have been studied extensively by mathematicians.[2-4] Here we will show that this procedure leads naturally to the introduction of a sesquilinear inner product which can then be used as the dynamical basis for a generalized theory of gravitation.[5,6] We will demonstrate that in addition to the gauge structure associated with the diffeomorphism group of the manifold, such a theory is invariant under an internal GL(4,*R*) symmetry which is associated with a change of frames in the *extended* tangent bundle.

In Sec. II, we briefly review the properties of hyperbolic complex numbers, and show how to extend the tangent bundle. In Sec. III, we introduce a metric and connection as the dynamical geometrical quantities. Finally in Sec. IV, we discuss the physical significance of the construction and present some conclusions.

II. HYPERBOLIC COMPLEX NUMBERS

It is known[2-4] that the field of complex numbers can be generalized in a natural way. These generalized complex numbers are of the form

$$a + \epsilon b , \tag{1}$$

where ϵ is a "number of a special kind" which obeys $\epsilon^2 = -1$ for ordinary complex numbers, $\epsilon^2 = 0$ for dual numbers and $\epsilon^2 = +1$ for double or hyperbolic complex numbers. Addition, subtraction and multiplication can be carried out in the usual way, although division requires special attention in the hyperbolic case, where the norm of a nonzero number

$$|a + \epsilon b | = (a^2 - b^2)^{1/2} , \tag{2}$$

can be zero. It is for this reason that hyperbolic complex numbers constitute a ring, but not a field. Nonetheless it is possible to generalize the concepts of holomorphism, analyticity, etc., in order to do analysis with hyperbolic complex numbers.

We wish to analyze the consequences of allowing the functions on a real four-dimensional manifold M to take their values in the ring of hyperbolic complex numbers. The physical motivation for this will be discussed in the last section. Had we chosen to use ordinary complex numbers, a complexification of the tangent bundle would have been the result. The consequences of this alternative have been discussed previously.[1] Given hyperbolic complex-valued functions of the form

$$f^H(x) = f^R(x) + \epsilon f^I(x) , \quad x \in M , \tag{3}$$

where $f^R(x)$ and $f^I(x)$ are real-valued functions on M, it is natural to define vectors A^H which take their values in the ring of hyperbolic complex numbers as well. These vectors map $f^H(x)$ into $(A^H f^H)(x)$ and can be interpreted in the usual way in terms of directional derivatives of hyperbolic complex-valued functions on M. To accommodate such vectors, it is necessary to enlarge the tangent space T_x at each point $x \in M$. We will work with a real representation for the hyperbolic complex numbers. In particular, we define an eight-dimensional vector space

$$T'_x = T_x \times T_x \tag{4}$$

so that elements of T'_x are ordered pairs of vectors (X,Y) such that $X,Y \in T_x$, with the multiplication law

$$(X,Y) := (\lambda X, \lambda Y) \tag{5}$$

and the addition law

$$(X,Y) + (X',Y') = (X + X', Y + Y') . \tag{6}$$

In general, the entire group GL(8,*R*) can act on T'_x, so that we can construct an associated GL(8,*R*) bundle $\mathscr{T}'(M)$ over M with typical fiber $T'_x \sim R^8$ by patching together direct products of the form $T'_x \times U_x$, where U_x is some neighborhood of M. Actually, we are constructing the fiber bundle $L'(M)$ associated to the frame bundle $L(M)$ with fiber GL(8,*R*) (see Kobayashi and Nomizu[7]):

$$L'(M) = L(M) X_{\mathrm{GL}(4,R)} \mathrm{GL}(8,R) , \tag{7}$$

where GL(4,R) has a natural (subgroup) right action on GL(8,R). $L'(M)$ can also be considered the bundle space for a principal GL(8,R) bundle. The bundle $\mathscr{T}'(M)$ is then a bundle with fiber R^8 associated to $L'(M)$ treated as a principal GL(8,R) bundle:

$$T'(M) = L'(M)X_{\mathrm{GL}(8,R)}R^8 . \tag{8}$$

Since we wish to consider T'_x as a real eight-dimensional representation for the space of hyperbolic complex-valued vectors, we must introduce the analog of a complex structure, i.e., hyperbolic complex structure E, say, and restrict the structure group of $\mathscr{T}'(M)$ to those transformations which preserve E. In the ordinary complex case, we introduce J such that $J^2 = -1$, and reduce the structure group to GL(4,ℂ). In order to obtain double numbers, we require

$$E^2 = +1 . \tag{9}$$

It is straightforward to show that the subgroup of GL(8,R) which preserves E is isomorphic to GL(4,R) ⊗ GL (4,R). In particular, let us take

$$E = \begin{pmatrix} 0 & I \\ I & 0 \end{pmatrix}, \tag{10}$$

where I is the four-dimensional unit matrix. Then the most general $M \epsilon \mathrm{GL}(8,R)$ such that

$$ME = EM \tag{11}$$

is of the form

$$M = \begin{pmatrix} A & B \\ B & A \end{pmatrix}, \tag{12}$$

where A,B are arbitrary 4×4 real matrices [i.e., elements of GL(4,R)]. Let us now choose the basis of T'_x in which E takes the form given in Eq. (10):

$$\begin{aligned} &\{e'_A\} = \{e'_\alpha, e'_{\bar\alpha}\}\,, \quad A = 1,\ldots,8\,, \\ &\alpha = 1,2,3,4\,, \quad \bar\alpha = \alpha + 4 \\ &\qquad\qquad\qquad\quad = 5,6,7,8\,. \end{aligned} \tag{13}$$

Thus

$$\begin{aligned} Ee'_\alpha &= e'_{\bar\alpha}\,, \\ Ee'_{\bar\alpha} &= e'_\alpha\,. \end{aligned} \tag{14}$$

In this basis, there exists a canonical choice for a *real* four-dimensional subspace; namely, the subspace spanned by $\{e'_\alpha, \alpha = 1,\ldots,4\}$. Vectors lying in the orthogonal subspace $\{e'_{\bar\alpha}, \alpha = 1,\ldots,4\}$ are then purely hyperbolic imaginary vectors. In other words, any hyperbolic complex-valued vector $A' \epsilon T_x$ can be decomposed as

$$A' = A^A e'_A = A^\alpha e'_\alpha + A^{\bar\alpha} e'_{\bar\alpha}\,, \tag{15}$$

so that in a four-dimensional representation, we would have

$$A = (A'^\alpha + \epsilon A'^{\bar\alpha})e_\alpha\,, \tag{16}$$

where $\{e_\alpha\}$ spans T_x. Note that the analog of complex conjugation also exists. The conjugation operator in our case (in the eight-dimensional real representation) takes the form

$$C = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \tag{17}$$

so that $\widetilde{A}' := CA' = A^\alpha e_\alpha - A^{\bar\alpha} e_{\bar\alpha}$, and in the four-dimensional hyperbolic complex representation, this corresponds to $\epsilon \to -\epsilon$:

$$\widetilde{A} = (A^\alpha - \epsilon A^{\bar\alpha})e_\alpha\,. \tag{18}$$

In the next section, we shall introduce the affine and metric structure which will serve as the *dynamical* foundation for a generalization of general relativity. The hyperbolic complex structure is fixed *a priori* and is not dynamical.

III. GEOMETRICAL STRUCTURE

It is possible to define an inner product g' on each eight-dimensional vector space T'_x. Naturally it must be symmetric:

$$g'(A',B') = g'(B',A') \quad \forall A',\ B' \epsilon T'_x\,, \tag{19}$$

but we do not as yet specify a signature. In component form we have

$$g'_{AB} = g'(e'_A, e'_B) = g'_{BA}\,, \quad A,B = 1,\ldots,8\,. \tag{20}$$

So far g' is independent of the hypercomplex structure E. We now impose the compatibility condition

$$g'(EA',EB') = -g'(A',B')\,, \tag{21}$$

which, as we shall see, leads to the sesquilinear, hyperbolic complex-valued metric we seek. Note the minus sign on the right-hand side of Eq. (21). For ordinary complex structure, this would be positive. In component form, Eq. (21) implies

$$g'_{\alpha\beta} = -g'_{\bar\alpha\bar\beta}\,, \tag{22a}$$

$$g'_{\alpha\bar\beta} = -g'_{\bar\alpha\beta}\,. \tag{22b}$$

Consequently, we have

$$g'(A',B') = g_{\alpha\beta}(A^\alpha B^\beta - A^{\bar\alpha}B^{\bar\beta}) + g_{\bar\alpha\beta}(A^{\bar\alpha}B^\beta - A^\alpha B^{\bar\beta})\,. \tag{23}$$

In addition to the symmetric bilinear form $g'(A',B')$, we can also define a symplectic form on T'_x, namely

$$g'(EA',B') = -g'(EB',A')\,, \tag{24}$$

where the antisymmetry follows from Eqs. (19) and (21). Thus we have the following hyperbolic complex-valued, sesquilinear form:

$$\begin{aligned} g(A,B) &= g'(A',B') + \epsilon g'(EA',B') \\ &= g_{\mu\nu}A^\mu \widetilde{B}^\nu\,, \end{aligned} \tag{25}$$

where

$$A^\mu = A'^\mu + \epsilon A'^{\bar\mu}\,, \tag{26a}$$

$$\widetilde{B}^\mu = B'^\mu - \epsilon B'^{\bar\mu}\,, \tag{26b}$$

and

$$g_{\mu\nu} = g'_{\mu\nu} + \epsilon g'_{\bar\mu\nu} \tag{27}$$

so that

$$g_{(\mu\nu)} = g'_{\mu\nu} \tag{28a}$$

and

$$g_{[\mu\nu]} = g'_{\bar\mu\nu}\,. \tag{28b}$$

In Eqs. (28a) and (28b) we use the standard notation $g_{(\mu\nu)} = \frac{1}{2}(g_{\mu\nu} + g_{\nu\mu})$ and $g_{[\mu\nu]} = \frac{1}{2}(g_{\mu\nu} - g_{\nu\mu})$.

In general relativity, the introduction of a metric reduces the structure group of the frame bundle from GL(4,R)

to the group of transformations which preserves the metric, namely, SO(3,1). In the present case, the subgroup of GL(4,*R*) ⊗ GL(4,*R*) which preserves the metric is isomorphic to GL(4,*R*). Thus for any given sesquilinear hyperbolic complex-valued metric, there exists an *internal* GL(4,*R*) gauge symmetry (analogous to local Lorentz invariance in general relativity), which corresponds to "orthogonal" (i.e., metric preserving) rotations of the generalized frames. It is remarkable that GL(4,*R*) appears as an *internal* symmetry group, which is not *a priori* related to coordinate transformations of the underlying space-time manifold.

In order to make the geometrical structure truly dynamical (i.e., so that we can construct a nontrivial Lagrangian containing derivatives), it is necessary to define covariant differentiation of the hyperbolic complex-valued vectors. That is, we need to define a connection which tells us how to parallel transport a vector in the fiber T'_x over $x \in M$ to the fiber $T'_{x'}$ over $x' \in M$. Since M is a four-dimensional real manifold, $(x' - x)$ locally defines a vector in T_x. Thus, the covariant derivative operator ∇ maps $T_x \times T'_x$ into T'_x:

$$\nabla : (X, A') \mapsto \nabla_X A' = A'^A(x) \nabla_X e'_A + (X A'^A(x)) e'_A \,. \tag{29}$$

Since $A'^A(x)$ is a real-valued function on M and X is a real vector, $(XA'^A)(x)$ is well defined:

$$XA^A = X^\mu \frac{\partial}{\partial x^\mu} A^A(x) \tag{30}$$

in a coordinate basis for T_x. It remains only to define

$$\nabla_X e'_A := X^\mu \Gamma^B_{\mu A} e'_B \,, \tag{31}$$

where $\Gamma^B_{\mu A}$ is the desired connection. Note that $\mu = 1,\ldots,4$; A,B = 1,...,8 so that the connection has $8^2 \times 4$ independent components. We now restrict the connection to be compatible with the hyperbolic complex structure E introduced earlier. Namely, we require $(\nabla E) = 0$, so that

$$\nabla_X(EA') = E(\nabla_X A') \,, \tag{32}$$

which in the basis of Eq. (31) yields the restriction

$$\Gamma^\alpha_{\mu\beta} = \Gamma^{\bar\alpha}_{\mu\bar\beta} \,, \tag{33a}$$

$$\Gamma^\alpha_{\mu\bar\beta} = \Gamma^{\bar\alpha}_{\mu\beta} \,. \tag{33b}$$

This is in fact a very strong restriction on the connection; it requires that the eigenspaces of the hyperbolic complex structure operator E be preserved under parallel transport. Of course, the real and hyperbolic complex imaginary subspaces (which are eigenspaces of the conjugation operator C) are not separately preserved. Thus, a pure real vector will in general become hyperbolic complex-valued under parallel transport. The condition (32) ensures that Γ defines a connection on the bundle $L'(M)$ of hyperbolic complex-valued frames; it now has only 2×4^3 independent components. In a four-dimensional hyperbolic complex-valued representation we have

$$W^\alpha_{\mu\beta} = \Gamma'^\alpha_{\mu\beta} + \epsilon \Gamma'^{\bar\alpha}_{\mu\beta} \,. \tag{34}$$

The metric and the connection are as yet independent geometrical objects defined on M. We can further restrict them in the standard way, by requiring the lengths of vectors to be preserved under parallel transport. The condition $(\nabla g') = 0$ can be written

$$\nabla_X(g'(A', B')) = g'(\nabla_X A', B') + g'(A', \nabla_X B') \,. \tag{35}$$

In component form, Eq. (35) gives the following condition on the metric:

$$\frac{\partial}{\partial x^\mu} g'_{AB} = \Gamma'^C_{\mu A} g'_{CB} + \Gamma'^C_{\mu B} g'_{AC} \,. \tag{36}$$

This can be reduced to a more familiar form[1] by expressing the metric and connection in terms of their hyperbolic complex-valued components $g_{\mu\nu}, W^\alpha_{\mu\beta}$:

$$\frac{\partial}{\partial x^\rho} g_{\mu\nu} - g_{\epsilon\nu} W^\epsilon_{\mu\rho} - g_{\mu\epsilon} \widetilde{W}^\epsilon_{\nu\rho} = 0 \,, \tag{37}$$

where $\widetilde{W}^\epsilon_{\nu\rho} = \Gamma'^\epsilon_{\nu\rho} - \epsilon \Gamma'^\epsilon_{\nu\rho}$ is the conjugate of $W^\epsilon_{\nu\rho}$, and $g_{\mu\nu}$ is defined in Eq. (27).

IV. DISCUSSION AND CONCLUSIONS

We have investigated the fiber bundle gauge structure which results when the tangent bundle of a real four-dimensional manifold is extended to admit hyperbolic complex-valued vectors. A sesquilinear (nonsymmetric) metric can be defined in terms of a real, bilinear form in an eight-dimensional representation for the extended tangent space over each point of M. The "hypercomplexified" tangent bundle defines a fiber bundle with typical fiber R^8 associated to the principal bundle $L'(M)$ of hyperbolic complex-valued frames, which has GL(4,*R*) ⊗ GL(4,*R*) as a structure group. The structure group of the bundle of hyperbolic complex-valued frames is reduced from GL(4,*R*) ⊗ GL(4,*R*) to GL(4,*R*) by the introduction of the metric g'. Technically, this metric is a cross section of the fiber bundle E(M,GL(4,*R*) ⊗ GL(4,*R*)/GL(4,*R*), GL(4,*R*) ⊗ GL(4,*R*),$L'(M)$) associated to $L'(M)$ (we use the notation of Kobayashi and Nomizu[7]).

A theory of gravitation has been developed using a nonsymmetric metric $g_{\mu\nu} = g'_{(\mu\nu)} + \epsilon g'_{[\mu\nu]}$, and $\Gamma^\lambda_{\mu\nu} = \Gamma'^\lambda_{\mu\nu} + \epsilon \Gamma'^\lambda_{[\mu\nu]}$, where $g'_{\mu\nu}$ and $\Gamma'^\lambda_{\mu\nu}$ are real valued, but $\epsilon^2 = -1$ $(+1)$ is the generator of ordinary complex (hyperbolic complex) numbers, respectively. When $\epsilon^2 = +1$, corresponding to the use of hyperbolic complex numbers, the theory is found to be free of ghost poles.[6,8] The physical interpretation of the theory is based on interpreting ϵ as a fermion number "charge," so that hyperbolic complex conjugation $\tilde{g}_{\mu\nu} = g'_{(\mu\nu)} - \epsilon g'_{[\mu\nu]} = g_{\nu\mu}$ corresponds to the fermion number conjugation operation which turns a fermion into an antifermion. In fact, a conserved Noether current does exist in the theory and the generator of the associated symmetry operation is proportional to ϵ. Thus the theory leads to a fundamental explanation of the stability of fermionic matter, provided that the conserved charge associated with the Noether current is identified with fermion number.[6] In other words, fermion number and its conservation are provided with a geometrical interpretation in the theory, which relates to the additional internal degrees of freedom of the extended frame bundle.

ACKNOWLEDGMENTS

We would like to thank Professor B. Kostant and Dr. M. Kalinowski for helpful discussions. One of us (G.K.)

would like to thank the Natural Sciences and Engineering Research Council of Canada for a University Research Fellowship. This work was supported by the Natural Sciences and Engineering Research Council of Canada.

---

[1] G. Kunstatter and R. Yates, J. Phys. A **14**, 847 (1981).

[2] I. M. Yaglom, *Complex Numbers in Geometry* (Academic, London, 1968).

[3] I. L. Kantor and A. S. Solodovnikov, *Hypercomplex Numbers* (in Russian) (Nauka, Moscow, 1973).

[4] B. A. Rosenfeld, *Noneuclidean Geometries* (in Russian) (Gostekisdat, Moscow, 1955).

[5] J. W. Moffat, Phys. Rev. D **19**, 3554 (1979).

[6] J. W. Moffat, Lectures given at the VIIth International School of Gravitation and Cosmology, Erice, 1981 (Proceedings to be published by World Scientific Publishing Co., Singapore, 1982).

[7] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1964), Vol. I.

[8] R. B. Mann and J. W. Moffat, Phys. Rev. D **26**, 1858 (1982).

Local symmetries in extended conformal gravity and conformal supergravity in superspace

Freydoon Mansouri
Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221

(Received 27 October 1982; accepted for publication 12 November 1982)

The extended conformal field theory of gravity is discussed, and its local symmetries are compared to those of Einstein and standard Weyl theories. It is then shown that the local symmetries of the extended conformal geometry are also natural for the construction of SU(N)-extended conformal supergravity theories in superspace. The natural superspace for the formulation of such theories is one with two sets of four-component fermionic variables which transform oppositely under scale transformations.

PACS numbers: 04.50. + h, 11.30.Pb, 11.10.Lm, 11.30.Ly

I. INTRODUCTION

Local space–time symmetries differ from internal symmetries in a number of ways. Perhaps the most significant difference is that local space–time symmetries are not arbitrary: whereas in, e.g., a grand unified theory the choice of a guage group is limited not by any known principle but by its *a posteriori* phenomenological implications, the very existence of a four-dimensional space–time imposes an upper limit to the extent of local space–time symmetries. For example, in an n-dimensional (psuedo) Riemannian manifold, the number of allowed one-parameter groups of conformal transformations is at most $(n + 1)(n + 2)/2$.[1] Moreover, this maximal group of conformal transformations is realized if, and only if, the corresponding manifold is conformally flat. In other words, if the space–time manifold of interest is not conformally flat, its local space–time symmetries *must* be *smaller* than the maximal group.

In four dimensions, $(n + 1)(n + 2)/2 = 15$, and the corresponding maximal group is SO(4,2)$\cong$SU(2,2). Thus in a conformally flat 4-manifold such as Minkowski space, i.e., one for which the components $C_{\mu\nu\rho}{}^{\lambda}$ of Weyl's curvature tensor vanish, SO(4,2) transformations can be realized as space–time symmetries. However, if the 4-manifold possesses *nontrivial* curvature, then the corresponding group of conformal transformations is, according to the above theorem, smaller than SO(4,2). Moreover, as part of the group of general coordinate transformations these (angle-preserving) transformations have to do with reparametrization invariance of the intrinsic properties of a manifold, and no gauge fields are associated with them. Gauge fields (connection) are, instead, associated with the *structure group* of a manifold. For a four-dimensional space–time manifold, the largest structure group is GL($4R$) or its affine extension IGL($4R$). These transformations are related to the freedom in choosing a basis at any point of the space–time manifold, and their algebra specifies the connection associated with manifold. So the question we must deal with is: What subgroup of GL($4R$) or IGL($4R$) transformations are conformal (angle-preserving)? Clearly GL($4R$) does not contain SO(4,2). In fact the largest conformal subgroup of GL($4R$) is SO(3,1) $\otimes$ SO(1,1), where SO(3,1) is the homogeneous Lorentz group and SO(1,1) is the one-parameter group of scale transformations. The only other freedom left is to work with IGL($4R$), in which case four more commuting generators become available.

There are two celebrated theories of gravitation which satisfy this requirement: Einsten's theory[2] with local space–time symmetry SO(3,1) and Weyl's theory[2] with local symmetry SO(3,1) $\otimes$ SO(1,1). In fact, one may regard[3] Weyl's theory as modification of Einstein's theory in which the local space–time symmetry is increased from SO(3,1) to SO(3,1) $\otimes$ SO(1,1), where SO(1,1) is the one-parameter group of scale transformations. It has recently been pointed out[4] that, consistent with the above stringent requirement, this local symmetry can be increased even further to $T'_4{}_\wedge$ SO(3,1) $\otimes$ SO(1,1), where the abelian group T'_4 is isomorphic to ordinary translations. One of the main objectives of this work is to describe in more detail how this can be carried out consistently.

The structure of local space–time symmetries is not only important for the construction of a pure gravity theory, it also specifies the basic group with respect to which the matter field multiplets which couple to gravity must transform. Thus in Einstein's theory matter field multiplets are linear representations of the Lorentz group SO(3,1). Since, in supergravity theories, gravity is coupled to matter fields, one can then ask whether the increase in the local space–time symmetry is compatible with the algebraic structure of conformal supergravity. The positive answer to this question will be one of the new results present in this paper. Another by-product of this investigation is that whereas the fermionic coordinates of the superspace are often taken to be dotted and undotted (two-component) spinors, the natural superspace from the point of view of the extended conformal symmetry is one with two sets of four-component spinors which transform oppositely under scale transformations. As a result, the structure of scale invariant actions in SU(N)-extended conformal supergravity in superspace become independent of N.

The plan of this paper is as follows: In Sec. II the extended conformal geometry is discussed and its consistency is demonstrated. This is mainly an elaboration of the results reported in Ref. 4. In Sec. III this analysis is applied to conformal supergravity in extended superspace. Section IV is devoted to a discussion of the results and concluding remarks.

0022-2488/83/040890-07$02.50

II. THE EXTENDED CONFORMAL GEOMETRY

A. The method

It will be recalled that Einstein's and Weyl's theories, in addition to the invariance under their respective local symmetries, are invariant under general coordinate transformations. A local symmetry such as SO(3,1) or SO(1,1) is a *linear* symmetry, which can be formulated as a Yang–Mills theory[5] in which the transformation laws of the gauge fields are linear and inhomogeneous:

$$B_\mu \rightarrow U B_\mu U^{-1} + (1/e) U \partial_\mu U^{-1}. \tag{2.1}$$

But, by their very definition, general coordinate transformations are not linear and cannot be represented directly in the form (2.1). Therefore, in trying to recast gravitation as a theory based on a local gauge principle, one must either treat general coordinate transformations separately as having nothing to do with (2.1), or to supplement (2.1) with the required conditions so that *all* the invariances could be based on a local gauge principle. Of course, as far as the gravitational field equations and their solutions are concerned, the choice between these alternatives is immaterial. But if the formulation of gravity as a gauge theory is regarded as a step toward its inclusion in a unified gauge theory, then the second alternative is more desirable.

It was with this application in mind that the method of nonlinear realizations of a gauge symmetry was developed.[6] In this approach one regards the general coordinate transformations as a four-parameter group of transformations with generators $\{X_i\}$, $i = 0,1,2,3$, *as if* these generators belonged to an ordinary Lie algebra. For example, in Einstein's theory in which the linear gauge symmetry is SO(3,1), one can start with a set of ten generators $\{X_{ij}, X_i\}$ and in analogy with Yang–Mills theory consider a quantity $\widehat{D}_\mu$ with values in this algebra:

$$\begin{aligned} \widehat{D}_\mu &= \partial_\mu + H_\mu^{\,ij} X_{ij} - K_\mu^{\,j} X_j \\ &\equiv D_\mu - K_\mu^{\,i} X_i = D_\mu - K_\mu . \end{aligned} \tag{2.2}$$

If we require that $\widehat{D}_\mu$ as a whole determine the covariance under the local gauge transformations generated by $\{X_{ij}, X_i\}$, then $\widehat{D}_\mu$ is the usual expression for the covariant derivative in Yang–Mills theory, and the resulting theory is a linear realization of a gauge symmetry. Alternatively, if we require that, in (2.2), D_μ and K_μ be separately covariant under the transformations generated by $\{X_{ij}, X_i\}$, then the resulting theories are nonlinear realizations of a local symmetry. The proof that in the second case one obtains nonlinear realizations is quite simple: Yang–Mills solution contains all the linear realizations, so that any other realization, if it exists, must be nonlinear. The existence of other possibilities can then be shown by explicit construction (see below).

Nonlinear realizations are distinguished by the nature of the constraints which make them nonlinear. In general relativity, the constraint must be such that the transformations generated by $\{X_i\}$ represent general coordinate transformations. The correct *differential* constraint for this nonlinear realization turns out to be[6]

$$\widehat{D}_\mu = 0. \tag{2.3}$$

For then one can solve for K_μ and regard the equality as a change of basis in a manifold with structure group SL(2,C):

$$K_\mu = K_\mu^{\,i} X_i = D_\mu . \tag{2.4}$$

As a result of this constraint, the gauge fields $K_\mu^{\,i}$ are identified with the tetrad matrices which satisfy with their inverses the orthonormality conditions

$$K_\mu^{\,i} K_i^{\,\nu} = \delta_\mu^\nu, \quad K_i^{\,\mu} K_\mu^{\,j} = \delta_i^j . \tag{2.5}$$

Moreover, the algebra is no longer the global Poincaré algebra. In fact, from (2.4) it follows that

$$\begin{aligned} [X_i, X_j] &= -K_i^{\,\mu} K_j^{\,\nu} [R_{\mu\nu}^{\,m\;\,n} X_{mn} + T_{\mu\nu}^{\,k} X_k] \\ &\equiv -R_{ij}^{\;mn} X_{mn} - T_{ij}^{\;\,k} X_k , \end{aligned} \tag{2.6}$$

where the components of the curvature and torsion tensors are given by

$$[D_\mu, D_\nu] = -R_{\mu\nu}^{\;ij} X_{ij} \tag{2.7}$$

$$T_{\mu\nu}^{\;\;\,k} = D_\mu K_\nu^{\,k} - D_\nu K_\mu^{\,k} . \tag{2.8}$$

To the commutator (2.6) we must add the remaining commutation relations of the local algebra, which remain the same as their global form:

$$[X_{ij}, X_{kl}] = f_{ij\,kl}^{\;m\,n} X_{mn}, \quad [X_{ij}, X_k] = f_{ij\;k}^{\;\;m} X_m \tag{2.9}$$

It now remains to show that, subject to the constraint (2.4), gauge transformations also account for general coordinate transformations. In Yang–Mills theory for every element g of a gauge group G, we have $D_\mu \rightarrow g D_\mu g^{-1}$, or, infinitesimally,

$$D_\mu \rightarrow (1 + \epsilon^A X_A) D_\mu (1 - \epsilon^A X_A), \tag{2.10}$$

where the index A runs over the generators of G. For the elements of the subgroup $H = \mathrm{SO}(3,1)$, which is realized linearly, the constraint (2.4) has no effect on (2.10), as expected. But for $A = i$, we get

$$\delta K_\mu^{\,i} = -D_\mu \epsilon^k + \epsilon^j K_j^{\,\lambda} T_{\mu\lambda}^{\;\;k}, \tag{2.11}$$

$$\delta H_\mu^{\,mn} = \epsilon^j K_j^{\,\lambda} R_{\lambda\mu}^{\;m\;\,n} = \epsilon^\lambda R_{\lambda\mu}^{\;m\;\,n}, \tag{2.12}$$

$$\delta R_{ij}^{\;mn} = -\epsilon^\lambda D_\lambda R_{ij}^{\;mn}; \tag{2.13}$$

That these transformations are equivalent to general coordinate transformations can be seen by computing the variation of a global tensor. For example, for the metric tensor

$$g_{\mu\nu} = \eta_{ij} K_\mu^{\,i} K_\nu^{\,j} \tag{2.14}$$

we get

$$\begin{aligned} \delta g_{\mu\nu} &= g'_{\mu\nu}(x) - g_{\mu\nu}(x) \\ &= -(g_{\nu\lambda} \partial_\mu \epsilon^\lambda + g_{\mu\lambda} \partial_\nu \epsilon^\lambda + \epsilon^\lambda \partial_\lambda g_{\mu\nu}). \end{aligned}$$

This is identical with the expression for $\delta g_{\mu\nu}$ computed from the familiar law of tensor analysis

$$g_{\mu\nu}(x) = \frac{\partial x'^\rho}{\partial x^\mu} \frac{\partial x'^\lambda}{\partial x^\nu} g'_{\rho\lambda}(x').$$

The above approach is quite general and not limited to the example just given. In particular, in Weyl's theory, where the linear gauge group is SO(3,1) ⊗ SO(1,1), we have an 11-parameter group of transformations with generators $\{X_{ij}, D, X_i\}$. Here the linear gauge symmetry is enlarged. But since (i) the enlargement falls within the restriction discussed in the Introduction[1] and (ii) the dim(G/H) is still $= 4$, the above discussion trivially extends to this case also.[3]

B. The extended conformal theory

We have seen that both Einstein's and Weyl's theories can be regarded as nonlinear realizations of a space–time gauge symmetry. In both cases, the local (nonlinear) transformations associated with the generators $\{X_i\}$ realize general coordinate transformations. In both theories the group SO(3,1) is a local gauge symmetry. The two theories differ in that in Weyl's theory the local gauge symmetry is enlarged to $SO(3,1) \otimes SO(1,1)$. We know, of course, that at energies which are small compared to Planck energy, SO(3,1) is the correct local symmetry of space–time, so that scale invariance must be broken at some stage. But it is, nevertheless, significant that such an enlarged local space–time symmetry exists and might play a role at very high energies. Given this premise, we want to show that the linear gauge symmetry $SO(3,1) \otimes SO(1,1)$ can be further enlarged. From the discussion given in the Introduction, it is clear that such a symmetry must be smaller than SO(4,2). If we demand, on physical grounds, that the extended conformal symmetry H have the nested property

$$SO(3,1) \subset SO(3,1) \otimes SO(1,1) \subset H,$$

then the only choice consistent with the four-dimensional symmetry of the geometry is

$$H = T'_4 \wedge SO(3,1) \otimes SO(1,1), \qquad (2.15)$$

where the abelian group T'_4 is isomorphic to translations in Minkowski space, so that $T'_4 \wedge SO(3,1)$ is locally isomorphic to the Poincaré group. Let $\{Y_k\}$ be the generators of T'_4. Then the algebra of the group H (not G) will be the set $\{X_{ij}, D, Y_k\}$. Since, as before, we want the generators $\{X_i\}$ to realize general coordinate transformations, there will be, in all, 15 generators in the local algebra. We have already seen in (2.6) that $\{X_i\}$ do not commute, so that, e.g., in Einstein's theory the algebra $\{X_{ij}, X_k\}$ is a deformation of the Poincaré algebra. It will be seen below that, in the same sense, the 15-parameter algebra

$$\{X_{ij}, D, Y_k, X_k\} \qquad (2.16)$$

may be *thought of* as a deformed SO(4,2) algebra, although this is not necessary.

We want to construct a nonlinear realization of the gauge algebra (2.16). Following the method outlined above, we start with the analog of (2.2) for this case:

$$\begin{aligned}\hat{D}_\mu &= \partial_\mu + H^{ij}_\mu X_{ij} + S_\mu D + C^k_\mu Y_k - K^i_\mu X_i \\ &\equiv D_\mu - K^i_\mu X_i = D_\mu - K_\mu. \end{aligned} \qquad (2.17)$$

The "interlocking" procedure specified by (2.3) and (2.4) is also applicable to this case without any alterations. But instead of Eqs. (2.5) and (2.7), we now have

$$\begin{aligned}&[X_i, X_j] \\ &\quad = -K^\mu_i K^\nu_j [R^{m}_{\mu\nu}{}^{n} X_{mn} + S_{\mu\nu} D + C_{\mu}{}^{k}_{\nu} Y_k + T_{\mu\nu}{}^k X_k] \\ &\quad = -[R_{ij}^{m\ n} X_{mn} + S_{ij} D + C^k_{ij} Y_k + T^k_{ij} X_k], \end{aligned} \qquad (2.18)$$

$$[D_\mu, D_\nu] = -R^{ij}_{\mu\nu} X_{ij} - S_{\mu\nu} D - C_{\mu}{}^{k}_{\nu} Y_k, \qquad (2.19)$$

where

$$R^{ij}_{\mu\nu} = H^{ij}_{\mu,\nu} - H^{ij}_{\nu,\mu} + f^{ij}_{kl\,mn} H_\mu{}^{lk} H^{m\ n}_{\nu}, \qquad (2.20)$$

$$S_{\mu\nu} = S_{\mu,\nu} - S_{\nu,\mu}, \qquad (2.21)$$

$$C_{\mu}{}^{k}_{\nu} = C_{\mu},{}^{k}{}_{\nu} - C_{\nu},{}^{k}{}_{\mu} + 2 f_l{}^k{}_{ij} H^{ij}_\mu C^l_\nu + S_\mu C^k_\nu - S_\nu C^k_\mu. \qquad (2.22)$$

To the commutator (2.18) we must add the commutators of the remaining generators. Since the subgroup H is realized linearly, its algebra is unaffected by the interlocking constraint. In addition to (2.9) we have

$$\begin{aligned}&[X_{ij}, D] = [Y_k, Y_j] = 0, \\ &[Y_j, D] = -Y_j, \quad [Y_j, X_{kl}] = f^m_{jkl} X_m, \\ &[X_j, D] = X_j. \end{aligned} \qquad (2.23)$$

The one remaining commutator $[X_i, Y_j]$ will be determined below by requiring that it be compatible with the constraint (2.4).

All the transformation laws of the fields are again given by (2.10), where now the index "A" runs over 15 parameters. Under the transformations specified by the parameters $\epsilon^j(x)$, we get, in addition to (2.11)–(2.13)

$$\delta S_\mu = \epsilon^j K^\lambda_j S_{\lambda\mu} = \epsilon^\lambda S_{\lambda\mu}, \qquad (2.24)$$

$$\delta C^k_\mu = \epsilon^j K^\lambda_j C_{\lambda}{}^{k}_{\mu} = \epsilon^\lambda C_{\lambda}{}^{k}_{\mu}, \qquad (2.25)$$

$$\delta C^k_{ij} = -\epsilon^\lambda D_\lambda C^k_{ij}. \qquad (2.26)$$

For the elements of the subgroup H which is realized linearly, the transformation laws of gauge fields are Yang–Mills transformations. For example, under scale transformations with parameter $\lambda(x)$,

$$\begin{aligned}&S_\mu \to S_\mu - \partial_\mu \lambda(x), \\ &H^{ij}_\mu \to H^{ij}_\mu, \quad C^k_\mu \to C^k_\mu. \end{aligned} \qquad (2.27)$$

The behavior of K^i_μ under these transformations is obtained from the identity[7] $[D_\mu, D] = 0$ and the constraint (2.4):

$$DK^i_\mu = K^i_\mu. \qquad (2.28)$$

The same identity also determines the scale dimension of various gauge fields:

$$DH^{ij}_\mu = DS_\mu = 0, \qquad (2.29)$$

$$DC^k_\mu = -C^k_\mu. \qquad (2.30)$$

Similarly, the action of the generators Y_j on various fields can be read off from the identity $[Y_j, D_\mu] = 0$:

$$Y_j H^{mn}_\mu = Y_j S_\mu = 0, \qquad (2.31)$$

$$Y_j C^k_\mu = f^k_{j\,mn} H^{m\ n}_{\mu} + \delta^k_j S_\mu. \qquad (2.32)$$

It follows from these that

$$Y_j Y_l C^k_\mu = 0. \qquad (2.33)$$

Let $\theta^k = \theta^k(x)$ be parameters corresponding to the infinitesimal gauge transformations generated by Y_k. Then from (2.10) we have

$$\begin{aligned}D_\mu &\to [1 + \theta^k(x) Y_k] D_\mu (1 - \theta^k Y_k) \\ &= D_\mu - D_\mu \theta^k Y_k. \end{aligned} \qquad (2.34)$$

This means that

$$\delta H^{ij}_\mu = \delta S_\mu = 0, \qquad (2.35)$$

$$\delta C^k_\mu = D_\mu \theta^k. \qquad (2.36)$$

To determine the action of Y_k on K^i_μ, we again start with the identity[7] $[D_\mu, Y_k] = 0$ and make use of the interlocking constraint $D_\mu = K^i_\mu X_i$ to obtain

$$(Y_k K^i_\mu) X_i - K^i_\mu [X_i, Y_k] = 0. \tag{2.37}$$

One solution of this equation is

$$[X_i, Y_k] = 0. \tag{2.38}$$

More generally, we must have

$$[X_i, Y_k] \subset X_l. \tag{2.39}$$

Here we consider only the special case (2.32). But in either case it is clear that the commutator $[X_i, Y_k]$ is different from what might have been inferred from a naive comparison with SO(4,2) algebra.[8] In other words, such an inference would be incompatible with (2.37). From (2.37) and (2.38) we have

$$Y_k K^i_\mu = 0. \tag{2.40}$$

This together with (2.31) and (2.35) shows that (a) H^{ij}_μ, S_μ, and K^i_μ are all annihilated by Y_k and that (b) under these transformations $\delta H^{ij}_\mu = \delta S_\mu = \delta K^i_\mu = 0$. In other words, the action of T'_4 on these fields is trivial. Therefore, this sector of the theory is self-contained. As long as we remain in this sector, any scale, Lorentz, and general coordinate invariant action will be, automatically, and trivially, invariant under T'_4 transformations.[9] Now recall that H^{ij}_μ, S_μ, and K^i_μ are fields which enter the gravity sector of Weyl theory. Therefore, the restriction to this sector makes the gravity sector of the extended theory equivalent to that of Weyl theory. Of course, if it turns out that space–time structure at extremely short distances has additional degrees of freedom which are describable by C^k_μ, then the above equivalence with Weyl's theory would no longer hold, and it is the extended theory which would then be viable.

Now we turn to the structure of matter multiplets in the extended theory. We know that in Einstein's theory, where the local symmetry is SO(3,1), matter fields transform as linear representations of this local symmetry. Similarly, in Weyl's theory, matter fields become linear representations of SO(3,1) ⊗ SO(1,1). Therefore, in the extended theory, matter fields must form linear representations of $T'_4 \wedge$ SO(3,1) ⊗ SO(1,1). Since $T'_4 \wedge$ SO(3,1) is locally isomorphic to the Poincaré group, its linear representations are well known. But, since T'_4 transformations are no longer translations (or conformal boosts) in Minkowski space, the eigenvalues qk of the generators Y^k do not have the interpretation of 4-momenta in space–time. In fact, these quantities, hereafter referred to a pseudomomenta, give rise to conservation laws which could make sense only at very short distances. In an originally scale invariant theory, we must select those representations of $T'_4 \wedge$ SO(3,1) which are the analogs of massless Poincaré states, that is those which correspond to $q^j q_j = 0$. Thus a matter field would be labeled by its helicity λ and its pseudomomentum $\mathbf{q}$.

We know that, at presently accessible energies, the correct local symmetry of space–time is the Lorentz group. Therefore, just as SO(3,1) ⊗ SO(1,1) symmetry must be broken down to SO(3,1), the linear gauge symmetry of the extended theory must be broken down, presumably dynamically, to SO(3,1). One possibility is that this symmetry breakdown takes place in two stages. In the first stage scale invariance is broken.[10] This allows for the appearance of dimensional parameters such as the gravitational coupling and the cosmological parameters in the gravity sector and a mass scale in the matter field sector. At this stage the local symmetry of space–time is $T'_4 \wedge$ SO(3,1), and matter fields will either remain massless or acquire masses of the order of Planck mass. In the next stage of symmetry breakdown, T'_4 symmetry is also broken, so that the remaining local space–time symmetry is SO(3,1). To see the effect of this symmetry breakdown, we note that if we restrict the gravity sector of the extended theory to the degrees of freedom of the Weyl theory, then, as we have seen, T'_4 symmetry is already acting on it trivially. So its breakdown will have no effect on the gravity sector. Its only effect would be (a) to reduce the local space–time symmetry of matter multiplets to SO(3,1), i.e., to that of Einstein's theory, and (b) to provide an additional mass scale for the matter field multiplets. This might provide a clue to the origin of the grand unification scale. In the presently popular (nonsupersymmetric) grand unified models the grand unification scale appears to be 3–4 orders of magnitude smaller than the Planck mass scale where, according to the above scenario, gravity and matter field sectors part company. Of course, it is not enough to know that an additional mass scale could arise naturally. The full implementation of the above idea requires the construction of concrete models in which the breakdown of T'_4 invariance does provide the correct numerical value for the grand unification (or some other) scale. Explicit models of this type will be reported elsewhere; cf. also the following section.

Instead of looking for a symmetry breaking scheme, one can, of course, break the T'_4 as well as the scale symmetry explicitly by simply starting from field theories which from the beginning violate these symmetries. That might be a satisfactory approach as long as one does not raise such questions as to how the dimensional parameters arise or how many scales there are.

III. APPLICATION TO EXTENDED CONFORMAL SUPERGRAVITY

The preceding work finds an immediate application to the construction of conformal supergravity theories in superspace. A minimal method of constructing such theories based on the supersymmetric extension of Weyl's theory has been given elsewhere.[6] It was constructed in the superspace of ordinary supergravity consisting of $(X^i, \theta^\alpha, \bar\theta^{\dot\alpha})$, where θ^α and $\bar\theta^{\cdot\alpha}$ are, respectively, undotted and dotted two component spinors.[11] Although such a formalism can be made internally consistent, the scale invariant actions which can be obtained that way have the unconventional feature that they involve fractional powers of the square of the curvature tensor. Moreover, these powers depend on the number N of the SU(N)-extended geometry. As a result, if one insists on scale invariance of the initial theory, the structure of the theory changes in a discontinuous manner when one goes from one N to the next.

It was later pointed out[12] that this peculiar feature can be avoided if one uses a superspace S with coordinates $Z^I = (X^i, \theta^\alpha, \eta^{\alpha'})$, where both θ^α and $\eta^{\alpha'}$ are four-component spinors under SL(2,C), but transform oppositely under scale transformations. Once such a superspace is constructed, one can also construct nonconformal supergravity theories in it. Before this can be done, one must make sure that the algebraic and geometrical properties of such an extended superspace are internally consistent. For one thing, this extension changes the dimension of S, and this might adversely affect the interlocking procedure. The proof of the consistency of the formalism used in Ref. 12 was left to another publication. Here we want to supply the missing proof and draw further conclusions from it.

To accomodate fermionic degrees of freedom, we will alter our notation from the previous sections in the following way: Consider a superspace S with coordinates $\{Z^I\} = \{X^i, \theta^\alpha, \eta^{\alpha'}\}$, where the indices $\{\alpha\}$ and $\{\alpha'\}$ stand, respectively, for $\{\alpha A\}$ and $\{\alpha' A\}$, and

$$\alpha, \alpha' = 1,\ldots,4, \quad A = 1,\ldots,N = \text{internal symmetry index},$$
$$I, J = i, j, \cdots (\text{even}), \;\; \alpha, \beta, \ldots, \alpha', \beta', \cdots \;\; (\text{odd}). \tag{3.1}$$

We have seen in previous sections that the geometries and invariances of Einstein, Weyl, and the extended theories can each be interpreted as a nonlinear realization of an appropriate gauge symmetry G, which is linear with respect to a subgroup H. In each case we have also seen that the corresponding algebraic structures have undergone significant deformations from Poincaré, inhomogeneous Weyl, and SO(4,2) algebras. In particular, in the case of the extended theory, there is the additional departure (2.38) from the SO(4,2) algebra. Therefore, in going to superspace S and constructing a nonlinear realization, we expect that our algebraic structure closely resembles SU(2,2|N), but, to satisfy all the consistency conditions, there will occur significant departures from this algebra. Let the generators of the super-Lie group G be the set

$$\{X_{\hat{A}}\} = \{X_{\hat{I}}, X_{\widehat{ij}}, \ldots, X_a\} \tag{3.2}$$

where the caret on top of an index implies that it is a group (algebra) index. Since we want the generators $\{X_{\hat{I}}\}$ to simulate general coordinate transformations in superspace, their number must be equal to $\{X^i, \theta^\alpha, \eta^{\alpha'}\}$:

$$\{X_{\hat{I}}\} = \{X_{\hat{i}}, X_{\hat{\alpha}}, Y_{\hat{\alpha}'}\} \tag{3.3}$$

This means that if we construct a nonlinear realization which is linear with respect to the subgroup with generators $\{\hat{A} \neq \hat{I}\}$, then $\dim(G/H)$ = dimension of S. It will be seen below that this equality of dimensions is crucial for implementing the interlocking scheme in superspace. Therefore, the choice of the (maximal) subgroup H (tangent space symmetry) is not purely a matter of taste. After the geometry has been set up, it is, of course, possible, as discussed in previous sections, to break the local symmetry H to a smaller subgroup.

It will be recalled that the algebra of SU(2,2|N) consists of the generators[13]

$$\{X_{\hat{A}}\} = \{X_{\hat{I}}, X_{\widehat{ij}}, D, Y_{\hat{i}}, X_a, \Gamma\}, \tag{3.4}$$

where $\{X_{\hat{I}}\}$ are given by (3.3) and where some of the important nonzero commutators are:

$$\begin{aligned} &[X_{\hat{i}}, D] = X_{\hat{i}}, \quad [Y_{\hat{i}}, D] = -Y_{\hat{i}}, \\ &[X_{\hat{\alpha}}, D] = \tfrac{1}{2}X_{\hat{\alpha}'}, \quad [Y_{\hat{\alpha}'}, D] = -\tfrac{1}{2}Y_{\hat{\alpha}'}, \end{aligned} \tag{3.5}$$

$$[X_{\hat{I}}, X_{\hat{J}}] = f_{\hat{I}\hat{J}}^{\hat{A}} X_{\hat{A}}, \tag{3.6}$$

$$[Y_{\hat{j}}, X_{\hat{k}}] = -2(\eta_{jk} D + M_{jk}), \tag{3.7}$$

$$[Y_j, X_{\hat{\alpha}}] = f_{j\hat{\alpha}}^{\hat{\beta}'} Y_{\hat{\beta}'}. \tag{3.8}$$

To construct a *local* nonlinear realization of the gauge symmetry associated with the algebra (3.4), which is linear with respect to the elements $\hat{A} \neq \hat{I}$, consider the quantity $\hat{D}_I$ defined as follows:

$$\begin{aligned} \hat{D}_I &= \partial_I + H_I^{\widehat{ij}} X_{\widehat{ij}} + S_I D + C_I^{\hat{k}} Y_{\hat{k}} \\ &\quad + H_I^a X_a + A_I \Gamma - K_I^{\hat{J}} X_{\hat{J}} \\ &= D_I - K_I, \end{aligned} \tag{3.9}$$

where

$$K_I \equiv K_I^{\hat{J}} X_{\hat{J}} = K_I^{\hat{j}} X_{\hat{j}} + K_I^{\hat{\alpha}} X_{\hat{\alpha}} + C_I^{\hat{\alpha}'} Y_{\hat{\alpha}'}. \tag{3.10}$$

Note that the right-hand side of (3.9) differs from the corresponding term in Ref. 12 by the additional term $C_I^{\hat{k}} Y_{\hat{k}}$. To construct the desired nonlinear realization, we (a) require that K_I and D_I be separately covariants under the local action of the group and (b) impose the interlocking constraint. To this end, we define a basis $\{\hat{X}_{\hat{I}}\}$ by

$$\hat{X}_{\hat{I}} = \hat{K}_{\hat{I}}^J D_J, \quad D_I = \hat{K}_I^{\hat{J}} \hat{X}_{\hat{J}}, \tag{3.11}$$

where

$$\hat{K}_{\hat{M}}^I \hat{K}_I^{\hat{J}} = \delta_{\hat{M}}^{\hat{J}}, \quad \hat{K}_M^{\hat{I}} \hat{K}_{\hat{I}}^J = \delta_M^J \tag{3.12}$$

and require that $D_{\hat{I}} = 0$, i.e.,

$$K_I = \hat{K}_I^{\hat{J}} \hat{X}_{\hat{J}} = D_I. \tag{3.13}$$

Then, as in the previous sections, we obtain

$$\begin{aligned} &[D_I, D_J\} \\ &\quad = -[R_{IJ}^{\widehat{ij}} X_{\widehat{ij}} + S_{IJ} D + C_{IJ}^{\hat{k}} Y_{\hat{k}} + R_{IJ}^a X_a + A_{IJ} \Gamma], \\ &\quad \equiv -R_{IJ}^{\hat{\Sigma}} X_{\hat{\Sigma}}, \quad \{\hat{\Sigma}\} = \{\hat{A} \,|\, \hat{A} \neq \hat{I}\}, \end{aligned} \tag{3.14}$$

where

$$R_{IJ}^{\widehat{ij}} = (-)^{\sigma_I \sigma_J} H_{I,J}^{\widehat{ij}} - H_{J,I}^{\widehat{ij}} + f_{\hat{k}\hat{l}\hat{m}\hat{n}}^{\widehat{ij}} H_I^{\hat{m}\hat{n}} H_J^{\hat{k}\hat{l}}, \tag{3.15}$$

$$S_{IJ} = (-)^{\sigma_I \sigma_J} S_{I,J} - S_{J,I}, \tag{3.16}$$

$$A_{IJ} = (-)^{\sigma_I \sigma_J} A_{I,J} - A_{J,I}, \tag{3.17}$$

$$\begin{aligned} C_{IJ}^{\hat{m}} &= (-)^{\sigma_I \sigma_J} C_{I,J}^{\hat{m}} - C_{J,I}^{\hat{m}} + f_{\hat{k}}{}^{\hat{m}}{}_{\widehat{ij}} (H_I^{\widehat{ij}} C_J^{\hat{k}} - C_I^{\hat{k}} H_J^{\widehat{ij}}) \\ &\quad + (C_I^{\hat{m}} S_J - S_I C_J^{\hat{m}}). \end{aligned} \tag{3.18}$$

From these it follows that

$$[X_{\hat{I}}, X_{\hat{J}}\} = -R_{\hat{I}\hat{J}}^{\hat{\Sigma}} X_{\hat{\Sigma}} - T_{\hat{I}\hat{J}}^{\hat{M}} X_{\hat{M}}, \tag{3.19}$$

where

$$R_{\hat{I}\hat{J}}^{\hat{\Sigma}} = (-)^{\sigma_{\hat{I}}(\sigma_J + \sigma_{\hat{J}})} K_{\hat{J}}^J K_{\hat{I}}^I R_{IJ}^{\hat{\Sigma}}, \tag{3.20}$$

$$T_{\hat{I}\hat{J}}^{\hat{M}} = (-)^{\sigma_{\hat{I}}(\sigma_J + \sigma_{\hat{J}})} K_{\hat{J}}^J K_{\hat{I}}^I T_{IJ}^{\hat{M}}, \tag{3.21}$$

$$T_{IJ}^{\hat{M}} = D_I K_J^{\hat{M}} - (-)^{\sigma_I \sigma_J} D_J K_I{}^{\hat{M}}. \tag{3.22}$$

Comparing the right-hand sides of the brackets (3.6) and (3.19), we find that, as expected, the constraint (3.13) has resulted in a locally deformed SU(2,2|N) algebra. As for the rest of the algebra (3.4), we note that the subalgebra H consisting of the elements $\{X_{\hat{A}} \,|\, \hat{A} \neq \hat{I}\}$ is realized linearly and is not affected by the constraint (3.13). In the language of fiber bundles,[7] these are generators of vertical (linear gauge) transformations. Moreover, since such transformations commute with the horizontal transformations generated by $\{D_I\}$, it follows that the commutators (3.5) also remain the same.

To check the remaining commutators (3.7) and (3.8), we again note the $\{ Y_{\hat{k}} \}$ are generators of vertical transformations, so that

$$[Y_{\hat{k}}, D_I] = 0. \tag{3.23}$$

Using the constraint (3.13), this translates into

$$(Y_{\hat{k}} K^{\hat{J}}_{I}) X_{\hat{J}} + K^{\hat{J}}_{I} [Y_{\hat{k}}, X_{\hat{J}}] = 0. \tag{3.24}$$

This is the analog for superspace of the condition (2.37) for the extended theory. Its possible solutions are

$$[Y_{\hat{k}}, X_{\hat{J}}] = 0 \tag{3.25}$$

or, more generally,

$$[Y_{\hat{k}}, X_{\hat{J}}] \subset X_{\hat{J}}. \tag{3.26}$$

Clearly, the commutator (3.7) is incompatible with either (3.25) or (3.26), although the commutator (3.8) is, in principle, consistent with (3.26). If we limit ourselves to the case (3.25), then it follows from (3.24) that

$$Y_{\hat{k}} K^{\hat{J}}_{I} = 0. \tag{3.27}$$

Except for the gauge transformations generated by $Y_{\hat{k}}$, the behavior of various fields under local gauge transformations was given in Ref. 12. As in Sec. II, it is based on insisting that all invariances, including general coordinate invariance in superspace, be derivable from the single rule $D_I \rightarrow g D_I g^{-1}$, $g \in G$, or, infinitesimally,

$$D_I \rightarrow (1 + \epsilon^{\hat{A}} X_{\hat{A}}) D_I (1 - \epsilon^{\hat{A}} X_{\hat{A}}). \tag{3.28}$$

For use in subsequent discussions we quote the results for $X_{\hat{A}} = X_{\hat{I}}$ and $X_{\hat{A}} = D$: Under local superspace translations generated by

$$\{ \epsilon^{\hat{J}} \} = \{ \epsilon^{\hat{J}}, \epsilon^{\hat{a}}, \epsilon^{\hat{\alpha}}, \epsilon^{\hat{\alpha}'} \} \tag{3.29}$$

we have

$$\delta K^{\hat{J}}_{I} = D_I \epsilon^{\hat{J}} - \epsilon^{M} T_{M}{}^{\hat{J}}_{I}, \tag{3.30}$$

$$\delta H^{\hat{\Sigma}}_{I} = \epsilon^{J} R^{\hat{\varepsilon}}{}_{JI}, \tag{3.31}$$

where

$$\epsilon^{M} = \epsilon^{\hat{J}} K^{M}_{\hat{I}}. \tag{3.32}$$

The variations of various fields under local scale transformations with parameter ϵ_D, as well as their scale dimensions, can be found from (3.13) and the identity $[D, D_I] = 0$. We have

$$\delta H^{\widehat{ij}}_{I} = \delta C^{\hat{k}}_{I} = \delta H^{a}_{I} = \delta A_I = 0, \tag{3.33}$$

$$\delta S_I = -\partial_I \epsilon_D, \tag{3.34}$$

$$D H^{\widehat{ij}}_{I} = D S_I = D H^{a}_{I} = D A_I = 0, \tag{3.35}$$

$$D C^{\hat{k}}_{I} = -C^{\hat{k}}_{I}, \tag{3.36}$$

$$D K^{\hat{J}}_{I} = K^{\hat{J}}_{I}, \quad D K^{\hat{\alpha}}_{I} = \tfrac{1}{2} K^{\hat{\alpha}}_{I}, \quad D C^{\hat{\alpha}'}_{I} = -\tfrac{1}{2} C^{\hat{\alpha}'}_{I}. \tag{3.37}$$

From these one can also compute the variation of $K = \det K^{\hat{J}}_{I}$ under local scale transformations. Since[4,14]

$$\delta K = (-)^{\sigma_I} K \delta K^{\hat{J}}_{I} K^{I}_{\hat{J}},$$

we get

$$\delta_D K = 4 \epsilon_D K. \tag{3.38}$$

Thus this determinant transforms in exactly the same way as the determinant of the vierbein field in four-dimensional space–time. It is this property which makes the form of scale invariant actions in our N-extended superspace independent of N.

Finally, consider the behavior of various fields under local transformations generated by $Y_{\hat{k}}$. Let $\rho^{\hat{k}}$ be the parameters of these transformations. Then from (3.28) we get

$$\delta_\rho H^{\widehat{ij}}_{I} = \delta S_I = \delta H^{a}_{I} = \delta A_I = 0, \tag{3.39}$$

$$\delta C^{\hat{k}}_{I} = D_I C^{\hat{k}}_{I}. \tag{3.40}$$

Moreover, from the identity $[Y_{\hat{j}}, D_I] = 0$, it follows that

$$Y_{\hat{n}} H^{\widehat{ij}}_{I} = Y_{n} S_I = Y_{\hat{n}} H^{a}_{i} = Y_{\hat{n}} A_I = 0, \tag{3.41}$$

$$Y_{\hat{n}} C^{\hat{k}}_{I} = \delta^{\hat{k}}_{n} S_I - f_{\hat{n}}{}^{k}_{\widehat{ij}} H^{\widehat{ij}}{}_{I}, \tag{3.42}$$

$$Y_{\hat{n}} Y_{\hat{m}} C^{\hat{k}}_{I} = 0. \tag{3.43}$$

It can be seen from (3.27) and (3.41) that with the exception of $C^{\hat{k}}_{I}$ the action of the generators $Y_{\hat{k}}$ on all fields is trivial, and any theory based on covariants which does not involve $C^{\hat{k}}_{I}$ is automatically invariant under the transformations generated by $Y_{\hat{k}}$. Therefore, all the results of Ref. 12, which were obtained by leaving out $Y_{\hat{k}}$ from the beginning, remain intact.

IV. CONCLUDING REMARKS

A number of general conclusions can be drawn from the results of the previous sections: Since we have ruled out SO(4,2) as the structure group of a nontrivial space-time, the best one can hope is a $T'_4 \wedge \mathrm{SO}(3,1) \otimes \mathrm{SO}(1,1)$ symmetry. The actual local symmetry could be, and usually is, smaller than this. Since we know that for distances larger than about 10^{-16} cm the Lorentz group is the correct local linear symmetry of space–time, we expect that any larger symmetry would be relevant, if at all, at shorter distances. But since in grand unified theories, in which gravity is not included, one already encounters very short distances, then it is possible that the extended symmetry would be relevant at such a unification scale. For one thing, contrary to SO(1,1) symmetry which, when broken, endows both the gravity and the grand unified sectors with a common scale, the Planck mass, e.g., the T'_4 symmetry seems to affect only the grand unified sector. Therefore, if the grand unification mass remains distinct from the Planck mass, the extended symmetry provides a rationale for the existence of another mass scale.

The application to conformal supergravity in superspace has both advantages and disadvantages. One of its desirable features is that the formulation in extended superspace provides a closer link to the gauge theories of internal symmetry. This is because the conformally invariant actions in extended superspace which bear a close relation to the conformally invariant actions in 4-space–time, are quadratic in the components of curvature and torsion for *all* N. Moreover, the dimensionless coupling constants in superspace remain dimensionless when all the fermionic coordinates are integrated out. This makes the identifications of these quantities with their 4-space–time counterparts in, say, Wess–Zumino gauge possible. This is to be contrasted with the formulations of conformal supergravity in the superspace of Poincaré supergravity, in which both the coupling constants and the fields acquire peculiar dimensions. On the other hand, one undesirable feature of working in the extended superspace is that there will, in general, be a larger number of auxiliary fields, and one needs more constraint equations to eliminate the unwanted fields. This can, in principle, complicate the solution of constraint equations. But since in this case the formalism is independent of N, the hope is that the solution for one nontrivial N will pave the way for obtaining the solutions for other N's.

ACKNOWLEDGMENTS

I would like to express my appreciation to the organizing committee of Aspen Center for Physics for its hospitality during the summer of 1981, where the first draft of this manuscript was written up. This work was supported, in part, by the Department of Energy under Contract No. DE-AS-2-76ER02978.

---

[1] See, e.g., L. P. Eisenhart, *Riemannian Geometry* (Princeton U. P., Princeton, N.J., 1925); E. A. Tagirov and I. T. Todorov, Acta Phys. Aust. **51**, 135 (1979).

[2] In this article Einstein's theory and Weyl's theory refer also to their respective generalizations, which have nonsymmetric connections. Thus by Einstein's theory we mean the Einstein–Cartan–Sciama–Kibble–Trautman theory.

[3] F. Mansouri, Phys. Rev. Lett. **42**, 1021 (1979).

[4] F. Mansouri, Phys. Rev. D **24**, 1056 (1981); Yale Report No. YTP81-12, 1981 (unpublished); *Proceedings of the VIth Workshop on Current Problems in Particle Theory,* edited by R. Casalbuoni *et al.* (Johns Hopkins Univ., Baltimore, Maryland, 1982).

[5] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[6] F. Mansouri, *Proceedings of the VIIth International Colloquium on Group Theory and Mathematical Physics, Austin, Texas,* edited by W. Beiglbock, A. Böhm, and E. Takasugi (Springer-Verlag, New York, 1978); F. Mansouri and C. Schaer, Phys. Lett. B **83**, 327 (1979).

[7] See, e.g., L. N. Chang, K. Macrae, and F. Mansouri, Phys. Rev. D **13**, 3192 (1976).

[8] A different approach based on taking the local linear symmetry to be the entire SO(4,2) algebra was given by M. Kaku, P. K. Townsend, and P. van Nieuwenhuizen, Phys. Lett. B **69**, 304 (1977); Phys. Rev. D **17**, 3179 (1978). This is incompatible with the theorem of Ref. 1, and its expressions for the components of various curvatures as well as the algebra itself are to be compared and contrasted with those given here.

[9] The analog of this case for conformal boost transformations in Minkowski space was demonstrated by G. Mack and A. Salam, Ann. Phys. (N.Y.) **53**, 174 (1969).

[10] As shown in Refs. 3 and 4, this can be done explicitly by requiring that, e.g., a quantity such as the scalar curvature to have a nonzero vacuum expectation value.

[11] For other approaches to superspace supergravity, see, e.g., J. Wess and B. Zumino, Phys. Lett. B **66**, 361 (1977); Phys. Lett. B **79**, 394 (1978); J. Wess, R. Grimm, and B. Zumino, Phys. Lett. B **73**, 415 (1978); L. Brink, M. Gell-Mann, P. Ramond, and J. Schwarz, Phys. Lett. B **74**, 336 (1978); Phys. Lett. B **76**, 417 (1978); S. J. Gates, Jr., and J. A. Shapiro, Phys. Rev. D **18**, 2768 (1978); Y. Neeman and T. Regge, Phys. Lett. B **74**, 31 (1978); Nuovo Cimento **1**, 1 (1978). W. Siegel, Nucl. Phys. B **142**, 301 (1978); J. Crispim-Ramao, A. Ferber, and P. G. O. Freund, Nucl. Phys. B **126**, 429 (1977). S. W. MacDowell, Phys. Lett. B **80**, 212 (1979); F. Mansouri, J. Math. Phys. **18**, 52 (1977).

[12] F. Mansouri and C. Schaer, Phys. Lett. B **101**, 51 (1981).

[13] See, e.g., P. G. O. Freund and I. Kaplansky, J. Math. Phys. **17**, 228 (1976).

[14] R. Arnowitt, P. Nath, and B. Zumino, Phys. Lett. B **56**, 81 (1975).

Quantum virial coefficients as cumulants of imaginary time-ordered Mayer diagrams

Antoine Royer
Centre de Recherche de Mathématiques Appliquées, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 18 August 1982; accepted for publication 15 October 1982)

The virial coefficients for a quantum gas (including quantum statistics) are expressed as sums of cumulants of connected (generalized) Mayer diagrams, the cumulants being built on the irreducible blocks of the diagrams. The Mayer diagrams are defined for the quantum case in terms of imaginary time-ordered exponentials, the quantum statistics being incorporated in the guise of multiparticle interactions. In order to extend Mayer diagrams to multiparticle interactions, we utilize terminology and methods from the theory of *hypergraphs*. The virial coefficients naturally separate into a quantum Boltzmann gas contribution, an ideal quantum gas contribution, and a final term expressing correlations between dynamics and statistics. In the classical limit, connected Mayer diagrams factorize into their irreducible blocks; the cumulants over irreducible blocks then vanish (by a basic property of cumulants), except for diagrams which are themselves irreducible, whence the classical result of Mayer (extended to multiparticle interactions). In the quantum case, the imaginary time ordering prevents the factorization into irreducible blocks by time entangling them. As a further illustration of the use of hypergraph-cumulant methods, we directly deduce the expressions of the virial coefficients in terms of Ursell–Kahn–Uhlenbeck cluster functions (the ideal quantum gas contribution naturally appears in that form).

PACS numbers: 05.30. − d, 02.10. + w

1. INTRODUCTION

There are several different manners of deriving and expressing the virial expansion of a gas (i.e., the expansion of the pressure in powers of the particle number density n).[1]

The most ancient method is that of Ursell, Kahn, and Uhlenbeck[2–6]: there, one first obtains, via the grand canonical formalism, the pressure as an expansion in powers of the fugacity (or activity) z, the coefficients of which are "cluster functions" closely related to cumulants.[5] The fugacity is then eliminated in favor of the density n by iteration. The resulting expressions of the virial coefficients are polynomials in the fugacity expansion coefficients. This method, applicable to both the classical and quantum cases, is very general and concise, but the expressions it yields for the virial coefficients are not very informative.

In the case of a classical gas interacting via pair forces, a much more physically informative, and mathematically convenient, form for the virial coefficients is in terms of Mayer diagrams.[7] To obtain the latter, the z expansion is first expressed as a sum of *connected* Mayer diagrams; after elimination of z, the virial coefficients turn out to be sums of *irreducible* diagrams, that is a *subclass* of the connected diagrams. This surprising simplification must have appeared somewhat miraculous to its discoverer ("due apparently to a numerical coincidence"[8]).

Later work clarified the mathematical and physical meaning of this "topological reduction"; it can be achieved in a much more illuminating manner by way of resummations,[9] i.e., one resums over classes of subdiagrams attached by a single vertex to a core irreducible diagram, thereby "renormalizing" fugacity vertices into density vertices.

In the case of a quantum Boltzmann gas (quantum dynamics but Boltzmann statistics), one can parallel the Mayer method insofar as expressing the fugacity expansion as a sum of connected quantum Mayer diagrams, the latter being defined in terms of imaginary time-ordered exponentials. However, when attempting resummations of the same kind as in the classical case, one is confronted with a new feature: because time-ordered noncommuting operators are involved, subdiagrams joined at a single vertex no longer factorize, because of the time entanglement, and the desired resummations cannot be performed. To be able to perform resummations, one must first explicitly expand the time-ordered exponentials entering the Mayer diagrams, thereby obtaining *Feynman* diagrams. The latter have a (imaginary) time dimension: each particle is represented by a line parallel to the imaginary time axis (taken as vertical), rather than by a single point, and the interaction between two particles now appears as a succession of horizontal lines joining the corresponding particle lines (one obtains a Mayer diagram by projecting a Feynman diagram onto the horizontal plane). Resummations can now be performed, but over classes of subdiagrams which are not only sufficiently weakly connected to, but also which do not *time overlap*, other parts of the diagram. One can thereby obtain various "renormalized" expansions,[10] but *not* the virial expansion. It is in fact impossible to arrive at the latter in such a manner, since the virial coefficients turn out to be *not* sums of diagrams, but sums of *products* of diagrams.

Another route to the virial expansion was indicated by Brout[11–13] and by Kubo[14] (independently). It consists in working directly with the (petit) canonical partition function, and making use of the properties of cumulants.[15] It is

0022-2488/83/040897-16$02.50

thereby possible to rederive the classical Mayer result in a simple and concise manner.

Kubo[14] hinted at the possibility of using this method to obtain the virial expansion of the quantum Boltzmann gas, but did not go into any detail. Brout[16] also considered the quantum gas, and obtained the free energy in terms of special diagrams containing cumulants of single particle state populations; these cumulants express the correlations between the different populations due to the fixed total number of particles in the canonical ensemble. Related results were given by Lee and Yang.[17] But again, it is not possible to extract simple expressions for the virial coefficients from these results.

Thus, in all existing treatments, the quantum virial coefficients have rather complicated and unilluminating expressions, whose relations to the classical Mayer expressions are not at all obvious.

In the present paper, we show that the virial coefficients for a quantum gas (including quantum statistics) can be expressed in a simple meaningful form directly related to the classical Mayer result.[7] This we achieve by extending the methods of Brout[11] and Kubo.[14]

Specifically, we obtain the virial coefficients as cumulants of connected (quantum) Mayer diagrams, the cumulants being built on the irreducible blocks of each diagram (i.e., the irreducible blocks play the role of the random variables in an ordinary cumulant). In the classical limit, time ordering is relaxed, and the Mayer diagrams factorize into their irreducible blocks, i.e., the latter become statistically independent; it then follows (by the basic property of cumulants of vanishing if their arguments are not statistically dependent)[14,15] that cumulants of diagrams containing more than one irreducible block vanish, whence the classical Mayer result.[7]

The quantum statistics are introduced in the guise of effective multiparticle interactions. We are therefore led to treat from the outset the case of nonpair forces. For this purpose, we generalize Mayer diagrams by making use of terminology and methods from the theory of hypergraphs.[18] These tools, we find, are more flexible, and require less of an effort of imagination than others that have been proposed for dealing with multiparticle interactions[19]; in fact, the latter bring about practically no complication, both conceptually and notationally, once the proper notions have been introduced.

In Sec. 2, we recall the definitions and basic properties of cumulants. In Sec. 3, generalized Mayer diagrams (hypergraphs) allowing for multiparticle interactions, are introduced, and some simple lemmas demonstrated. The basic statistical mechanical formulas to be used are listed in Sec. 4. In Sec. 5, the virial expansion for the quantum Boltzmann gas is obtained. The quantum statistics are added in Sec. 6. In Sec. 7, we show how cumulant methods can be used to obtain directly the expressions of the virial coefficients in terms of the Ursell–Kahn–Uhlenbeck cluster functions. Two appendices contain technical details.

The paper is essentially self-contained; the only result not proven is the explicit expressions of cumulants, which are well known.[14,15]

1.1 Notation (multisets and partitions)

We denote $\mathbb{N} = \{0,1,2,...\}$ the set of natural numbers, and $\mathbb{N}_+ = \{1, 2, 3,...\}$. For any $n \in \mathbb{N}_+$, we denote

$$\underline{n} \equiv \{1, 2, ..., n\}.$$

A *set* S is a collection of objects, all *distinct* from one another. A *multiset* M is a collection of objects not necessarily all distinct from one another;[20] e.g., let the set

$$S = \{a, b, c, d\}; \tag{1.1a}$$

then

$$M = \{a, a, a, b, c, c\} \tag{1.1b}$$

is a multiset with elements in S. We denote $\mathscr{M}(S)$ the set of all multisets with elements in S. A multiset $M \in \mathscr{M}(S)$ may be represented by an element of $\mathbb{N}^S$ (the set of functions from S to $\mathbb{N}$):

$$M \leftrightarrow \{m_X, X \in S\} \in \mathbb{N}^S, \tag{1.2}$$

where $m_X \in \mathbb{N}$ is the number of times the element $X \in S$ is contained in M.

The set relations and operations (identity, inclusion, union, intersection, etc.) are extended to multisets in an obvious manner: Let $M \leftrightarrow \{m_s, s \in S\}$ and $M' \leftrightarrow \{m'_s, s \in S\}$ be two multisets in $\mathscr{M}(S)$; then

$$\begin{aligned} &M \subset M' \Leftrightarrow m_s \leqslant m'_s \quad \text{for all } s \in S, \\ &M \cup M' \leftrightarrow \{\text{Max}(m_s, m'_s), \quad s \in S\}, \\ &M \cap M' \leftrightarrow \{\text{Min}(m_s, m'_s), \quad s \in S\}. \end{aligned} \tag{1.3}$$

We also define a direct sum

$$M + M' \leftrightarrow \{m_s + m'_s, s \in S\}.$$

The number of elements (cardinality) of a set S is denoted $|S|$; likewise the number of members of a multiset M (counting repetitions) is denoted $|M|$ [e.g., in example (1.1), $|S| = 4$, $|M| = 6$].

It is often convenient to be able to regard a multiset M as a set. This can be done by establishing a one-to-one correspondence between the members of M and the elements of an *indexing set* I; e.g., continuing example (1.1), let

$$I = \underline{6} = \{1, 2, 3, 4, 5, 6\} \tag{1.1c}$$

and define the function $X: I \to S$ as $X_1 = X_2 = X_3 = a$, $X_4 = b$, $X_5 = X_6 = c$; then

$$M = \{X_i, \quad i \in I\}, \tag{1.4}$$

which may now be handled as a set, since each member has been given a distinct *name*.

To illustrate the notation, let $\{\lambda_X, X \in S\}$ be a set of algebraic objects; we have

$$\begin{aligned} \exp\Big(\sum_{X \in S} \lambda_X\Big) &= \sum_{\{m_X\} \in \mathbb{N}^S} \prod_{X \in S} (\lambda_X)^{m_X} / m_X! \\ &= \sum_{M \in \mathscr{M}(S)} (1/M!) \prod_{X \in M} \lambda_X, \end{aligned} \tag{1.5}$$

where

$$M! \equiv \prod_{X \in S} (m_X!) \quad \text{if } M \leftrightarrow \{m_X, X \in S\}. \tag{1.6}$$

It is understood that in $\Pi_{X \in M}$ (or $\Sigma_{X \in M}$), repetitions of elements must be taken into account; e.g., with M as in (1.1b), we have

$$\prod_{X\in M}\lambda_X=(\lambda_a)^3\lambda_b(\lambda_c)^2. \tag{1.1d}$$

We define a *partition* $P(M)$ of a multiset M as a multiset of nonempty sub-multisets of M, such that their direct sum equals M:

$$P(M)=\{M_1,M_2,...,M_p\},\quad M\supset M_j\neq\emptyset,$$
$$M_1+M_2+\cdots+M_p=M. \tag{1.7}$$

Another partition $P'(M)$ is a *subpartition* of $P(M)$, denoted

$$P'(M)\leqslant P(M) \tag{1.8}$$

if it can be obtained from $P(M)$ by further partitioning the members of $P(M)$ (note that $P'\leqslant P$ implies $|P'|\geqslant|P|$); $P(M)$ is then called a *superpartition* of $P'(M)$.

When we specialize the above definitions to sets, we obtain the usual notions: a partition of a set S is a set $P(S)=\{S_1,S_2,...,S_p\}$ of nonempty subsets $S_j\subset S$ such that $S_i\cap S_j=\emptyset$ if $i\neq j$ and $\cup_{j=1}^{p}S_j=S$; $P'(S)=\{S'_1,...,S'_{p'}\}$ is a subpartition of $P(S)$ if each S'_j is entirely contained in one of the S_j. We denote $\pi(S)$ the set of all (distinct) partitions of S.

Let the multiset M be indexed by the set I: $M=\{X_i,\ i\in I\}$. To each partition $P(I)=\{I_1,I_2,...,I_p\}$ of I corresponds a partition of M,

$$P_{P(I)}(M)=\{M_1,M_2,...,M_p\},$$

where (1.9)

$$M_j=\{X_i,\quad i\in I_j\}$$

(note that M_j and $M_{j'}$, $j\neq j'$, may have a nonempty intersection, and even be identical, in contradistinction to I_j and $I_{j'}$; also, to a given $P(M)$ may correspond several different partitions of I). We denote

$$\pi(M)=\{P_{P(I)}(M),\ P(I)\in\pi(I)\} \tag{1.10}$$

the collection of all partitions of M *regarded as a set*. Note that $\pi(M)$ it itself a multiset.

Examples: (i) Let $M=\{a,a,a,a,b,b\}=\{X_i,\ i\in\underline{6}\}$; to the partition $P(\underline{6})=\{(1,5),(2,6),(3,4)\}$ corresponds $P(M)=\{(a,b),(a,b),(a,a)\}$.

(ii) Let $M=\{a,a,b\}=\{X_i,\ i\in\underline{3}\}$; we have

$$\pi(\underline{3})=\{[1,2,3)],[(1,2),(3)],[(1,3),(2)],[(1),(2,3)],[(1),(2),(3)]\},$$

whence

$$\pi(M)=\{[(a,a,b)],[(a,a),(b)],[(a,b),(a)],[(a),(a,b)],[(a),(a),(b)]\}.$$

2. CUMULANTS

Let S be a set of stochastic variables, and denote $\langle\ \rangle$ the statistical averaging operation. For all multisets $M\in\mathscr{M}(S)$, $M\leftrightarrow\{m_X,\ X\in S\}\in\mathbb{N}^S$, we define *cumulants* $\langle\Pi_{X\in M}X\rangle_c=\langle\Pi_{X\in S}X^{m_X}\rangle_c$ and *anticumulants* $\langle\Pi_{X\in M}X\rangle_a$ through the relations

$$\left\langle\exp\Big(\sum_{X\in S}\lambda_X X\Big)-1\right\rangle_a=\exp[\langle e^{\Sigma\lambda_X X}-1\rangle]-1, \tag{2.1a}$$

$$\left\langle\exp\Big(\sum_{X\in S}\lambda_X X\Big)-1\right\rangle_c=\ln\left\langle\exp\Big(\sum_{X\in S}\lambda_X X\Big)\right\rangle, \tag{2.1c}$$

where the λ_X, $X\in S$, are arbitrary constants. The left-hand sides are shorthands for [see (1.5)]

$$\left\langle\exp\Big(\sum\lambda_X\Big)X-1\right\rangle_{c,a}=\sum_{M\in\mathscr{M}(S)}(1/M!)\Big(\prod_{X\in M}\lambda_X\Big)\left\langle\prod_{X\in M}X\right\rangle_{c,a}. \tag{2.2}$$

The right-hand sides of (2.1) are to be similarly expanded in powers of the λ_X, and corresponding coefficients identified. There result the explicit expressions[14,15]

$$\left\langle\prod_{X\in M}X\right\rangle_{a,c}=\sum_{P\in\pi(M)}A^{(a,c)}_{|P|}\prod_{M_i\in P}\left\langle\prod_{X\in M_i}X\right\rangle, \tag{2.3a,c}$$

the sum being over all partitions $P=\{M_1,M_2,...,M_{|P|}\}\in\pi(M)$ of the multiset M, and

$$A^{(a)}_p=1, \tag{2.4a}$$

$$A^{(c)}_p=(p-1)!(-)^{p-1}, \tag{2.4c}$$

The first few cumulants and anticumulants are given by (we sometimes abbreviate $\langle X_1X_2\cdots\rangle\equiv\langle 12\cdots\rangle$) (we *define* $\langle 1\rangle_c=\langle 1\rangle_a=\langle 1\rangle=1$)

$$\langle X\rangle_c=\langle X\rangle_a=\langle X\rangle, \tag{2.5}$$

$$\langle X_1X_2\rangle_c=\langle X_1X_2\rangle-\langle X_1\rangle\langle X_2\rangle,$$

$$\langle 123\rangle_c=\langle 123\rangle-\langle 12\rangle\langle 3\rangle-\langle 13\rangle\langle 2\rangle$$
$$-\langle 1\rangle\langle 23\rangle+2\langle 1\rangle\langle 2\rangle\langle 3\rangle, \tag{2.6c}$$

...

and

$$\langle X_1X_2\rangle_a=\langle X_1X_2\rangle+\langle X_1\rangle\langle X_2\rangle,$$

$$\langle 123\rangle_a=\langle 123\rangle+\langle 12\rangle\langle 3\rangle+\langle 13\rangle\langle 2\rangle \tag{2.6a}$$
$$+\langle 1\rangle\langle 23\rangle+\langle 1\rangle\langle 2\rangle\langle 3\rangle.$$

...

Because $\langle\ \rangle$ acts linearly, so do $\langle\ \rangle_c$ and $\langle\ \rangle_a$ in view of (2.3):

$$\left\langle(Y_1+Y_2)\prod_{X\in M}X\right\rangle_b=\left\langle Y_1\prod_{X\in M}X\right\rangle_b+\left\langle Y_2\prod_{X\in M}X\right\rangle_b, \tag{2.7}$$

where $\langle\ \rangle_b$ stands for $\langle\ \rangle$, $\langle\ \rangle_c$, or $\langle\ \rangle_a$. Thus all the ordinary algebraic manipulations can be done inside $\langle\ \rangle_c$ or $\langle\ \rangle_a$, so that (2.2) is not just a notational convenience.

Sometimes, it is necessary to specify explicitly on which variables $\langle\ \rangle_a$ or $\langle\ \rangle_c$ operates; e.g., the left-hand side of (2.1c) should more accurately be written as

$$\left\langle\exp\Big(\sum\lambda_X X\Big)-1\right\rangle_{c\{X\}}, \tag{2.8}$$

where $\{X\}$ indicates that $\langle\ \rangle_c$ operates on the X's, not the λ's which are scalars with respect to that operation.

Cumulants and anticumulants are reciprocal, in the sense that[21]

$$\left\langle\prod_{X\in M}X\right\rangle_{ca}=\left\langle\prod_{X\in M}X\right\rangle_{ac}=\left\langle\prod_{X\in M}X\right\rangle, \tag{2.9}$$

where $\langle\Pi_{X\in M}X\rangle_{ca}$ is given by (2.3a) with $\langle\ \rangle$ replaced by $\langle\ \rangle_c$ everywhere (likewise with $\langle\ \rangle_{ac}$); e.g., $\langle X_1X_2\rangle_{ca}=\langle X_1X_2\rangle_c+\langle X_1\rangle_c\langle X_2\rangle_c=\langle X_1X_2\rangle$.

The construction (2.3)–(2.6) may be applied to any family $\{A_M,\ M\in\mathscr{M}(S)\}$ of objects indexed by multisets; that is, we may define "Ursell clusters" A^c_M as

$$A_1^c = A_1, \quad A_{12}^c = A_{12} - A_1 A_2, \tag{2.10}$$

$$A_{123}^c = A_{123} - A_{12}A_3 - A_{13}A_2 - A_1 A_{23} + 2A_1A_2A_3.$$

$$\cdots$$

It is possible to use the same notation as in (2.3)–(2.6) if we formally denote

$$A_M \equiv \Big\langle \prod_{X\in M} X \Big\rangle, \tag{2.11}$$

where the X's are here purely formal algebraic objects, and $\langle\ \rangle$ a formal "averaging" operation with the convention $\langle 1 \rangle = 1$. We then have

$$A_M^c = \Big\langle \prod_{X\in M} X \Big\rangle_c \tag{2.10$'$}$$

The present discussion applies to both the cases where the X's are true stochastic variables, or just a notational device.

Two sets S and S' of stochastic variables are said $\langle\ \rangle$-independent (or statistically independent) if

$$\Big\langle \Big(\prod_{X\in M} X\Big)\Big(\prod_{Y\in M'} Y\Big)\Big\rangle = \Big\langle \prod_{X\in M} X\Big\rangle\Big\langle \prod_{Y\in M'} Y\Big\rangle \tag{2.12}$$

for all $M\in\mathscr{M}(S)$ and $M'\in\mathscr{M}(S')$; otherwise S and S' are $\langle\ \rangle$-dependent. We say that M or $\langle \Pi_{X\in M} X\rangle$ is $\langle\ \rangle$-linked if the members of the multiset M do not belong to two or more $\langle\ \rangle$-independent sets; otherwise M is $\langle\ \rangle$-unlinked.

The basic property of cumulants is:

Lemma 2.1[14,15]: A cumulant vanishes if its arguments belong to two or more independent sets, i.e., $\langle (\Pi_{X\in M} X)(\Pi_{Y\in M'} Y)\rangle_{c\{X,Y\}} = 0$ if $M\in\mathscr{M}(S)$ and $M'\in\mathscr{M}(S')$, where S and S' are $\langle\ \rangle$-independent.

Proof: Immediate from definition (2.1c):

$$\exp[\langle \exp(\Sigma_{X\in S} X + \Sigma_{Y\in S'} Y) - 1\rangle_c] \overset{(2.1c)}{=} \langle e^{\Sigma X + \Sigma Y}\rangle$$

$$\overset{(2.12)}{=} \langle e^{\Sigma X}\rangle\langle e^{\Sigma Y}\rangle \overset{(2.1c)}{=} \exp[\langle e^{\Sigma X} - 1\rangle_c + \langle e^{\Sigma Y} - 1\rangle_c],$$

implying that cumulants mixing X's and Y's vanish.

The following lemma, obvious from considering (2.3), will also be useful:

Lemma 2.2: Let $P'(M) = \{M_1', M_2', \ldots, M_p'\}$ be a partition of the multiset M and denote

$$Y_j \equiv \prod_{X\in M_j'} X, \quad j = 1 \ldots p. \tag{2.13}$$

If in $\langle \Pi_{X\in M} X\rangle_{c\{X\}}$, Eq. (2.3c), we delete all terms $P\in\pi(M)$ which are not superpartitions of $P'(M)$ [i.e., $\Sigma_{P\in\pi(M)}$ is replaced by $\Sigma_{P\in\pi(M), P\geqslant P'}$], there results

$$\Big\langle \prod_{j=1}^{p} Y_j \Big\rangle_{c\{Y\}} \equiv \Big\langle \prod_{X\in M} X \Big\rangle_{c\{Y\}}. \tag{2.14}$$

Example: Let $Y_1 = X_1X_2$, $Y_2 = X_3$; by suppressing in $\langle X_1X_2X_3\rangle_{c\{X\}}$ all terms such as $\langle X_1X_3\rangle\langle X_2\rangle$ which do not preserve the integrity of the Y's (i.e., such that X's belonging to the same Y appear in different $\langle\ \rangle$ factors), we obtain

$$\langle X_1X_2X_3\rangle - \langle X_1X_2\rangle\langle X_3\rangle$$
$$= \langle Y_1Y_2\rangle - \langle Y_1\rangle\langle Y_2\rangle$$
$$= \langle Y_1Y_2\rangle_{c\{Y\}} \equiv \langle X_1X_2X_3\rangle_{c\{Y\}}.$$

Corollary: Let

$$f_X(x) = a_{X1}x + a_{X2}x^2 + \cdots, \quad X\in S, \tag{2.15}$$

where the a_{Xi} are scalars. If in $\langle \Pi_{X\in S} f_X(X)\rangle_{c\{X\}}$, we delete all terms in which any given $X\in S$ appears inside more than one $\langle\ \rangle$ factor, we get $\langle \Pi_{X\in S} f_X(X)\rangle_{c\{f\}}$ [remark: in (2.15), S serves as indexing set].

Example: If in $\langle f(X_1)f(X_2)\rangle_{c\{X\}}$, we retain only the terms of the form $\langle X_1^m X_2^n\rangle$ and $\langle X_1^m\rangle\langle X_2^n\rangle$, we get

$$\langle f(X_1)f(X_2)\rangle - \langle f(X_1)\rangle\langle f(X_2)\rangle \equiv \langle f(X_1)f(X_2)\rangle_{c\{f\}}.$$

Remark: Lemma 2.2 and its corollary also apply to anticumulants, as they do not depend on the value of A_p [Eqs. (2.4)].

2.1 Leveling operators

Sometimes, in order to be able to use cumulant methods, it is useful to express a given quantity as an exponential preceded by an operator which *selects* the given quantity out of the expanded exponential. As a trivial example, we can write any quantity X as $X = L_1 e^X$ where L_1 suppresses all terms X^m with $m \neq 1$; more generally,

$$\prod_{i\in \underline{N}} X_i = L'\exp\Big(\sum_{i\in \underline{N}} X_i\Big), \tag{2.16}$$

where L' suppresses all terms in the expansion of the rhs which do not contain each index $1,2,\ldots,N$ exactly once. More interestingly, we have[14]

$$\prod_{i\in \underline{N}} (1 + X_i) = 1 + \sum_{E\subset \underline{N}} \Big(\prod_{i\in E} X_i\Big) = L\exp\Big(\sum_{i\in \underline{N}} X_i\Big), \tag{2.17}$$

where L, which may be called a "leveling operator," suppresses all terms which contain repeated indices. The summation $\Sigma_{E\subset \underline{N}}$ is over all subsets of $\underline{N}$, i.e.,

$$\sum_{E\subset \underline{N}}\Big(\prod_{i\in E} X_i\Big) = \sum_{i=1}^{N} X_i + \sum_{i<j} X_iX_j + \sum_{i<j<k} X_iX_jX_k + \cdots. \tag{2.18}$$

Lemma 2.3: Let $\{X_i, i\in \underline{N}\}$ be a set of stochastic variables. We have

$$\Big\langle \prod_{i\in \underline{N}} X_i \Big\rangle_{a\{X\}} = L'\exp\Big[\sum_{E\subset \underline{N}} \Big\langle \prod_{i\in E} X_i\Big\rangle\Big] \tag{2.19}$$

$$= \Big(\prod_{i\in \underline{N}} \langle X_i\rangle\Big) L\exp\Big[\sum_{E\subset \underline{N}, |E|\geqslant 2} \Big\langle \prod_{i\in E} \overline{X}_i\Big\rangle\Big], \tag{2.20}$$

where $\overline{X}_i \equiv X_i/\langle X_i\rangle$.

Proof: We have, in view of (2.1), (2.16), and (2.9),

$$\Big\langle \prod_{i\in \underline{N}} X_i \Big\rangle_{a\{X\}}$$
$$= L'\Big\langle \exp\Big(\sum_{i\in \underline{N}} X_i\Big)\Big\rangle_a = L'\exp(\langle e^{\Sigma X_i} - 1\rangle_{ac})$$
$$= L'\exp(\langle e^{\Sigma X_i} - 1\rangle). \tag{2.21}$$

We may now insert L inside $(\cdots)$ in the last line, since its action will simply be redundant with that of L'; we thereby get (2.19) in view of (2.17). If we replace each X_i by $\overline{X}_i$ in (2.19), we obtain

$$\Big\langle \prod_{i} \overline{X}_i \Big\rangle_a = L'\exp\Big(\sum_{i\in \underline{N}} 1_i + \sum_{E\subset \underline{N}, |E|\geqslant 2} \Big\langle \prod_{i\in E} \overline{X}_i\Big\rangle\Big), \tag{2.22}$$

where each 1_i is to be replaced by 1 after L' has acted; (2.22) is clearly equivalent to (2.20).[22] Q.E.D.

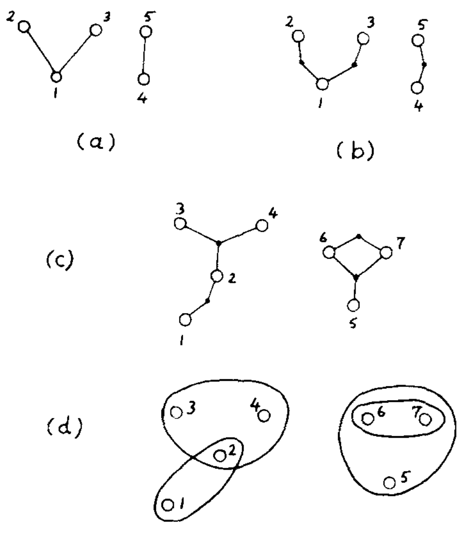


FIG. 1. (a) Mayer diagram representing $f_{12}f_{13}f_{45}$; (b) the generalized Mayer diagram representing the same product; (c) generalized Mayer diagram representing the product $f_{12}f_{234}f_{567}f_{67}$; and (d) the standard representation of the corresponding hypergraph $\{(1,2),(2,3,4),(5,6,7),(6,7)\}$.

3. GENERALIZED MAYER DIAGRAMS (HYPERGRAPHS)

Given a product $\Pi_{i<j} f_{ij}$, it is conveniently represented by a Mayer diagram, wherein each index i is drawn as a small circle labeled i, and each f_{ij} as a line joining circles i and j [Fig. 1(a)]. We shall need a slight generalization of these diagrams or graphs.

A *hypergraph* [18] is a multiset

$$H = \{E_1, E_2, \ldots, E_n\} \tag{3.1}$$

whose members E_j are nonempty sets. The elements of

$$V(H) \equiv \bigcup_{j=1}^{n} E_j \tag{3.2}$$

are called the *vertices* of H, the E_j are called the *edges*. We represent H by a diagram wherein each vertex $v_i \in V(H)$ is drawn as a small circle labeled v_i, and each edge E_j as a dot with lines joining it to each vertex it contains [Fig. 1(b) and (c)]. These *generalized Mayer diagrams* clearly reduce to ordinary Mayer diagrams when each E_j contains only two vertices, except that the bonds have slightly more personality [Fig. 1(b)].

The standard representation of a hypergraph is as shown in Fig. 1(d), each edge being drawn as a curve encircling the vertices it contains. Another possible representation is in terms of simplexes.[19] However, the representation in Fig. 1(b) and (c) as a *bipartite graph* (i.e., a graph with two kinds of vertices, no lines between vertices of the same kind) is much easier to visualize. Also, it makes visually manifest the symmetry existing between edges and vertices; this latter symmetry often allows one to *dualize* definitions and theorems by simply interchanging edges and vertices.

The multiset relations and operations ($=, \subset, \cup, \cap, +$) apply to hypergraphs (Fig. 2). H' is a *subhypergraph* of H if $H' \subset H$. A hypergraph is *simple* if no two of its edges are identical.

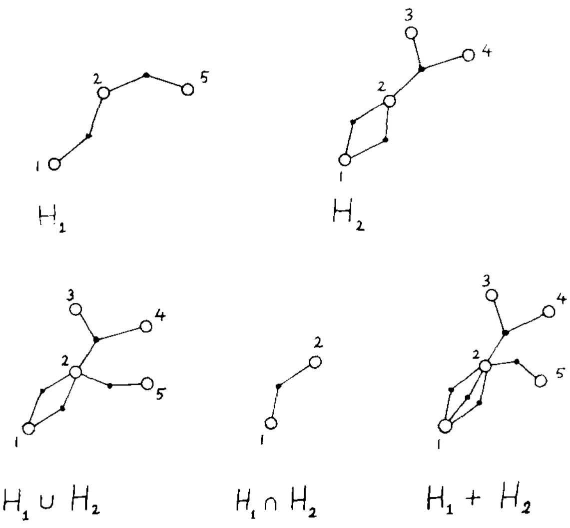


FIG. 2. Union, intersection, and direct sum of two hypergraphs.

We denote $\mathscr{H}(S)$ the set of hypergraphs with vertex set *equal* to the set S, and $\mathbb{H}(S) = \cup_{E\subset S}\mathscr{H}(E)$ the set of hypergraphs with vertex set *contained* in S. We add a subscript s to indicate restriction to simple hypergraphs [e.g., $\mathscr{H}_s(S)$ is the set of simple hypergraphs with vertex set S].

Given a family $\{f_E, E\subset S\}$ of objects labeled by subsets of S, and a hypergraph $H\in\mathscr{H}(S)$, we denote

$$H^f \equiv \prod_{E\in H} f_E. \tag{3.3}$$

This is the natural generalization of the product $\Pi_{i<j} f_{ij}$ mentioned at the beginning.

It is often necessary to assign to the edges of a hypergraph additional characteristics or labels; we shall globally call these "flavors." We define a *flavored hypergraph* as a multiset

$$H = \{(E_i, \phi_i), i\in I\}, \tag{3.4}$$

wherein each *edge* (E_i, ϕ_i) is a couple: $E_i \subset S$ is called the *value* of the edge, and the label ϕ_i its *flavor*. In the diagram of H, the dot representing the edge (E_i, ϕ_i) is labeled ϕ_i (Fig. 3). We allow for *flavorless* edges, simply denoted by their value.

We denote $\mathscr{H}(S;\phi^1,\phi^2,\ldots,\phi^m)$ and $\mathbb{H}(S;\phi^1,\phi^2,\ldots,\phi^m)$ the sets of hypergraphs *containing* the flavors $\phi^1,\phi^2,\ldots,\phi^m$, and having vertex sets equal to, and contained in S, respectively.

Given a (unflavored) hypergraph $H\in\mathscr{H}(S)$, we denote $H^{(\phi)}\in\mathscr{H}(S;\phi)$ the flavored hypergraph obtained by assigning

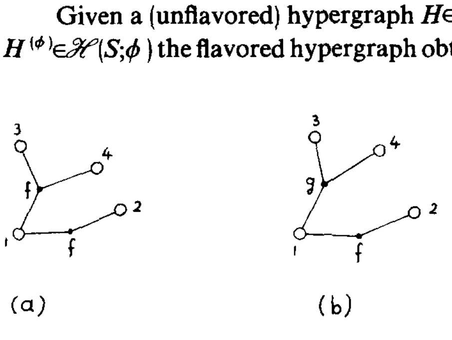


FIG. 3. (a) 1-flavored hypergraph $\{[(1,2),f], [(1,3,4),f]\}$ representing the product $f_{12}f_{134}$; (b) 2-flavored hypergraph $\{[(1,2),f], [(1,3,4),g]\}$ representing the product $f_{12}g_{134}$.

the same flavor ϕ to every edge of H. For instance, the product (3.3) is naturally represented by the flavored hypergraph $H^{(f)} = \{(E,f), E\in H\} \in \mathscr{H}(S;f)$ [Fig. 3(a)].

An m-flavored hypergraph $H\in\mathscr{H}(S;\phi^1,\phi^2,\dots,\phi^m)$ can always be written as a union

$$H = H_1^{(\phi^1)} \cup H_2^{(\phi^2)} \cup \cdots \cup H_m^{(\phi^m)} \tag{3.5}$$

of 1-flavored hypergraphs $H_i^{(\phi^i)} \in \mathbb{H}(S;\phi^i)$. We have, e.g.,

$$\mathbb{H}(S;\phi^1,\phi^2) = \{H_1\cup H_2; H_1\in\mathbb{H}(S;\phi^1), H_2\in\mathbb{H}(S;\phi^2)\}. \tag{3.6}$$

Given two families $\{f_E, E\subset S\}$ and $\{g_E, E\subset S\}$ of objects labeled by subsets of S, and a hypergraph $H = H_1^{(\phi^1)}\cup H_2^{(\phi^2)} \in\mathscr{H}(S;\phi^1,\phi^2)$, we denote [Fig. 3(b)]

$$H^{f,g} \equiv H_1^f H_2^g = \prod_{E\in H_1} f_E \prod_{E'\in H_2} g_{E'}. \tag{3.7}$$

A flavored hypergraph is *simple* if no two of its edges are identical; i.e., two edges can have identical values, or identical flavors, but *not* both [thus H in (3.5) is simple iff each H_i, $i = 1\cdots m$, is simple].

To illustrate the notation, let $\{u_E, E\subset \underline{N}\}$ and $\{C_E, E\subset\underline{N}\}$ be two families of objects labeled by subsets of $\underline{N} = \{1,2,\dots,N\}$, and denote

$$f_E = e^{u_E} - 1, \quad g_E = e^{C_E} - 1. \tag{3.8}$$

We have

$$\exp\left(\sum_{E\subset\underline{N}} u_E\right) = 1 + \sum_{H\in\mathbf{H}(\underline{N})} (1/H!)H^u \tag{3.9}$$

$$= \prod_{E\subset\underline{N}} (1+f_E) = 1 + \sum_{H\in\mathbf{H}_s(\underline{N})} H^f, \tag{3.10}$$

where [see (1.5)]

$$H! = \prod_{E\subset\underline{N}} (m_E!), \tag{3.11}$$

where m_E is the number of times $E\subset\underline{N}$ is contained in H. Note that in (3.10), the sum is over *simple* hypergraphs. We deduce

$$\exp\left[\sum_{E\subset\underline{N}} (u_E + C_E)\right] = 1 + \sum_{H\in\mathbf{H}_s(\underline{N})} (H^f + H^g) + \sum_{H\in\mathbf{H}_s(\underline{N};f,g)} H^{f,g} \tag{3.12}$$

from expanding both $e^{\Sigma u}$ and $e^{\Sigma C}$ in the manner (3.10).

Let

$$V(H)\cup H = \{h_1, h_2, \dots, h_n\} \tag{3.13}$$

so that each h_i stands for either an edge or a vertex. We write h_i—h_j if h_i is a vertex and h_j an edge containing it, or vice versa (i.e., one is a dot, the other a circle joined to it by a line); h_i and h_j are then said to be *incident* on each other. The *degree* $|h_i|$ is the number of elements of $V(H)\cup H$ incident on h_i.[23] Two edges (vertices) are *adjacent* if there is a vertex (edge) incident on both. A *path* is a sequence of *distinct* elements h_1—h_2—⋯—h_k, each incident on the preceding. A *cycle* is a closed path $\overbrace{h_1\text{—}h_2\text{—}\cdots\text{—}h_k}$. We write h_i----h_j if there exists a path between h_i and h_j; this is clearly an equivalence relation, whose equivalence classes define the *connected components* of H. H is *connected* if it has a single connected component, i.e., if h_i----h_j for each pair (h_i, h_j). We write $E_1 \overset{(v)}{\text{----}} E_2$ if there exists a path between edges E_1 and E_2 not containing vertex v; $\overset{(v)}{\text{----}}$ is clearly an equivalence relation between edges.[24]

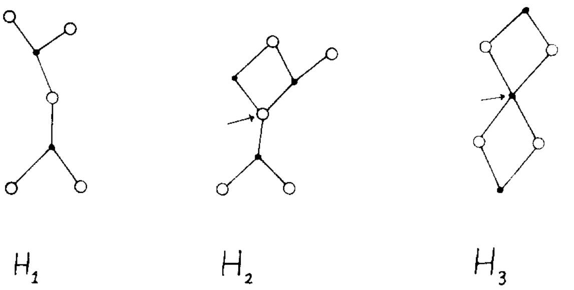


FIG. 4. H_1 is a tree, since every vertex (whence also every edge) is a pure articulation; H_2 is not irreducible since it contains an articulation vertex (arrow); H_3 is irreducible since it contains no articulation vertex [even though it contains an articulation edge (arrow)].

We shall speak of *removing* an edge or a vertex. This means that in the diagram of H, we remove the corresponding dot or circle, *and* the lines emanating from it.

Let H be connected. h_i is an *articulation* of order m, $m\geqslant 2$, if upon its removal, H breaks into m connected components; it is a pure *articulation* if $m = |h_i|$. h_i is dangling if $|h_i| = 1$. H is a tree if each h_i is either a pure articulation or dangling, or equivalently, if H contains no cycles. H is a *star*, or *irreducible*, or *2-(v) connected* if it contains no articulation vertex (it may however contain articulation edges) (Fig. 4).

3.1 Partitions of hypergraphs

A partition of a hypergraph H is a partition of the multiset H, i.e., a multiset $P(H) = \{H_j, j\in J\}$ of subhypergraphs $H_j \subset H$, such that $\Sigma_{j\in J} H_j = H$; the latter equality implies

$$\prod_{H_j\in P(H)} H_j^f = H^f. \tag{3.14}$$

A partition is illustrated as in Fig. 5, with dashed lines delineating the different parts.

We define a partition operator $\hat{P}$: hypergraphs→partitions, as an operator which, when acted on a hypergraph, partitions it according to some prescribed rule. We denote $\hat{P}_1 H$ the partition of any hypergraph H into its connected components, and $\hat{P}_2 H$ the partition into maximal irreducible blocks, or stars (i.e., maximal with respect to the property of possessing no articulation vertex). Given any vertex v, we

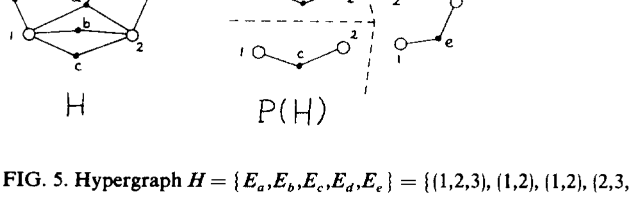


FIG. 5. Hypergraph $H = \{E_a, E_b, E_c, E_d, E_e\} = \{(1,2,3), (1,2), (1,2), (2,3,4), (1,5)\}$ and partition $P(H) = \{(E_a,E_b), (E_c), (E_d,E_e)\}$.

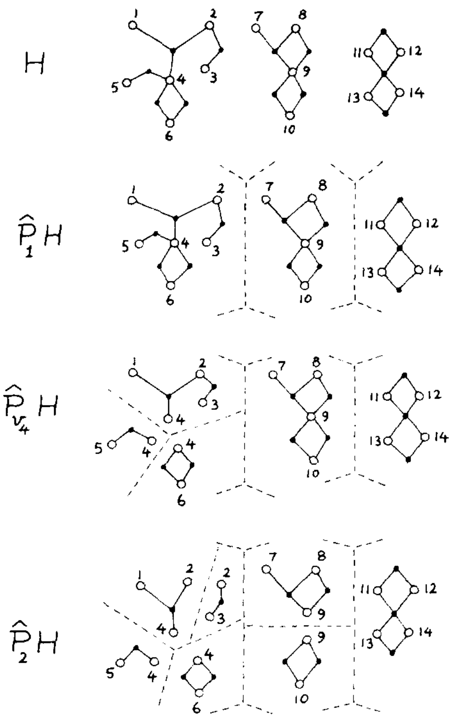


FIG. 6. The partition operators $\hat{P}_1$, $\hat{P}_v$ and $\hat{P}_2$.

denote $\hat{P}_v H$ the partition of H into the equivalence classes of the relation $\overset{(v)}{---}$ between edges. Examples of the above partition operations are given in Fig. 6.

The following two lemmas are proved in Appendix A:
Lemma 3.1: Given any hypergraph H, we have

$$\sum_{E\in H} (|E| - 1) \geqslant |V(H)| - |\hat{P}_1 H|, \tag{3.15}$$

and dually,

$$\sum_{v\in V(H)} (|v| - 1) \geqslant |H| - |\hat{P}_1 H|, \tag{3.16}$$

the equalities holding iff H is a *forest* (i.e., if each connected component of H is a tree).

Remark: $|H|$ is the number of edges, $|V(H)|$ the number of vertices, and $|\hat{P}_1 H|$ the number of connected components of H.

Lemma 3.2: Let H be any hypergraph, and $P(H) = \{H_j, j\in J\}$ any partition of H. We have

$$\sum_{H_j\in P(H)} (|V(H_j)| - |\hat{P}_1 H_j|) \geqslant |V(H)| - |\hat{P}_1 H|, \tag{3.17}$$

the equality holding iff $P(H)$ is a superpartition of $\hat{P}_2 H$.

3.2 Cumulants of hypergraphs

We shall meet cumulants of products over hypergraphs, i.e., quantities of the form

$$\langle H^f\rangle_{c\{f\}} = \Big\langle \prod_{E\in H} f_E \Big\rangle_{c\{f\}}$$

$$= \sum_{P\in\pi(H)} (-)^{|P|-1}(|P|-1)! \prod_{H_j\in P} \langle H_j^f\rangle, \tag{3.18}$$

where H is any hypergraph [$\pi(H)$ is the multiset of all partitions of H (see Sec. 1.1)].

Let the partition of H into its irreducible blocks be

$\hat{P}_2 H = \{S_j, j\in J\}$.

We denote by[25]

$$\langle H^f\rangle_{c\{f,\mathrm{irr}\}} \equiv \Big\langle \prod_{S_j\in\hat{P}_2 H} S_j \Big\rangle_{c\{S\}} \tag{3.19}$$

$$= \sum_{P\in\pi(H), P>\hat{P}_2 H} (-)^{|P|-1}(|P|) - 1)! \prod_{H_i\in P} \langle H_i^f\rangle, \tag{3.20}$$

the cumulant built on the irreducible blocks of H [(3.20) differs from (3.18) in that the sum over partitions is restricted to superpartitions of $\hat{P}_2 H$].

Lemma 3.3: Let the stochastic variables f_E be such that

$$\langle H^f\rangle \equiv \Big\langle \prod_{E\in H} f_E \Big\rangle = \mathrm{const} \times \epsilon^{|V(H)| - |\hat{P}_1 H|} \tag{3.21}$$

for all hypergraphs H, where ϵ is some (small) parameter. We then have

$$\langle H^f\rangle_{c\{f\}} = \langle H^f\rangle_{c\{f,\mathrm{irr}\}} + (\mathrm{remainder}), \tag{3.22}$$

where

$$\langle H^f\rangle_{c\{f,\mathrm{irr}\}} = (\mathrm{const})\epsilon^{|V(H)| - |\hat{P}_1 H|} \tag{3.23}$$

and (remainder) is of order higher than $|V(H)| - |\hat{P}_1 H|$ in ϵ.

Proof: Immediate from Lemma 3.2. Note that (remainder) consists of the terms P in (3.18) which are not superpartitions of $\hat{P}_2 H$.

4. STATISTICAL MECHANICAL FORMULAS[1,13]

We consider a gas of N particles in a volume $\mathscr{V}$ at temperature T. The (canonical) partition function is defined as

$$Z(T, \mathscr{V}, N) = \mathrm{Tr}\, e^{-\beta H}, \tag{4.1}$$

where H is the Hamiltonian of the gas, and $\beta^{-1} = kT$, where k is Boltzmann's constant. The trace Tr is over symmetrized states, symmetric for bosons, antisymmetric for fermions. The free energy is

$$A(T, \mathscr{V}, N) = -kT \ln Z(T, \mathscr{V}, N) \tag{4.2}$$

and the equation of state of the gas is given by

$$P = -\partial A/\partial\mathscr{V} = n^2\partial a/\partial n, \tag{4.3}$$

where P is the pressure,

$$n = N/\mathscr{V} \tag{4.4}$$

is the particle number density, and

$$a = A/N \tag{4.5}$$

is the free energy per particle.

The grand partition function is

$$Z_{\mathrm{gr}}(T, \mathscr{V}, \mu) = \sum_{N=0}^{\infty} z^N Z(T, \mathscr{V}, N), \tag{4.6}$$

where μ is the chemical potential, and

$$z = e^{\beta\mu} \tag{4.7}$$

is called the fugacity. The grand thermodynamic potential

$$J(T, \mathscr{V}, \mu) = kT \ln Z_{\mathrm{gr}}(T, \mathscr{V}, \mu). \tag{4.8}$$

We have

$$P\mathscr{V} = J(T, \mathscr{V}, \mu), \tag{4.9}$$

$$n\mathscr{V} = \partial J / \partial \mu. \tag{4.10}$$

The equation of state must here be deduced by eliminating μ from (4.9) in favor of n by use of (4.10).

We are interested in the virial expansion

$$P/(nkT) = 1 + \sum_{j=1}^{\infty} n^j B_{j+1}, \tag{4.11}$$

i.e., the expansion of the pressure in powers of the density. The virial coefficients B_j are functions of the temperature, and their explicit form is our main object. The virial expansion may be arrived at by two different routes: the more usual[1–10] consists in first obtaining an expansion in powers of z for the grand thermodynamic potential (4.8), and then deducing (4.11) by use of (4.9) and (4.10). The other method is to consider the free energy (4.5) in the thermodynamic limit $N \to \infty$, $\mathscr{V} \to \infty$ with $N/\mathscr{V} = n$, and deduce its density expansion

$$-\beta a = -\beta a^0 + \sum_{k=1}^{\infty} n^k \overline{B}_{k+1}, \tag{4.12}$$

where $a^0 = kT \ln(n)$ is the free energy for the classical ideal gas. In view of (4.3), there follows

$$B_j = -(j-1)\overline{B}_j. \tag{4.13}$$

This latter method has heretofore only been applied to the classical gas.[11–14] We shall here apply it to the quantum case, and thereby obtain new expressions for the virial coefficients, much more meaningful than those obtained via the grand canonical formalism.

5. QUANTUM BOLTZMANN GAS

In this section, we consider a gas of particles obeying quantum dynamics, but Boltzmann statistics. The Hamiltonian is taken as

$$H = K + U, \tag{5.1}$$

where

$$K = \sum_{i=1}^{N} K_i \tag{5.2}$$

is the sum of kinetic energy operators, and U is the interaction potential. We take U of the form

$$U = \sum_{i<j} U_{ij} + \sum_{i<j<k} U_{ijk} + \cdots \tag{5.3}$$

$$= \sum_{E \subset \underline{N}} U_E, \tag{5.3'}$$

where the $U_{ij \cdots m} = U(\vec{r}_i, \vec{r}_j, ..., \vec{r}_m)$ are assumed to have the cluster property of vanishing when $\mathrm{Max}|\vec{r}_i - \vec{r}_j|$ becomes larger than some distance; $\vec{r}_i$ is the position coordinate of the ith particle. In (5.3'), the sum is over all subsets of $\underline{N} = \{1,2,...,N\}$, and we define

$$U_E = 0 \quad \text{if } |E| < 2. \tag{5.4}$$

The partition function is

$$Z(T, \mathscr{V}, N) = (N!)^{-1} \mathrm{tr}\, e^{-\beta H}, \tag{5.5}$$

where the trace tr (small t) is over *unsymmetrized* states. Equation (5.5) is the high temperature and/or low density limit of the fully quantum expression (4.1); the $(N!)^{-1}$ is a remnant of the quantum statistics ("correct Boltzmann counting").[26]

We rewrite (5.5) as

$$Z(T, \mathscr{V}, N) = (N!)^{-1} \mathrm{tr}\left\{ e^{-\beta K} T_{\leftarrow} \exp\left[-\int_0^{\beta} d\tau \sum_{E \subset \underline{N}} U_E(\tau) \right] \right\} \tag{5.6}$$

$$= Z_N^{(0)} \left\langle \exp\left(\sum_{E \subset \underline{N}} u_E \right) \right\rangle, \tag{5.6'}$$

where

$$U_E(\tau) = e^{\tau K} U_E e^{-\tau K} \tag{5.7}$$

and $T_{\leftarrow}$ orders the $U_E(\tau)$ such that the "imaginary times" τ increase from right to left. In (5.6'), we introduced the "averaging" operation ($\langle 1 \rangle = 1$)

$$\langle (\cdots) \rangle = (\mathrm{tr}\, e^{-\beta K})^{-1} \mathrm{tr}\{ e^{-\beta K} T_{\leftarrow} (\cdots) \} \tag{5.8}$$

and denote

$$u_E = -\int_0^{\beta} d\tau\, U_E(\tau). \tag{5.9}$$

As to $Z_N^{(0)}$, it is the ideal Boltzmann gas partition function:

$$Z_N^{(0)} = (N!)^{-1} \mathrm{tr}\, e^{-\beta K} = (N!)^{-1} (\mathscr{V}/\lambda^3)^N, \tag{5.10}$$

where

$$\lambda = (2\pi\hbar^2 \beta / m)^{1/2} \tag{5.11}$$

is the thermal wavelength (m the mass of the particles).

Under the protection of $T_{\leftarrow}$ contained in $\langle\ \rangle$, the $U_E(\tau)$ can be treated as commuting variables. It is understood that $T_{\leftarrow}$ in (5.8) does not act outside $\langle\ \rangle$, i.e., it orders only the operators inside $\langle\ \rangle$.

The excess free energy per particle is, in view of (4.2), (4.5), and (2.1c),

$$-\beta(a - a^{(0)}) = N^{-1} \ln \langle e^{\Sigma_{E \subset \underline{N}} u_E} \rangle \tag{5.12}$$

$$= N^{-1} \langle e^{\Sigma u_E} - 1 \rangle_{c\{u\}}, \tag{5.13}$$

where $a^{(0)} = -N^{-1} kT \ln Z_N^{(0)} = \ln(n/\lambda^3)$ is the ideal gas free energy per particle. We now introduce (3.8)–(3.10) into (5.13). Noticing that

$$\langle H_1^u H_2^u \rangle = \langle H_1^u \rangle \langle H_2^u \rangle \tag{5.14}$$

if the hypergraphs H_1 and H_2 have no common vertices (i.e., particles), we deduce, in view of the cluster property Lemma 2.1 of cumulants,

$$\langle H^u \rangle_{c\{u\}} = \langle H^f \rangle_{c\{u\}} = 0 \tag{5.15}$$

if H is not connected. We thus get

$$-\beta(a - a^{(0)}) = N^{-1} \sum_{k=1}^{N} \binom{N}{k} \sum_{H \in \mathscr{H}_{s,c}(\underline{k})} \langle H^f \rangle_{c\{u\}} \tag{5.16}$$

$$= N^{-1} \sum_{k=1}^{N} \binom{N}{k} \sum_{H \in \mathscr{H}_c(\underline{k})} (1/H!) \langle H^u \rangle_{c\{u\}}, \tag{5.16'}$$

where $\mathscr{H}_c(\underline{k})$ is the set of all *connected* hypergraphs with vertex set $\underline{k} = \{1,2,...,k\}$, $\mathscr{H}_{s,c}(\underline{k})$ the set of all *simple* such hypergraphs, and $\binom{N}{k} = N!/[k!(N-k)!]$ are binomial coefficients.

Now, given any *connected* hypergraph H with k vertices $[V(H) = \underline{k}]$, we have

$$\begin{aligned}\langle H^u\rangle &= (\lambda/\mathscr{V})^k \operatorname{tr}_{V(H)} e^{-\beta K} T_{\leftarrow} H^u \\ &= (\lambda/\mathscr{V})^k \int_{\mathscr{V}} d\vec{r}_1 \cdots d\vec{r}_k \langle \vec{r}_1 \cdots \vec{r}_k | e^{-\beta K} T_{\leftarrow} H^u | \vec{r}_1 \cdots \vec{r}_k\rangle \\ &= \text{const} \times (1/\mathscr{V})^{k-1}. \end{aligned} \tag{5.17}$$

We used the fact that $\langle \vec{r}_1 \ldots \vec{r}_k | e^{-\beta K} T_{\leftarrow} H^u | \vec{r}_1 \ldots \vec{r}_k \rangle$ vanishes unless the coordinates $\vec{r}_1 \ldots \vec{r}_k$ are all clustered together (because H is connected and the u_E have the cluster property); thus the trace over $k-1$ of the particles yields a finite result independent of the volume $\mathscr{V}$ and of the coordinates of the last particle; the trace over the latter yields a factor $\mathscr{V}$. Given then *any* hypergraph H with c connected components, i.e., $\hat{P}_1 H = \{H_1, H_2, \ldots, H_c\}$, we have

$$\begin{aligned}\langle H^u\rangle = \left\langle \prod_{i=1}^{c} H_i^u \right\rangle &= \prod_{i=1}^{c} \langle H_i^u \rangle \\ &= \text{const} \times (1/\mathscr{V})^{|V(H)| - |\hat{P}_1 H|}, \end{aligned} \tag{5.18}$$

where we applied (5.17) to each $H_i \in \hat{P}_1 H$ and used $\Sigma_{i=1}^{c}(|V(H_i)| - 1) = |V(H)| - c$.

We can now apply Lemma 3.3, with $\epsilon = (1/\mathscr{V})$, to evaluate (5.16) in the limit $N \to \infty$, $\mathscr{V} \to \infty$ with $N/\mathscr{V} = n$. We have, for any $H \in \mathscr{H}_c(\underline{k})$ [i.e., $|\hat{P}_1 H| = 1$ and $|V(H)| = k$]

$$\langle H^u \rangle_{c\{u\}} = \langle H^u \rangle_{c\{u,\text{irr}\}} + (\text{remainder}), \tag{5.19}$$

where

$$\langle H^u \rangle_{c\{u,\text{irr}\}} = \text{const} \times (1/\mathscr{V})^{k-1} \tag{5.20}$$

and (remainder) is of higher order in $(1/\mathscr{V})$. Thus, when $\langle H^u \rangle_{c\{u\}}$ is multiplied by $N^{-1}\binom{N}{k} \cong N^{k-1}/k!$, and the limit N, $\mathscr{V} \to \infty$ taken, only the first term of (5.19) survives. On then comparing (5.16) with (4.12), we deduce

$$\begin{aligned}\overline{B}_k^{\text{Boltzmann}} \\ &= (k!)^{-1} \sum_{H \in \mathscr{H}_c(\underline{k})} (1/H!) \operatorname*{Lim}_{\mathscr{V} \to \infty} \mathscr{V}^{k-1} \langle H^u \rangle_{c\{u,\text{irr}\}}, \qquad (5.21') \\ &= (k!)^{-1} \sum_{H \in \mathscr{H}_{s,c}(\underline{k})} \operatorname*{Lim}_{\mathscr{V} \to \infty} \mathscr{V}^{k-1} \langle H^f \rangle_{c\{f,\text{irr}\}}. \qquad (5.21'')\end{aligned}$$

The second line follows from the corollary to Lemma 2.2, because a given u_E cannot appear inside two different $\langle\ \rangle$ factors in any term of $\langle H^u \rangle_{c\{u,\text{irr}\}}$ [since two irreducible blocks of H share at most one vertex, and $|E| \geqslant 2$ (or else $u_E = 0$)].

Equations (5.21) constitute one of the main results of the paper. In (5.21′), we have a sum over u-edged connected hypergraphs (Feynman diagrams), while in (5.21″), the sum is over f-edged connected simple hypergraphs (Mayer diagrams); in either case, we have cumulants built on the irreducible blocks.

5.1 Classical case

In the classical case,

$$\begin{aligned} f_E &= e^{-\beta U_E} - 1, \\ \langle\langle \cdots \rangle\rangle &= \mathscr{V}^{-N} \int_{\mathscr{V}} d\vec{r}_1 \cdots d\vec{r}_N (\ldots). \end{aligned} \tag{5.22}$$

Here, $\langle H^f \rangle$ factorizes according to the irreducible blocks of H, i.e., if $\hat{P}_2 H = \{S_j, j = 1 \cdots s\}$,

$$\langle H^f \rangle = \left\langle \prod_{j=1}^{s} S_j^f \right\rangle = \prod_{j=1}^{s} \langle S_j^f \rangle; \tag{5.23}$$

thus the S_j^f are $\langle\ \rangle$-independent variables. To demonstrate (5.23), it suffices to show that $\langle H_1^f H_2^f \rangle = \langle H_1^f \rangle \langle H_2^f \rangle$ if H_1 and H_2 are connected hypergraphs having a single common vertex. Let then $V(H_1) = \{1, 2, \ldots, j\}$, $V(H_2) = \{j, j+1, \ldots, j+m\}$, j being the common particle. We have

$$\begin{aligned} \langle H_1^f H_2^f \rangle \\ &= \mathscr{V}^{-(j+m)} \int_{\mathscr{V}} d\vec{r}_1 \cdots d\vec{r}_{j+m} H_1^f(\vec{r}_1 \cdots \vec{r}_j) H_2^f(\vec{r}_j \cdots \vec{r}_{j+m}) \\ &= \mathscr{V}^{-(j+m)} \int_{\mathscr{V}} d\vec{r}_j \mathscr{V}^{j-1} \langle H_1^f \rangle \mathscr{V}^m \langle H_2^f \rangle = \langle H_1^f \rangle \langle H_2^f \rangle, \end{aligned} \tag{5.24}$$

where we used $\int_{\mathscr{V}} d\vec{r}_j = \mathscr{V}$ and (with $V(H) = \underline{n}$)

$$\begin{aligned} \int_{\mathscr{V}} d\vec{r}_1 \cdots d\vec{r}_{n-1} H^f(\vec{r}_1 \cdots \vec{r}_n) \\ &= \mathscr{V}^{-1} \int d\vec{r}_1 \cdots d\vec{r}_n H^f = \mathscr{V}^{n-1} \langle H^f \rangle \end{aligned} \tag{5.25}$$

(at large $\mathscr{V}$, the lhs is independent of $\mathscr{V}$ and of $\vec{r}_n$ if H is connected).

The factorization (5.23) implies, in view of Lemma 2.1, that $\langle H^f \rangle_{c\{\text{irr}\}}$ vanishes if H has two or more irreducible blocks; i.e., it is nonzero only if H is itself irreducible, in which case $\langle H^f \rangle_{c\{\text{irr}\}} = \langle H^f \rangle$. One then recovers from (5.21″) the classical Mayer result (extended to multiparticle interactions)

$$\overline{B}_k^{\text{classical}} = (k!)^{-1} \sum_{H \in \mathscr{H}_{s,\text{irr}}(\underline{k})} \operatorname*{Lim}_{\mathscr{V} \to \infty} \mathscr{V}^{k-1} \langle H^f \rangle, \tag{5.26}$$

where $\mathscr{H}_{s,\text{irr}}(\underline{k})$ is the set of all simple irreducible hypergraphs with vertex set $\underline{k}$.

In the quantum case, the time-ordering operation $T_{\leftarrow}$ inside $\langle\ \rangle$ time entangles the different irreducible blocks of a diagram, and thereby prevents the factorization (5.23) to happen.

6. QUANTUM STATISTICS

We now incorporate quantum statistics. The quantum statistical partition function (4.1), wherein the trace Tr is over symmetrized states, may be expressed in terms of the trace tr over unsymmetrized states as[27]

$$Z = (N!)^{-1} \operatorname{tr} e^{-\beta H} S_{\underline{N}}, \tag{6.1}$$

where the symmetrizer or antisymmetrizer (acting on the set $\underline{N}$)

$$S_{\underline{N}} = \sum_{\mathscr{P}} \epsilon^{|\mathscr{P}|} \mathscr{O}_{\mathscr{P}}. \tag{6.2}$$

The sum $\Sigma_{\mathscr{P}}$ is over all permutations of the N particles,

$$\epsilon = \begin{cases} 1 \text{ for bosons,} \\ -1 \text{ for fermions,} \end{cases} \tag{6.3}$$

and $|\mathscr{P}|$ is 0 or 1 according as the permutation $\mathscr{P}$ is even or odd. The operator $\mathscr{O}_{\mathscr{P}}$ permutes the particle coordinates in a wave function, i.e., if $\mathscr{P}$ is the permutation $\{1,2,...,N\}\rightarrow\{i_1,i_2,...,i_N\}$, then

$$\mathscr{O}_{\mathscr{P}}\,\Psi(\vec{r}_1,\vec{r}_2,...,\vec{r}_N)=\Psi(\vec{r}_{i_1},\vec{r}_{i_2},...,\vec{r}_{i_N}). \tag{6.4}$$

Let $(i_1,i_2,...,i_k)$ denote the cyclic permutation $\{i_1,i_2,...,i_k\}\rightarrow\{i_k,i_1,...,i_{k-1}\}$. The parity of $(i_1,i_2,...,i_k)$ is $(k-1)\bmod 2$, since it takes $k-1$ transpositions to displace i_k from the last to the first position. We denote $\mathscr{C}(E)$ the set of distinct cycles involving the set of indices $E\subset\underline{N}$; there are $(|E|-1)!$ such cycles, i.e.,[28]

$$|\mathscr{C}(E)|=(|E|-1)!. \tag{6.5}$$

E.g., $\mathscr{C}(\underline{3})=\{(1,2,3),(1,3,2)\}$. We denote

$$\begin{aligned} C_E &= \epsilon^{|E|-1}\sum_{c\in\mathscr{C}(E)}\mathscr{O}_c \quad \text{if } |E|\geqslant 2,\\ &=0 \qquad\qquad\qquad\qquad \text{if } |E|<2. \end{aligned} \tag{6.6}$$

Since every permutation on $\underline{N}$ can be written as a product of cycles, with each index $i\in\underline{N}$ appearing *at most once* in the product [e.g.,

$$\begin{pmatrix}123456\\321564\end{pmatrix}=(13)(2)(465)],$$

we have the following equivalent expressions for $S_{\underline{N}}$:

$$S_{\underline{N}}=\sum_{P\in\pi(\underline{N})}\prod_{E\in P}C_E \tag{6.7}$$

$$=L\prod_{E\subset\underline{N}}(1+C_E)=L\exp\Bigl(\sum_{E\subset\underline{N}}C_E\Bigr). \tag{6.8}$$

In (6.7), the sum is over all partitions of the set $\underline{N}$. In (6.8), L is the "leveling" operator (see Sec. 2.1), which suppresses all terms wherein any index appears more than once [e.g., $L\,C_{12}C_{345}=C_{12}C_{345}$ but $L\,C_{12}C_{234}=0$]; we used

$$L\,e^{C_E}=1+C_E. \tag{6.9}$$

On comparing (6.7) with (2.3) and (2.10), we deduce that

$$C_E=S_E{}^c. \tag{6.10}$$

Example: $C_{\underline{2}}=\epsilon\mathscr{O}_{(1,2)}$, $C_{\underline{3}}=\epsilon^2[\mathscr{O}_{(1,2,3)}+\mathscr{O}_{(1,3,2)}]$ and $S_{\underline{3}}=L\exp(C_{12}+C_{13}+C_{23}+C_{123}+C_{132})=1+C_{12}+C_{13}+C_{23}+C_{123}+C_{132}$ (all other terms in the expanded exponential contain repeated indices, and are killed by L).

Note that under the protection of L, permutators with common arguments can be treated as commuting variables, since L acting on their product will yield zero anyhow.

The operators C_E may be regarded as effective "exchange interactions." These multiparticle interactions do not have the cluster property; for instance

$$\langle\vec{r}_1',\vec{r}_2',\vec{r}_3'|\mathscr{O}_{(123)}|\vec{r}_1,\vec{r}_2,\vec{r}_3\rangle =\delta(\vec{r}_1'-\vec{r}_3)\delta(\vec{r}_2'-\vec{r}_1)\delta(\vec{r}_3'-\vec{r}_2) \tag{6.11}$$

is nonzero if

$$\vec{r}_1'=\vec{r}_3,\quad \vec{r}_2'=\vec{r}_1,\quad \vec{r}_3'=\vec{r}_2, \tag{6.12}$$

which can be even if $\vec{r}_1$, $\vec{r}_2$, and $\vec{r}_3$ are arbitrarily far from one another. However, $\mathscr{O}_{(123)}$ has a partial cluster property, in the sense that if in (6.11) each $\vec{r}_i'$ is close to $\vec{r}_i$, i.e., if

$$\operatorname{Max}|\vec{r}_i'-\vec{r}_i|\leqslant\lambda, \tag{6.13}$$

where λ is some finite distance, then (6.12) implies that the $\vec{r}_i$'s and $\vec{r}_i'$'s must all be clustered together for (6.11) not to vanish. Now, in

$$\operatorname{tr}e^{-\beta H}S=\int d\vec{r}_1\cdots d\vec{r}_N d\vec{r}_1'\cdots d\vec{r}_N'\,\langle\vec{r}_1\cdots\vec{r}_N|e^{-\beta H}|\vec{r}_1'\cdots\vec{r}_N'\rangle\langle\vec{r}_1'\cdots\vec{r}_N'|S|\vec{r}_1\cdots\vec{r}_N\rangle$$

the propagator $\langle\vec{r}_1\cdots\vec{r}_N|e^{-\beta H}|\vec{r}_1'\cdots\vec{r}_N'\rangle$ vanishes when $\operatorname{Max}|\vec{r}_i'-\vec{r}_i|$ is much larger than the thermal wavelength λ; thus condition (6.13) is satisfied in our case, and the interactions C_E *effectively have* the cluster property (at finite temperature).

Using (5.7) and (6.8), we rewrite (6.1) as [compare (5.6)]:

$$\begin{aligned} Z(T,\mathscr{V},N)&=(N!)^{-1}\operatorname{tr}\{e^{-\beta K}(T_{\leftarrow}e^{\Sigma_{E\subset\underline{N}}u_E})(Le^{\Sigma_{E\subset\underline{N}}C_E})\}\\ &=Z^{(0)}\Bigl\langle\exp\Bigl[\sum_{E\subset\underline{N}}(u_E+C_E)\Bigr]\Bigr\rangle, \end{aligned} \tag{6.14}$$

where $Z^{(0)}$ is again the ideal Boltzmann gas partition function, and here the averaging operation

$$\langle(\cdots)\rangle=(\operatorname{tr}e^{-\beta K})^{-1}\operatorname{tr}\{e^{-\beta K}\Lambda(\cdots)\}, \tag{6.15}$$

where the operator Λ does three things: (i) it puts all C_E's to the right of all u_E's, (ii) it time orders the U_E's (iii) it suppresses all terms wherein two or more C_E's have common indices (i.e., Λ acts $T_{\leftarrow}$ on the u_E's and L on the C_E's). It is understood that Λ acts only *inside* $\langle\ \rangle$.

Expression (6.14) is nearly identical to (5.6'), the differences being that u_E is replaced by u_E+C_E and $\langle\ \rangle$ is generalized. Since the two properties we used to get (5.21), viz. $\langle\ \rangle$-independence of vertex-disjoint hypergraphs and cluster property of interactions, are again present here, we obtain anew (5.21), but with f_E replaced by

$$F_E=e^{u_E}(1+C_E)-1 \tag{6.16}$$

[more precisely, $F_E=e^{u_E+C_E}-1$; but inside $\langle H^F\rangle_{c\{F,\text{irr}\}}$, F_E becomes (6.16) in view of (6.9)].

Equation (5.21) with $\langle\ \rangle$ given by (6.15) and f_E replaced by F_E is the solution to our problem. An alternative form follows from (3.12):

$$\overline{B}_k=\overline{B}_k^{\text{Boltzmann}}+\overline{B}_k^{\text{ideal}}+\overline{B}_k^{\text{corr}}, \tag{6.17}$$

where $\overline{B}_k^{\text{Boltzmann}}$ is given by (5.21), $\overline{B}_k^{\text{ideal}}$ is given by (5.21) with f_E replaced by C_E, and

$$\overline{B}_k^{\text{corr}}=(k!)^{-1}\sum_{H\in\mathscr{H}_{s,c}(\underline{k};f,C)}\operatorname*{Lim}_{\mathscr{V}\rightarrow\infty}\mathscr{V}^{k-1}\langle H^{f,C}\rangle_{c\{f,C,\text{irr}\}}, \tag{6.18}$$

where $\mathscr{H}_{s,c}(\underline{k};f,C)$ is the set of all simple connected 2-flavored hypergraphs with vertex set $\underline{k}$. The physical significance of $\overline{B}_k^{\text{Boltzmann}}$ and $\overline{B}_k^{\text{ideal}}$ is obvious; as to $\overline{B}_k^{\text{corr}}$, it contains, so to speak, the correlations between dynamics and statistics.

The physically pleasing decomposition (6.17) can be effected directly on the free energy ln Z, before the limit N, $\mathscr{V}\to\infty$ is taken, which may be useful for certain purposes (e.g., dealing with Bose–Einstein condensation); we have, from (6.14) and (2.1c),

$$\begin{aligned}\ln Z(N,\mathscr{V},T) &= \ln Z^{(0)} + \Big\langle \exp\Big[\sum_{E\subset \underline{N}} (u_E + C_E)\Big]\Big\rangle_{c\{u,C\}} \\ &= \ln Z^{(0)} + \langle e^{\Sigma u_E} - 1\rangle_{c\{u\}} + \langle e^{\Sigma C_E} - 1\rangle_{c\{C\}} \\ &\quad + \langle (e^{\Sigma u_E} - 1)(e^{\Sigma C_E} - 1)\rangle_{c\{u,C\}}, \qquad (6.19)\end{aligned}$$

where we used the identity $(e^{a+b}-1) = (e^a - 1) + (e^b - 1) + (e^a - 1)(e^b - 1)$. The four terms in (6.19) are the ideal Boltzmann, quantum Boltzmann, ideal quantum, and dynamics-statistics correlations, respectively.

Remarks: 1. As already mentioned, the time ordering $T_{\leftarrow}$ inside $\langle\ \rangle$ prevents different irreducible blocks in a product H^f from being $\langle\ \rangle$-independent (by time entangling them). The leveling operator L in (6.15) has a similar effect; e.g., $LC_{12}C_{23} = 0$, so that $\langle C_{12}C_{23}\rangle = 0 \neq \langle C_{12}\rangle\langle C_{23}\rangle$ and $\langle C_{12}C_{23}\rangle_{c\{\mathrm{irr}\}} = -\langle C_{12}\rangle\langle C_{23}\rangle \neq 0$ (beware that L operates only *inside* $\langle\ \rangle$).[29]

2. The time ordered exponentials contained in (5.21), (6.17), etc., may be expanded in the interactions $U_E(\tau)$, leading to imaginary-time Feynman diagrams; or, one may revert to ordinary exponentials, e.g.,

$$\begin{aligned}&\mathrm{tr}\, e^{-\beta K} T_{\leftarrow} f_{12} f_{13} \\ &\quad = \mathrm{tr}\, e^{-\beta K} T_{\leftarrow} (e^{u_{12}+u_{13}} - e^{u_{12}} - e^{u_{13}} + 1) \\ &\quad = \mathrm{tr}\big[e^{-\beta(K+U_{12}+U_{13})} - e^{-\beta(K+U_{12})} \\ &\qquad - e^{-\beta(K+U_{13})} + e^{-\beta K}\big].\end{aligned}$$

6.1 Ideal quantum gas

Let us consider the ideal gas term

$$\overline{B}_k^{\mathrm{ideal}} = (k!)^{-1} \sum_{H\in\mathscr{H}_{s,c}(\underline{k})} \operatorname*{Lim}_{\mathscr{V}\to\infty} \mathscr{V}^{k-1} \langle H^C\rangle_{c\{C,\mathrm{irr}\}}, \qquad (6.20)$$

wherein

$$\langle(\cdots)\rangle = (\mathrm{tr}\, e^{-\beta K})^{-1} \mathrm{tr}\{e^{-\beta K} L(\cdots)\}. \qquad (6.21)$$

Because of the leveling L in (6.21), the only hypergraphs which contribute to (6.20) are those whose irreducible blocks all consist of a single edge (all other irreducible blocks are obviously killed by L), i.e., *trees*. Thus

$$\overline{B}_k^{\mathrm{ideal}} = (k!)^{-1} \sum_{T\in\mathscr{T}(\underline{k})} \operatorname*{Lim}_{\mathscr{V}\to\infty} \mathscr{V}^{k-1} \langle T^C\rangle_{c\{C\}}, \qquad (6.22)$$

where $\mathscr{T}(\underline{k})$ is the set of all trees with vertex set $\underline{k}$; we noticed that $\langle T^C\rangle_{c\{C,\mathrm{irr}\}} = \langle T^C\rangle_{c\{C\}}$ since each edge of T is an irreducible block by itself.

Now, $\langle T^C\rangle_{c\{C\}}$ consists of products $\Pi_{T_j\in P(T)}\langle \Pi_{E\in T_j} C_E\rangle$ [see (2.3)]; but because of L, $\langle \Pi_{E\in T_j} C_E\rangle \neq 0$ only if the edges $E\in T_j$ are all disjoint from one another, in which case $\langle \Pi_{E\in T_j} C_E\rangle = \Pi_{E\in T_j}\langle C_E\rangle$. It follows that

$$\langle T^C\rangle_{c\{C\}} = \alpha(T) \prod_{E\in T} \langle C_E\rangle, \qquad (6.23)$$

where $\alpha(T)$ is an integer whose value depends on the tree T. In Appendix B it is shown that[30]

$$\alpha(T) = (-)^{|T|-1} \prod_{v\in V(T)} (|v| - 1)! \qquad (6.24)$$

($|T|$ is the number of edges of T, and $|v|$ the number of edges incident on the vertex v). We thus have

$$\overline{B}_k^{\mathrm{ideal}} = (k!)^{-1} \sum_{T\in\mathscr{T}(\underline{k})} \alpha(T) \operatorname*{Lim}_{\mathscr{V}\to\infty} \mathscr{V}^{k-1} \prod_{E\in T} \langle C_E\rangle. \qquad (6.25)$$

In view of (6.5),

$$\begin{aligned}\langle C_{\underline{n}}\rangle &= \epsilon^{n-1}(n-1)!\langle \mathscr{O}_{(1,2,\ldots,n)}\rangle \\ &= \epsilon^{n-1}(n-1)!\Big(\sum_{\mathbf{k}} e^{-\beta\hbar^2k^2/2m}\Big)^{-1}\Big(\sum_{\mathbf{k}} e^{-n\beta\hbar^2k^2/2m}\Big) \\ &= \epsilon^{n-1}(n-1)!(\mathscr{V}/\lambda^3)^{-(n-1)} n^{-3/2}, \qquad (6.26)\end{aligned}$$

where we used

$$\begin{aligned}&\mathrm{tr}_{\underline{n}}\, e^{-\beta K} \mathscr{O}_{(1,2,\ldots,n)} \\ &\quad = \sum_{\mathbf{k}_1,\mathbf{k}_2,\ldots,\mathbf{k}_n} e^{-\beta\hbar^2(k_1^2+k_2^2+\cdots+k_n^2)/2m} \langle \vec{k}_n|\vec{k}_1\rangle\langle\vec{k}_1|\vec{k}_2\rangle \\ &\qquad \cdots\langle\vec{k}_{n-1}|\vec{k}_n\rangle = \sum_{\mathbf{k}} e^{-\beta n\hbar^2k^2/2m} \qquad (6.27)\end{aligned}$$

(since $\langle\vec{k}'|\vec{k}\rangle = \delta_{\vec{k}\vec{k}'}$), and $\Sigma_{\mathbf{k}} e^{-n\beta\hbar^2k^2/2m} = (\mathscr{V}/\lambda^3)n^{-3/2}$ in the large volume limit. Thus, noting that $\Sigma_{E\in T}(|E|-1) = |V(T)| - 1$ by Lemma 3.1, we finally have (here $|V(T)| = k$)

$$\begin{aligned}\overline{B}_k^{\mathrm{ideal}} &= \epsilon^{k-1}\lambda^{3(k-1)}(k!)^{-1} \sum_{T\in\mathscr{T}(\underline{k})} (-)^{|T|-1} \\ &\quad \times \prod_{E\in T} |E|^{-3/2}(|E|-1)! \prod_{v\in V(T)} (|v|-1)!. \qquad (6.28)\end{aligned}$$

In view of (6.10), the quantities $\langle C_E\rangle$ are essentially the Ursell cluster functions for the ideal quantum gas, so that (6.25) is in fact the expression of $\overline{B}_k^{\mathrm{ideal}}$ in terms of Ursell clusters. One can readily infer that (6.25) also holds in the nonideal case, provided the $\langle C_E\rangle$ are replaced by the nonideal Ursell cluster functions. This will now be shown explicitly, as a further illustration of the use of hypergraph-cumulant methods.

7. VIRIAL COEFFICIENTS IN TERMS OF URSELL CLUSTER FUNCTIONS

In this section, we express the $\overline{B}_k$'s in terms of Ursell cluster functions. We proceed by direct evaluation of the canonical partition function, without recourse to the grand canonical fugacity expansion as an intermediate step.

Let us denote by

$$Q_E = (\mathrm{tr}\, e^{-\beta H} S)_E = |E|!Z(E,\mathscr{V},T) \qquad (7.1)$$

the partition function for the gas consisting of the set E of particles, multiplied by $|E|!$. The Ursell cluster functions are the quantities Q_E^c obtained by the construction (2.10), i.e., if we formally denote

$$Q_E = \Big\langle \prod_{i\in E} X_i \Big\rangle, \qquad (7.2)$$

then

$$Q_E^c = \Big\langle \prod_{i\in E} X_i \Big\rangle_{c\{X\}} \tag{7.3}$$

Alternatively, if we collectively denote R_E the set of coordinates $\{\vec{r}_i, i\in E\}$ and write

$$Q_E = \int_{\mathscr{V}} dR_E \; W_E(R_E), \tag{7.4}$$

where

$$\begin{aligned} W_E(R_E) &= \langle R_E | e^{-\beta H_E} S_E | R_E \rangle \\ &= e^{-\beta U(R_E)} \quad \text{(classical case)}, \end{aligned} \tag{7.5}$$

then

$$Q_E^c = \int_{\mathscr{V}} dR_E \; W_E^c(R_E). \tag{7.6}$$

It is usually the W_E^c which are termed "cluster functions." Since $W_E(R_E)$ factorizes if the set of coordinates R_E separates into distant subsets, $W_E^c(R_E)$ correspondingly vanishes (Lemma 2.1), i.e., it has the cluster property; it follows that

$$Q_E^c \text{ is proportional to the volume } \mathscr{V}. \tag{7.7}$$

As is well known, Q_E^c can be expressed as the sum of all connected (generalized) Mayer diagrams with vertex set E.

We will now express the $\overline{B}_k$ in terms of the Q_E^c. We have, on using (7.2), (2.9), and Lemma 2.3,

$$\begin{aligned} Z_N &= (N!)^{-1}\langle X_1X_2\cdots X_N\rangle = (N!)^{-1}\langle X_1X_2\cdots X_N\rangle_{ca} \\ &= Z_N^{(0)} L \exp\Big(\sum_{E\subset \underline{N}} D_E\Big), \end{aligned} \tag{7.8}$$

where L is the leveling operator, and we defined

$$\begin{aligned} D_E &= 0 && \text{if } |E|<2 \\ &= (\lambda^3/\mathscr{V})^{|E|} Q_E^c && \text{if } |E|\geqslant 2, \quad (7.9) \\ &= \text{const}\times(1/\mathscr{V})^{|E|-1} && \quad (7.10) \end{aligned}$$

the last equality by (7.7) [we used $\langle X_i\rangle = (\lambda^3/\mathscr{V})$ and $(N!)^{-1}\Pi_{i\in\underline{N}}\langle X_i\rangle = Z_N^{(0)}$]. Introducing the "averaging operation"

$$\langle (\cdots)\rangle = L(\cdots) \tag{7.11}$$

(i.e., just "leveling"), we get from (7.8)

$$\begin{aligned} \ln(Z_N/Z_N^{(0)}) &= \Big\langle \exp\sum_{E\subset\underline{N}} D_E - 1 \Big\rangle_{c\{D\}} & (7.12) \\ &= \sum_{H\in \mathbb{H}(\underline{N})} (1/H!)\langle H^D\rangle_{c\{D\}}, & (7.13) \end{aligned}$$

where $\mathbb{H}(\underline{N})$ is the set of all hypergraphs with vertex set *contained* in $\underline{N}$. Since $\langle H_1^D H_2^D\rangle = \langle H_1^D\rangle\langle H_2^D\rangle$ if H_1 and H_2 have no common vertices, $\langle H^D\rangle_{c\{D\}}$ vanishes if H is not connected, and we can replace $\mathbb{H}(\underline{N})$ by $\mathbb{H}_c(\underline{N})$ the set of all *connected* hypergraphs with vertex set in $\underline{N}$. We thus get

$$-\beta(a-a^{(0)}) = N^{-1}\sum_{k=1}^{N}\binom{N}{k}\sum_{H\in\mathscr{H}_c(\underline{k})}(1/H!)\langle H^D\rangle_{c\{D\}}. \tag{7.14}$$

Now, because of (7.10), H^D and $\langle H^D\rangle_{c\{D\}}$ are proportional to

$$(1/\mathscr{V})^{\Sigma_{E\in H}(|E|-1)}. \tag{7.15}$$

It then follows from Lemma 3.1 that in the limit $N\rightarrow\infty$, $V\rightarrow\infty$, $\langle H^D\rangle_{c\{D\}}$ multiplied by $N^{-1}\binom{N}{k}\cong N^{k-1}/k!$, where $k=|V(H)|$, vanishes unless H is a tree. We thus obtain, on comparing (7.14) with (4.12),

$$\begin{aligned} \overline{B}_k &= (k!)^{-1}\sum_{T\in\mathscr{T}(\underline{k})}\operatorname*{Lim}_{\mathscr{V}\rightarrow\infty}\mathscr{V}^{k-1}\langle T^D\rangle_{c\{D\}} & (7.16) \\ &= (k!)^{-1}\sum_{T\in\mathscr{T}(\underline{k})}\alpha(T)\operatorname*{Lim}_{\mathscr{V}\rightarrow\infty}\mathscr{V}^{k-1}T^D. & (7.17) \end{aligned}$$

We noticed that $T! = 1$, since a tree has no multiple edges (of degree $\geqslant 2$), and we used $(\mathrm{B}1)^{31}$; $\alpha(T)$ is given in (6.24).

Equation (7.17) is the sought expression. To make contact with more standard notation,[1] let us denote

$$\mathscr{U}_E = \operatorname*{Lim}_{\mathscr{V}\rightarrow\infty}\mathscr{V}^{-1}\lambda^{3|E|}Q_E^c, \tag{7.18}$$

$$V_E = k!\,\overline{B}_{|E|}. \tag{7.19}$$

Equation (7.17) may then be rewritten

$$V_E = \sum_{T\in\mathscr{T}(E)}\alpha(T)T^{\mathscr{U}}. \tag{7.20}$$

The converse of this relation is the well-known Husimi relation[4,6]

$$\mathscr{U}_E = \sum_{T\in\mathscr{T}(E)} T^V. \tag{7.21}$$

To terminate, let us briefly show how the activity expansion of $\ln Z_{gr}$ is deduced by use of cumulants.[5] Let us formally denote

$$Q_{\underline{N}} = \langle X^N\rangle \tag{7.22}$$

(this notation is self-sufficient because $Q_{\underline{N}}$ depends only on $|\underline{N}| = N$). We then have

$$Z_{gr} = \sum_{N=0}^{\infty}(z^N/N!)Q_{\underline{N}} = \langle e^{zX}\rangle, \tag{7.23}$$

whence

$$\begin{aligned} \ln Z_{gr} &= \langle e^{zX}-1\rangle_{c\{X\}} = \sum_{N=1}^{\infty}(z^N/N!)\langle X^N\rangle_{c\{X\}} \\ &= \sum_{N=1}^{\infty}(z^N/N!)Q_{\underline{N}}^c, \end{aligned} \tag{7.24}$$

where we used the fact that $Q_E = Q_{E'}$ if $|E| = |E'|$ to write $\langle X^N\rangle_c = Q_{\underline{N}}^c$ (e.g., $\langle X^2\rangle_c = \langle X^2\rangle - \langle X\rangle^2 = Q_{\underline{2}} - Q_{\underline{1}}Q_{\underline{1}} = Q_{12} - Q_1Q_2 = Q_{\underline{2}}^c$). Equation (7.24) is the well-known activity expansion of the grand thermodynamic potential.

APPENDIX A: DEMONSTRATION OF LEMMAS 3.1 AND 3.2

We say that the connected hypergraph $H = \{E_1, E_2, \ldots, E_n\}$ is in *proper order* if

$$H^{(j)} \equiv \{E_1, E_2, \ldots, E_j\} \tag{A1}$$

is connected for each $j = 1,2,\ldots,n$. Proper order can always be achieved if H is connected (for choose E_1 arbitrarily; among the remaining edges, at least one is adjacent to E_1, or else H would not be connected; call it E_2; etc.) We shall always assume proper order. Obviously:

Lemma A.1: H is a tree iff $H^{(j)}$ and E_{j+1} have a single common vertex, i.e, $|V(H^{(j)})\cap E_{j+1}| = 1$, for each $j = 1,2,...,n-1$.

Given a hypergraph H, let us denote

$$\bar{\sigma}(H) \equiv \sum_{E\in H} (|E|-1). \tag{A2}$$

Lemma A.2: Let H be connected; we have

$$\bar{\sigma}(H) \geqslant |V(H)| - 1, \tag{A3}$$

the equality holding iff H is a tree.

Proof: Let $H = \{E_1, E_2, ..., E_n\}$ be properly ordered, and denote

$$b_j = \bar{\sigma}(H^{(j)}) - (|V(H^{(j)})| - 1).$$

We must show that $b_n \geqslant 0$, and $b_n = 0$ iff H is a tree. Clearly $b_1 = 0$. Let E_{j+1} have m_j common vertices with $H^{(j)}$; then $|V(H^{(j+1)})| = |V(H^{(j)})| + |E_{j+1}| - m_j$. It follows that

$$\begin{aligned} b_{j+1} &= b_j + (|E_{j+1}| - 1) - [\,|V(H^{(j+1)})| - |V(H^{(j)})|\,] \\ &= b_j + m_j - 1, \end{aligned}$$

whence $b_n = \Sigma_{i=1}^{n-1}(m_i - 1)$. Since H is connected, each $m_i \geqslant 1$; thus $b_n \geqslant 0$, the equality holding iff each $m_i = 1$, i.e., iff H is a tree (by Lemma A.1). Q.E.D.

Proof of Lemma 3.1: Let H be any hypergraph, $\hat{P}_1 H = \{H_j, j\in J\}$ its partition into connected components: $\bar{\sigma}(H) = \Sigma_{j\in J}\bar{\sigma}(H_j) \geqslant \Sigma_{j\in J}(|V(H_j)| - 1) = |V(H)| - |J|$. Q.E.D.

A partition $P(H) = \{H_i, i\in I\}$ naturally defines a new hypergraph

$$G\,[P(H)] \equiv \{V(H_i), i\in I\}. \tag{A4}$$

The diagram of $G\,[P(H)]$ is obtained from that of H by coalescing together the dots representing the edges of H_i, for each $i\in I$ (Fig. 7).

Clearly, if H is connected, and

$$\hat{P}_2 H = \{S_1, S_2, ..., S_m\}, \tag{A5}$$

where S_i, $i = 1,...,m$, are the maximal stars of H, then $G\,[\hat{P}_2 H] = \{V(S_i), i = 1,...,m\}$ is a tree. More generally, we have

Lemma A.3: Let H be connected, $P(H) = \{H_i, i = 1,..., p\}$ a partition of H; $G\,[P(H)] = \{V(H_i), i = 1,..., p\}$ is a tree iff (i) each H_i is connected and (ii) $P(H)$ is a superpartition of $\hat{P}_2 H$.

Proof: (if) is obvious. (only if): first note that when two nonadjacent dots in a connected hypergraph are coalesced together, there automatically results a cycle (Fig. 8); it is then easy to see that $G\,[P(H)]$ is not a tree if the H_i are not all connected. Suppose then that the H_i are all connected, but $P(H)$ is not a superpartition of $\hat{P}_2 H = \{S_j, j\in J\}$. This means that at least one of the S_j, say S_1, intersects several of the H_i. Denote $J_1 = \{j | j\in J, H_j \cap S_1 \neq \emptyset\}$. Each H_{j1}, $j_1 \in J_1$, has at least two common vertices with the union of the other H_{j1}'s (otherwise S_1 would contain an articulation vertex, or not be connected); this implies that when the $V(H_j)$'s are properly ordered, the right-most $V(H_{j1})$ shares at least two vertices with the union of $V(H_j)$'s on its left, implying that $G\,[P(H)]$ is not a tree in view of Lemma A.1. Q.E.D.

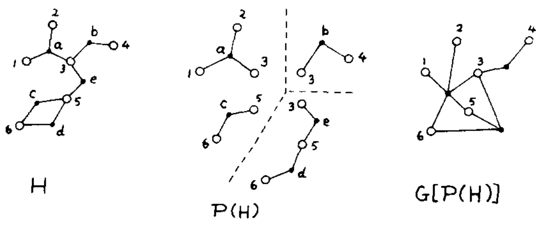


FIG. 7. The hypergraph $G[P(H)] = \{(1,2,3,5,6), (3,5,6), (3,4)\}$, where $H = \{E_a, E_b, E_c, E_d, E_e\} = \{(1,2,3), (3,4), (5,6), (5,6), (3,5)\}$ and the partition $P(H) = \{(E_a, E_c), (E_b), (E_d, E_e)\}$.

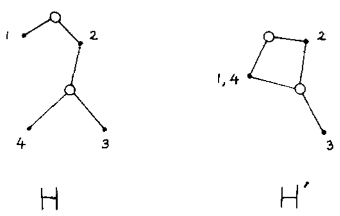


FIG. 8. If one coalesces together the two nonadjacent edges 1 and 4 in the hypergraph H, one obtains H', which contains a cycle.

Given the partition $P(H) = \{H_i, i = 1,..., p\}$, denote

$$\sigma[P(H)] \equiv \bar{\sigma}(G\,[P(H)]) = \sum_{i=1}^{p} (|V(H_i)| - 1). \tag{A6}$$

Lemma A.4: Let H be connected, $P(H) = \{H_i, i = 1,..., p\}$ a partition of H. We have

$$\sigma[P(H)] \geqslant |V(H)| - 1, \tag{A7}$$

the equality holding iff (i) each H_i, $i = 1,..., p$, is connected and (ii) $P(H)$ is a superpartition of $\hat{P}_2 H$.

Proof: This follows from Lemmas A.3 and A.2.

We extend the domain of application of partition operators $\hat{P}$ from hypergraphs to partitions as follows: if $P(H) = \{H_i, i\in I\}$ and $\hat{P}H_i = \{H_{ij}, j\in J_i\}$, then

$$\hat{P}P(H) \equiv \bigcup_{i\in I} \{\hat{P}H_i\} \equiv \{H_{ij_i}, j_i \in J_i, i\in I\}. \tag{A8}$$

Lemma A.5: Let the partition $P'(H) = \{H'_k, k\in K\}$ be such that each H'_k is connected. Let $P(H) = \{H_i, i\in I\}$ be another partition of H. Then

$$P(H) \geqslant P'(H) \Leftrightarrow \hat{P}_1 P(H) \geqslant P'(H). \tag{A9}$$

Proof: ($\Leftarrow$) is obvious since $P(H) \geqslant \hat{P}_1 P(H)$ and $\geqslant$ is transitive. ($\Rightarrow$): Let $\hat{P}_1 H_i = \{H_{ij}, j\in J_i\}$. $P(H) \geqslant P'(H) \Leftrightarrow$ each H'_k is contained in an H_i; let $H'_k \subset H_1$ say. Since H'_k is connected, it must be entirely contained in a single connected component of H_1, i.e., $H'_k \subset H_{1j}$ for some $j\in J_1$. Thus $P'(H) \leqslant \hat{P}_1 P(H) = \{H_{ij_i}, j_i \in J_i, i\in I\}$. Q.E.D.

Lemma A.6: Let H be any hypergraph, $P(H) = \{H_i, i\in I\}$ any partition of H:

$$\sigma[\hat{P}_1 P(H)] = \sum_{H_i \in P(H)} \sigma[\hat{P}_1 H_i]. \tag{A10}$$

Proof: Let $\hat{P}_1 H_i = \{H_{ij}, j\in J_i\}$. Then $\hat{P}_1 P(H) = \{H_{ij_i}, j_i \in J_i, i\in I\}$ and

$$\sigma[\hat{P}_1 P(H)] = \sum_{i\in I} \sum_{j_i \in J_i} (|V(H_{ij_i})| - 1) = \sum_{i\in I} \sigma[\hat{P}_1 H_i].$$

Q.E.D.

Let us now denote, for any partition $P(H) = \{H_i, i\in I\}$ of any hypergraph H,

$$\sigma'[P(H)] \equiv \sigma[\widehat{P}_1 P(H)] = \sum_{H_i\in P(H)} (|V(H_i)| - |\widehat{P}_1 H_i|), \quad \text{(A11)}$$

where $|\widehat{P}_1 H_i|$ is the number of connected components of H_i. The second equality in (A11) follows from (A10) and

$$\sigma[\widehat{P}_1 H] = \sum_{H_j\in \widehat{P}_1 H} (|V(H_j)| - 1) = |V(H)| - |\widehat{P}_1 H| \quad \text{(A12)}$$

for any H.

Lemma A.7: Let $P(H) = \{H_i, i\in I\}$ be any partition of any hypergraph H. We have

$$\sigma'[P(H)] = \sum_{H_j\in \widehat{P}_1 H} \sigma'[P(H)\circ H_j], \quad \text{(A13)}$$

where we denote, for any subhypergraph $H'\subset H$,

$$P(H)\circ H' \equiv \{H_i\cap H' | H_i\in P(H)\} \quad \text{(A14)}$$

the partition induced by $P(H)$ on H'.

Proof: Let $\widehat{P}_1 H = \{\overline{H}_j, j\in J\}$. We have $\sigma'[P(H)] = \Sigma_{i\in I}(|V(H_i)| - |\widehat{P}_1 H_i|) = \Sigma_{i\in I}(\Sigma_{j\in J}|V(H_i\cap\overline{H}_j)| - \Sigma_{j\in J}|\widehat{P}_1(H_i\cap\overline{H}_j)|) = \Sigma_{j\in J}\Sigma_{i\in I}(|V(H_i\cap\overline{H}_j)| - |\widehat{P}_1(H_i\cap\overline{H}_j)|) = \Sigma_{j\in J}\sigma'[P(H)\circ\overline{H}_j]$, where we used

$$|\widehat{P}_1 H'| = \sum_{H_j\in \widehat{P}_1 H} |\widehat{P}_1(H_j\cap H')| \quad \text{(A15)}$$

for any subhypergraph $H'\subset H$ (see Fig. 9).

Lemma A.8: Let H be connected, $P(H)$ any partition of H. We have

$$\sigma'[P(H)] \geqslant |V(H)| - 1, \quad \text{(A16)}$$

the equality holding iff $P(H)$ is a superpartition of $\widehat{P}_2 H$.

Proof: By Lemma A.4, $\sigma'[P(H)] = \sigma[\widehat{P}_1 P(H)] \geqslant |V(H)| - 1$, the equality holding iff (i) each element of $\widehat{P}_1 P(H)$ is connected and (ii) $\widehat{P}_1 P(H) \geqslant \widehat{P}_2 H$; but (i) is always satisfied (by definition of $\widehat{P}_1$), and (ii) $\Leftrightarrow P(H) \geqslant \widehat{P}_2 H$ by Lemma A.5.

Proof of Lemma 3.2: Let $P(H)$ be any partition of any hypergraph H: we have

$$\sigma'[P(H)] \overset{\text{(A13)}}{=} \sum_{H_j\in \widehat{P}_1 H} \sigma'[P(H)\circ H_j] \overset{\text{(A16)}}{\geqslant} \sum_{H_j\in \widehat{P}_1 H} (|V(H_j)| - 1)$$
$$= |V(H)| - |\widehat{P}_1 H|,$$

the equality holding iff $P(H_j) \geqslant \widehat{P}_2 H_j$ for each $H_j\in\widehat{P}_1 H$, i.e., iff $P(H) \geqslant \widehat{P}_2 H$. Q.E.D.

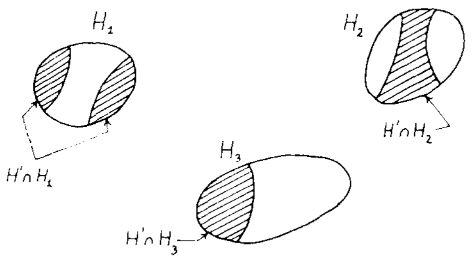


FIG. 9. The blobs represent the connected components of some hypergraph H, $\widehat{P}_1 H = \{H_1, H_2, H_3\}$. The hatched sections represent a subhypergraph $H'\subset H$. Clearly, the number of connected components of H' is $|\widehat{P}_1 H'| = \Sigma_{j=1}^{3} |\widehat{P}_1(H_j\cap H')|$ [where $|\widehat{P}_1(H_j\cap H')|$ is the number of connected components of $H_j\cap H'$].

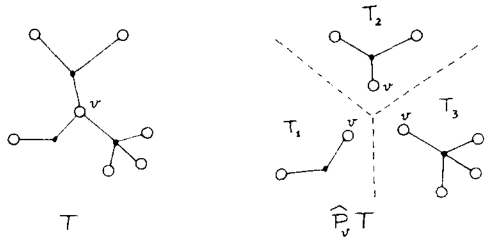


FIG. 10. Tree T wherein each $T_i\in\widehat{P}_v T$ consists of a single edge.

APPENDIX B: DEMONSTRATION OF EQUATIONS (6.23), (6.24)

Lemma B.1: Let $\{f_E, E\subset S\}$ be stochastic variables indexed by subsets of some set S; $\langle\ \rangle$ some averaging operation. Denote $\langle(\cdots)\rangle \equiv \langle L(\cdots)\rangle$ where the leveling operator L suppresses products $\Pi_{i\in I} f_{E_i}$ wherein the sets E_i, $i\in I$, are not all disjoint from one another. Let T be a tree. Denoting $T^f \equiv \Pi_{E\in T} f_E$, we have

$$\langle T^f\rangle_{c\{f\}} = \alpha(T) \prod_{E\in T} \langle f_E\rangle, \quad \text{(B1)}$$

where

$$\alpha(T) = (-)^{|T|-1} \prod_{v\in V(T)} (|v| - 1)! \quad \text{(B2)}$$

Proof: Let $v\in V(T)$ be a vertex of degree $|v| = n \geqslant 2$, and let

$$\widehat{P}_v T = \{T_1, T_2, \ldots, T_n\}, \quad \text{(B3)}$$

where the partition operator $\widehat{P}_v$ is defined in Sec. 3.1. In the special case that each $T_i\in\widehat{P}_v T$ consists of a single edge E_i (Fig. 10), we have

$$\langle T^f\rangle_{c\{f\}} = (-)^{n-1}(n-1)! \prod_{i=1}^{n} \langle f_{E_i}\rangle \quad \text{(B4)}$$

(all other terms in the cumulant vanish because of L inside $\langle\ \rangle$). We will show that in general

$$\langle T^f\rangle_{c\{f\}} = (-)^{n-1}(n-1)! \prod_{i=1}^{n} \langle T_i^f\rangle_{c\{f\}}. \quad \text{(B5)}$$

(B1) will then follow immediately by application of (B5) to each vertex of T in succession, and use of $\Sigma_{v\in V(T)}(|v| - 1) = |T| - 1$, by Lemma 3.1.

Note that (B5) can be written as

$$\langle T^f\rangle_{c\{f\}} = \Big\langle \prod_{i\in n} T_i^f \Big\rangle^{P}_{c\{T\}}, \quad \text{(B5')}$$

where

$$\langle(\cdots)\rangle^{P} \equiv \langle L_v(\cdots)\rangle_{c\{f\}} \equiv \langle L_v(\cdots)\rangle\!\!\!/\!\!/, \quad \text{(B6)}$$

where L_v suppresses products wherein the vertex v appears more than once, and

$$\langle(\cdots)\rangle\!\!\!/\!\!/ \equiv \langle(\cdots)\rangle_{c\{f\}}. \quad \text{(B7)}$$

We now prove (B5) by induction, i.e., we show that if (B5) is true for all proper subtrees of T (i.e., all $T'\subset T$ such that $T' \neq T$), then it is also true for T; this will prove our assertion since (B5) is true in the case that each $T_i \in \widehat{P}_v T$ consists of a single edge [Eq. (B4)].

We have from (2.9) and (2.3a),

$$0 = \langle T^f \rangle = \langle T^f \rangle_{c\{f\}a\{f\}} = \langle T^f \rangle_{a\{f\}} \tag{B8}$$

$$= \sum_{P \in \pi_c(T)} \prod_{\Theta \in P} \langle \Theta^f \rangle \tag{B8$'$}$$

$$= \sum_{P_1 \in \pi_c(T_1)} \sum_{P_2 \in \pi_c(T_2)} \cdots \sum_{P_n \in \pi_c(T_n)} \left(\prod_{i \in \underline{n}} \prod_{T_{ij} \in P_i}{}' \langle T_{ij}^f \rangle \right) \left\langle \prod_{i \in \underline{n}} T_{i1}^f \right\rangle_{a\{T\}}, \tag{B9}$$

where $\pi_c(T)$ denotes the set of all partitions of T into *connected* subtrees [all other partitions contribute nothing to $\langle T^f \rangle_{a\{f\}}$ since $\langle \Theta^f \rangle = \langle \Theta^f \rangle_{c\{f\}} = 0$ if Θ is not connected, by Lemma 2.1]; the first equality in (B8) follows from $LT^f = 0$ (since $n \geqslant 2$, v is repeated in T^f). In (B9), T_{i1} is the part of $P_i \in \pi_c(T_i)$ containing v, and the primed product $\Pi'_{T_{ij} \in P_i}$ is over the other parts (excluding T_{i1}). [The identity of (B9) with (B8′) is easily shown by induction, by using the obvious identity

$$\sum_{P \in \pi_c(T + \{E\})} \prod_{\Theta \in P} \langle \Theta^f \rangle = \sum_{P \in \pi_c(T)} \left[\prod_{\Theta \in P} \langle \Theta^f \rangle \langle f_E \rangle + \sum_{\Theta \in P}^{(c)} \prod_{\Theta' \neq \Theta} \langle \Theta'^f \rangle \langle \Theta^f f_E \rangle \right], \tag{B10}$$

where the superscript (c) on the last sum indicates restriction to the $\Theta \in P$ such that $\Theta + \{E\}$ is connected. Let E be linked to T_1 say, and not contain v. Assuming then that (B8′) = (B9), and applying (B10) while noticing that $\langle T_{ij}^f f_E \rangle = 0$ if $i \neq 1$, we obtain again (B8) = (B9) with T replaced by $T + \{E\}$ in (B8′) and T_1 replaced by $T_1 + \{E\}$ in (B9). And since (B8′) = (B9) is obviously true if each T_i consists of a single edge, it is true in general.]

Consider now $\langle \Pi_{i \in \underline{n}} T_{i1}^f \rangle_{a\{T\}}$ in the case that at least one of the T_{i1} is not equal to T_i; then $\Sigma_{i \in \underline{n}} T_{i1}$ is a proper subtree of T, so that we can apply our induction hypothesis (B5′), $\langle \Pi_{i \in \underline{n}} T_{i1}^f \rangle = \langle \Pi_{i \in \underline{n}} T_{i1}^f \rangle_{c\{T\}}$, whence

$$\left\langle \prod_{i \in \underline{n}} T_{i1}^f \right\rangle_{a\{T\}} = \left\langle \prod_{i \in \underline{n}} T_{i1}^f \right\rangle_{c\{T\}a\{T\}} = \left\langle \prod_{i \in \underline{n}} T_{i1}^f \right\rangle = 0, \tag{B11}$$

where we used (2.9) and noted that $L_v \Pi_{i \in \underline{n}} T_{i1} = 0$ (since $n \geqslant 2$). There thus remains in (B9) only the terms wherein each $T_{i1} = T_i$, i.e., we have

$$0 = \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle_{a\{T\}} = \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle + \sum_{P \in \pi(\underline{n})}{}' \prod_{S \in P} \left\langle \prod_{i \in S} T_i^f \right\rangle, \tag{B12}$$

where the primed sum Σ' excludes the trivial partition; we can accordingly apply the induction hypothesis (B5′) to the factors $\langle \Pi\, T_i^f \rangle$ in that sum, and thereby get

$$0 = \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle + \sum_{P \in \pi(\underline{n})}{}' \prod_{S \in P} \left\langle \prod_{i \in S} T_i^f \right\rangle_{c\{T\}}. \tag{B13}$$

Comparing this with

$$0 = \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle = \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle_{c\{T\}a\{T\}}$$

$$= \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle_{c\{T\}} + \sum_{P \in \pi(\underline{n})}{}' \prod_{S \in P} \left\langle \prod_{i \in S} T_i^f \right\rangle_{c\{T\}}, \tag{B14}$$

we deduce

$$\langle T^f \rangle \equiv \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle = \left\langle \prod_{i \in \underline{n}} T_i^f \right\rangle_{c\{T\}}. \qquad \text{Q.E.D.} \tag{B15}$$

---

[1] For reviews see, e.g., (a) K. Huang, *Statistical Mechanics* (Wiley, New York, 1963); (b) C. A. Croxton, *Liquid State Physics, A Statistical Mechanical Introduction* (Cambridge U.P., Cambridge, 1974), Chap. 1; (c) G. E. Uhlenbeck and G. W. Ford, *The Theory of Linear Graphs with Applications to the Theory of the Virial Development of the Properties of Gases*, in *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1962), Vol. I; (d) G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics* (American Mathematical Society, Providence, RI, 1963).

[2] H. D. Ursell, Proc. Cambridge Phil. Soc. **23**, 685 (1927).

[3] B. Kahn and G. E. Uhlenbeck, Physica **5**, 399 (1938).

[4] K. Husimi, J. Chem. Phys. **18**, 682 (1950).

[5] S. Ono, J. Chem. Phys. **19**, 504 (1951).

[6] F. Y. Wu, J. Math. Phys. **4**, 1438 (1963); S. Sherman, J. Math. Phys. **6**, 1189 (1965); G. Stell, J. Math. Phys. **6**, 1193 (1965).

[7] J. E. Mayer, J. Chem. Phys. **5**, 67 (1937); J. E. Mayer and M. G. Mayer, *Statistical Mechanics* (Wiley, New York, 1940).

[8] J. E. Mayer, *Equilibrium Statistical Mechanics* (Pergamon, New York, 1968), p. 74.

[9] E. E. Salpeter, Ann. Phys. **5**, 183 (1958); M. S. Green, J. Math. Phys. **1**, 391 (1960); for a review, see, e.g., G. Stell, *Cluster Expansions for Classical Systems in Equilibrium*, in *The Equilibrium Theory of Classical Fluids*, edited by H. L. Frisch and J. L. Lebowitz (Benjamin, New York, 1964).

[10] For a review see, e.g., C. Bloch, *Diagram Expansions in Quantum Statistical Mechanics*, in *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1965), Vol. III.

[11] R. Brout, Phys. Rev. **115**, 824 (1959).

[12] R. Brout and P. Carruthers, *Lectures on the Many Electron Problem* (Gordon and Breach, New York, 1969).

[13] R. Balescu, *Equilibrium and Non-Equilibrium Statistical Mechanics* (Wiley, New York, 1975).

[14] R. Kubo, J. Phys. Soc. Jpn. **17**, 1100 (1962).

[15] For more mathematical discussions of cumulants, see, e.g., H. Cramer, *Random Variables and Probability Distributions*, 3rd ed. (Cambridge U.P., Cambridge, 1970); R. N. Bhattacharya and R. Ranga Rao, *Normal Approximation and Asymptotic Expansions* (Wiley, New York, 1976); D. R. Brillinger, *Time Series, Data Analysis and Theory*, expanded edition (Holden-Day, San Francisco, 1981).

[16] Section 2.5 of Ref. 12, and references therein.

[17] T. D. Lee and C. N. Yang, Phys. Rev. **113**, 1165 (1959); **116**, 25 (1959).

[18] (a) C. Berge, *Graphes et Hypergraphes* (Dunod, Paris, 1970) [English translation: *Graphs and Hypergraphs* (North-Holland, Amsterdam, 1973); (b) J. E. Graver and M. E. Watkins, *Combinatorics with Emphasis on the Theory of Graphs* (Springer, Berlin, 1977).

[19] E.g., G. Stell, *Generating Functionals and Graphs*, in *Graph Theory and Theoretical Physics*, edited by F. Harary (Academic, New York, 1967); or, *Correlation Functions and their Generating Functionals*, in *Phase Transitions and Critical Phenomena*, edited by C. Domb and M. S. Green (Academic, New York, 1976), Vol. 5B.

[20] A multiset is also sometimes called a *selection*; see, e.g., Ref. 18(b), p. 5.

[21] Equation (2.9) follows immediately from definitions (2.1):

$$\langle e^{\Sigma X} \rangle_{ca} \overset{(2.1a)}{=} \exp(\langle e^{\Sigma X} - 1 \rangle_c) \overset{(2.1c)}{=} \langle e^{\Sigma X} \rangle,$$

$$\langle e^{\Sigma X} - 1 \rangle_{ac} \overset{(2.1c)}{=} \ln \langle e^{\Sigma X} \rangle_a \overset{(2.1a)}{=} \langle e^{\Sigma X} - 1 \rangle.$$

[22] This is readily seen after replacing each $\exp(1_i)$ by $(1 + 1_i)$, as is permissible under the protection of L'.

[23] In the case of an edge E, $|E|$ is indeed the number of vertices contained in E, in our set theoretic notation. By duality, we also denote $|v|$ the number of edges containing the vertex v (note that one can represent a hypergraph

H by its dual[18] $H^* = (V_1, V_2, \ldots, V_n)$, where $n = |V(H)|$ and V_i is the set of edges containing the vertex v_i).

[24] More generally, given a set of vertices $W \subset V(H)$, we write $E \overset{W}{--} E'$ if there exists a path between edges E and E' not containing any vertex of W; $\overset{W}{--}$ is clearly an equivalence relation between edges.

[25] In detail,

$$\langle H^f \rangle_{c\{f,\text{irr}\}} = \sum_{P \in \pi(J)} (-)^{|P|-1}(|P|-1)! \prod_{J_i \in P} \left\langle \prod_{j \in J_i} S_j^f \right\rangle,$$

where $S_j^f = \Pi_{E \in S_j} f_E$.

[26] See, e.g., Ref. 1(a), Sec. 7.6 (Gibbs paradox).

[27] See, e.g., Ref. 10, p. 31.

[28] E.g., if $E = \underline{k} = \{1,2,\ldots,k\}$, we have $(k-1)!$ different cycles on $\underline{k}$ corresponding to the $(k-1)!$ distinct permutations of $k-1$ indices in the basic cycle $(1,2,\ldots,k)$, keeping one index fixed [since $(i_1,i_2,\ldots,i_k)$ and a cyclic permutation of it, e.g., $(i_k,i_1,\ldots,i_{k-1})$ represent the same cycle].

[29] However, there are certain cases where different irreducible blocks *are* $\langle\ \rangle$-independent. E.g., let H contain an articulation vertex v which, for some partition $H = H_1 + H_2$ with $V(H_1) \cap V(H_2) = v$, is incident on only C-edges of H_1 and only f-edges of H_2. Now, C- and f- edges, although not commuting, are not entangled by Λ in (6.15), which puts all C's to the right of all f's; and since furthermore all f-edges of H_1 commute with all f-edges of H_2 as they have no common vertex, and likewise for C-edges, H_1 and H_2 are completely de-entangled. It follows that $\langle H_1^{f,C} H_2^{f,C} \rangle = \langle H_1^{f,C} \rangle \langle H_2^{f,C} \rangle$ and $\langle H^{f,C} \rangle_{c\{\text{irr}\}} = 0$.

[30] To adapt (B1) to the evaluation of $\langle T^C \rangle_{c\{C\}}$, let $\langle (\cdots) \rangle = (\text{tr}\, e^{-\beta K})^{-1} \text{tr}\{e^{-\beta K}(\cdots)\}$, so that (6.21) $= \langle L(\cdots) \rangle = \langle (\cdots) \rangle'$.

[31] Note that here $\langle (\cdots) \rangle = (\cdots)$, so that

$$\langle T^D \rangle'_{c\{D\}} = \alpha(T) \prod_{E \in T} \langle D_E \rangle = \alpha(T) \prod_{E \in T} D_E = \alpha(T) T^D.$$

On a relation between percolation and phase transition in the gauge invariant Ising model

Rossana Marra
Istituto di Fisica, Università di Salerno, 84100-Salerno, Italy

(Received 12 January 1982; accepted for publication 6 August 1982)

We point out a connection between the coexistence of phases in the Z_2 lattice gauge model and the existence of infinite clusters for a suitable associated system of duplicate currents.

PACS numbers: 05.50. + q, 11.15. − q

1. INTRODUCTION

Computer simulations by Monte Carlo methods[1] strongly suggest the existence of a first-order phase transition in the gauge invariant Ising model (GIIM) in four dimensions. In Ref. 2 it is pointed out that this transition could be interpreted as coexistence of phases at the singular point arising from different boundary conditions (closed and free b.c.). More precisely the first-order phase transition appears to be associated to a spontaneous breaking of the symmetry given by duality, as well as in the Ising model the coexistence of two phases for $\beta \geqslant \beta_c$, in zero external field, is related to a spontaneous breaking of the reflection symmetry for the spin variables.

A useful picture of Ising phase transition was obtained in terms of percolative phenomena, e.g., the existence of infinite clusters of spins "up" and "down" in typical configurations of pure phases.[3–6] More recently Aizenman[7] related the long range order in Ising model to some percolation phenomenon in an associated "current system." The Aizenman procedure allows one to write the correlation functions for the Ising model in terms of some conditional probabilities on a suitable current systems. This method is applicable to other systems and in particular to GIIM, as stated in Ref. 7.

In this paper we start from the Aizenman results to investigate the relation between the GIIM phase coexistence and the existence of infinite clusters in the current model. In particular in Sec. 2 we describe, for sake of completeness, the Aizenman procedure for GIIM. In Sec. 3 the absence of first-order phase transition in the GIIM when there are no infinite clusters of currents is shown.

2. THE CURRENT EXPANSION FOR THE GAUGE INVARIANT ISING MODEL

Let $\mathbb{L}^d$ be the set of links $l = (n,n')$ between nearest neighbor sites in the unit lattice $\mathbb{Z}^d$. We associate to each link l a spin $\sigma_l = \pm 1$ variable and denote σ^L a spin configuration on $\mathbb{L}^d$, i.e., $\sigma^L = \{\sigma_l\}_{l \in \mathbb{L}^d}$. We call plaquette each set of four links which are edges of unit squares and for each plaquette p we write $\sigma_p = \Pi_{l \subset p} \sigma_l$. Finally Let $H_\Lambda(\sigma^L)$ be the energy of the system in $\Lambda \subset \mathbb{L}^d$ for the configuration σ^L. The GIIM is defined by assigning the Hamiltonian

$$H_\Lambda(\sigma^L) = - \sum_{p \subset (\Lambda \cup B\Lambda)} \beta_p \sigma_p, \qquad (2.1)$$

where $B\Lambda$ is the set of the links in $\mathbb{L}^d \setminus \Lambda$ which form plaquettes with the links in Λ.

We call Z^c_Λ, Z^0_Λ the partition functions with, respectively, the $(\pm)$-boundary conditions (closed)[8] and the free boundary conditions (i.e., $\beta_p = 0 \ \forall p \not\subset \Lambda$); we define

$$Z^\epsilon_\Lambda = \frac{1}{2^{|\Lambda|}} \sum_{\sigma^{\Lambda_\epsilon}} \prod_{p \subset \Lambda_\epsilon} \exp \beta_p \sigma_p, \qquad (2.2)$$

where $\epsilon = c, 0$; $\Lambda_0 = \Lambda$, $\Lambda_c = \Lambda \cup B\Lambda$.

Following Ref. 7, for each plaquette we use

$$\exp \beta_p \sigma_p = \sum_{n \in N} \frac{(\beta_p \sigma_p)^n}{n!},$$

and take the product on p and the average on spin configurations. Then we get

$$Z^\epsilon_\Lambda = \sum_{n^{\Delta_\epsilon} : (\partial n_1)^\Lambda = \varnothing} W(n^{\Delta_\epsilon}), \qquad (2.3)$$

where Δ_0 is the set of plaquettes contained in Λ and Δ_c is the set of plaquettes contained in $\Lambda \cup B\Lambda$. n^Δ is a current configurations, that is $n^\Delta = \{n(p)\}_{p \in \Delta}$ with $n(p)$ an integer for each p. For a configuration n^Δ, $(\partial n)^{\Lambda'}$ is the set

$$(\partial n)^{\Lambda'} = \left\{ l \supset \Lambda' : \sum_{\substack{p \supset l \\ p \in \Delta'}} n(p) = 1 \pmod 2 \right\}. \qquad (2.4)$$

Finally

$$W(n^\Delta) = \prod_{p \subset \Lambda} \frac{(\beta_p)^{n(p)}}{n(p)!}. \qquad (2.5)$$

By small modifications of the above argument we have, for the expectations of the gauge invariant observables, $\sigma_V = \Pi_{l \subset V} \sigma_l$ (V union of circuits in $\mathbb{L}^d$),

$$\langle \sigma_V \rangle^\epsilon_\Lambda = \frac{\displaystyle\sum_{n^{\Delta_\epsilon} : (\partial n)^\Lambda = V} W(n^{\Delta_\epsilon})}{\displaystyle\sum_{n^{\Delta_\epsilon} : (\partial n)^\Lambda = \varnothing} W(n^{\Delta_\epsilon})} \qquad (2.6)$$

Now we need some definitions.

Two plaquettes are called *adjacent* if they have a common link. A *chain* is a finite sequence $(p_1, \ldots, p_n)$ of distinct plaquettes in $\mathbb{L}^d$ such that p_i and p_{i+1} are adjacent for each $i \in \{1, \ldots, n\}$. A subset A of a plaquettes is *connected* if $\forall (p_i, p_j), p_i, p_j \in A$, there is a chain of elements in A, having p_i, p_j as terminal plaquettes.

Given $V \subset \mathbb{L}^d$ we call *surface* for V a connected subset of plaquettes s.t. $\Delta A = V$, where

$$\Delta A = \left\{ l \subset \mathbb{L}^d : \sum_{p \supset l} \chi_A(p) = 1 \pmod 2 \right\}$$

0022-2488/83/040913-04$02.50

with χ_A the characteristic function of A.

Given a current configuration on $\mathbb{L}^d$, n, we denote by $J(n\neq 0)$ the set of plaquettes for which $n(p)\neq 0$; a *cluster in n* is a maximal connected component of $J(n\neq 0)$; V belongs to a cluster in n if there is a cluster which contains a surface for V.

Now we claim that the following relations hold:

$$((\langle\sigma_\Gamma\rangle_\Lambda^0)^2 = \frac{\displaystyle\sum\nolimits'_{\substack{n_1{}^{\Delta_0}:(\partial n_1)^\Lambda=\varnothing\\ n_2{}^{\Delta_0}:(\partial n_2)^\Lambda=\varnothing}} W(n_1{}^{\Delta_0})W(n_2{}^{\Delta_0})}{\displaystyle\sum_{\substack{n_1{}^{\Delta_0}:(\partial n_1)^\Lambda=\varnothing\\ n_2{}^{\Delta_0}:(\partial n_2)^\Lambda=\varnothing}} W(n_1{}^{\Delta_0})W(n_2{}^{\Delta_0})}, \tag{2.7}$$

where Σ' means the sum on the configurations $n_1{}^\Delta$, $n_2{}^\Delta$ s.t. Γ belongs to a cluster in $m^\Delta = n_1{}^\Delta + n_2{}^\Delta$ (Γ contour in $\mathbb{L}^d$).

On the other hand, for closed boundary conditions,

$$((\langle\sigma_\Gamma\rangle_\Lambda^c)^2 = \frac{\displaystyle\sum_{\substack{n_1{}^{\Delta_0}:(\partial n_1)^\Lambda=\varnothing\\ n_2{}^{\Delta_0}:(\partial n_2)^\Lambda=\varnothing}}{}^{*}\, W(n_1{}^\Delta)W(n_2)^\Delta}{\displaystyle\sum_{\substack{n_1{}^{\Delta_0}:(\partial n_1)^\Lambda=\varnothing\\ n_2{}^{\Delta_0}:(\partial n_2)^\Lambda=\varnothing}} W(n_1{}^\Delta)W(n_2{}^\Delta)}. \tag{2.7'}$$

The above sum Σ^* is taken over all the configurations $(n_1{}^\Delta, n_2{}^\Delta)$ s.t. either Γ belongs to a cluster in m^Δ or there is $\Gamma^*\subset B\Lambda$ s.t. $\Gamma\cup\Gamma^*$ belongs to a cluster in m^Δ.

Proof:

$$((\langle\sigma_\Gamma\rangle_\Lambda^0)^2$$

$$= (Z_\Lambda^0)^{-2} \sum_{n_1{}^{\Delta_0}:(\partial n_1)^\Lambda=\varnothing} W(n_1{}^{\Delta_0}) \sum_{n_2{}^{\Delta_0}:(\partial n_2)^\Lambda=\varnothing} W(n_2{}^{\Delta_0})$$

$$= (Z_\Lambda^0)^{-2} \sum_{m^{\Delta_0}:(\partial m)^\Lambda=\varnothing} W(m) \sum_{n_1{}^{\Delta_0}:(\partial n_1)^\Lambda=\Gamma} \binom{m}{n},$$

where

$$m = n_1 + n_2;\quad \binom{m}{n} = \prod_{p\in\Delta_0} \binom{m(p)}{n(p)}.$$

We observe that all configurations $m = n_1 + n_2$ are s.t. Γ belongs to a cluster in m. In fact an easy inductive argument shows that there exists a path for Γ s.t. $n_1(p)\neq 0$ for all the plaquettes of the path. Now we use the following lemma:

Lemma: If Ξ belongs to a cluster in m then

$$\sum_{n:\partial n=\Xi} \binom{m}{n} = \sum_{n:\partial n=\varnothing} \binom{m}{n}, \tag{2.8}$$

where Ξ is a finite collection of contours in $\mathbb{L}^d$. In Ref. 9 is proven essentially the same lemma for the case of an Ising model by using graphical methods; we give an alternative proof of the lemma in Appendix A, which works also for the GIIM. The lemma and the above remark conclude the proof for free boundary conditions; small modifications of the argument give the result for closed boundary conditions.

3. MAIN RESULT

We associate to each plaquette p on $\mathbb{L}^d$ two natural numbers $n_1(p)$, $n_2(p)$. The configuration space is $\Omega = \times_{p\in S^d}(N\times N)$ where S^d is the set of plaquettes on $\mathbb{L}^d$. Let $\mathfrak{B}$ be the Borel σ-algebra on Ω. We introduce in $\mathfrak{B}$ the probability measure ν ("free measure") as the infinite volume limit of the measures

$$\nu_\Delta(A) = \frac{\sum_{\omega\in A} W(\omega)}{\sum_{\omega\in\Omega_\Delta} W(\omega)} \quad \forall A\in\mathfrak{B}_\Delta, \tag{3.1}$$

where

$$\omega\equiv(n_1,n_2),\quad W(\omega) = W(n_1)W(n_2).$$

Given $\Lambda\subset\mathbb{L}^d$ we consider the event M_Λ

$$M_\Lambda = \{\omega\in\Omega:(\partial n_1)^\Lambda = (\partial n_2)^\Lambda = \varnothing\}$$

and the conditional probability

$$\nu(A\,|M_\Lambda)\equiv\mu_\Lambda(A),$$

and we denote by $\mu(A)$

$$\mu(A) = \nu(A\,|M) = \lim_{\Lambda\nwarrow\mathbb{L}_d} \nu(A\,|M_\Lambda), \tag{3.2}$$

where $M = \cap_\Lambda M_\Lambda$.

We introduce also a different conditional probability:

$$\mu_\Lambda^0(A)\equiv\nu^0(A\,|M_\Lambda) = \nu(A\,|M_\Lambda\cap H),$$

where H is the "boundary condition"

$$H = \{\omega\in\Omega:n_1(p) + n_2(p) = 0\ \ \forall p\in S^d\setminus\Delta_0\}$$

with Δ_0 defined in the previous section.

We define $\mu^0(A)$ by

$$\mu^0(A) = \nu^0(A\,|M) = \lim_{\Lambda\nwarrow\mathbb{L}^d} \nu^0(A\,|M_\Lambda). \tag{3.3}$$

Now consider the event

$$A_\Gamma = \{\omega\in\Omega:\Gamma \text{ belongs to a cluster in } n_1 + n_2\};$$

then relations (2.7) and (2.7′) imply

$$((\langle\sigma_\Gamma\rangle^c)^2 = \mu(A_\Gamma), \tag{3.4}$$

$$((\langle\sigma_\Gamma\rangle^0)^2 = \mu^0(A_\Gamma). \tag{3.5}$$

Our goal is to evaluate the difference

$$((\langle\sigma_p\rangle^c)^2 - ((\langle\sigma_p\rangle^0)^2$$

in terms of some "percolation probability." To this end we consider the family $\mathbb{T}_p$ of all finite connected subsets $K\subset S^d$ such that a path for p belongs to K. For any $K\subset S^d$ consider the event

$$A_p^K = \{\omega\in\Omega:K \text{ is a cluster to which belongs } p \text{ in } n_1 + n_2\}.$$

We observe that $A_p^K\cap A_p^{K'} = \varnothing$.

Finally we denote by A_p^∞ the event

$$A_p^\infty = \{\omega\in\Omega:p \text{ belongs to an infinite cluster in } n_1 + n_2\}.$$

The above definitions imply

$$\mu(A_p) = \sum_{K\in\mathbb{T}_p} \mu(A_p^K) + \mu(A_p^\infty), \tag{3.6}$$

$$\mu^0(A_p) = \sum_{K\in\mathbb{T}_p} \mu^0(A_p^K) + \mu^0(A_p^\infty). \tag{3.7}$$

Now we state the following

Theorem: For the gauge invariant Ising model the following inequality holds:

$$((\langle\sigma_p\rangle^c)^2-((\langle\sigma_p\rangle^0)^2\leqslant\mu(A_p^\infty). \tag{3.8}$$

Proof: Call BK $\forall K\in\mathbb{T}_p$, the set of plaquettes $p\in S^d\setminus K$ which are adjacent to plaquettes of K; we consider, for any fixed $K\in\mathbb{T}_p$, a d-dimensional cube Δ in such a way that $\Delta_0\supset K\cup BK$. The theorem is proven if for any $K\in\mathbb{T}_p$.

$$\mathfrak{H}\equiv\frac{\mu_{\Delta'}(A_p^K)}{\mu_{\Delta'}^0(A_p^K)}\leqslant 1\quad\forall\Delta'\supset\Delta. \tag{3.9}$$

We rewrite $\mathfrak{H}$ as

$$\mathfrak{H}=\mu_\Delta(H)\frac{\mu_\Delta(A_p^K)}{\mu_\Delta(A_p^K\cap H)}. \tag{3.10}$$

Now we consider

$$\mu_\Delta(H)=\frac{\sum\limits_{\omega\in(\Omega_\Delta\cap H)}W(\omega)}{\sum\limits_{\omega\in\Omega_\Delta}W(\omega)}.$$

If one uses the expression for $Z_{\Delta_0}^c$ and $Z_{\Delta_0}^0$, it is easy to see that

$$\mu_\Delta(H)=\frac{(Z_{\Delta_0}^0)^2}{(Z_{\Delta_0}^c)^2}. \tag{3.11}$$

Consider now

$$\frac{\mu_\Delta(A_p^K)}{\mu_\Delta(A_p^K\cap H)}=\frac{\sum\limits_{\omega\in(A_p^K\cap\Omega_\Delta)}W(\omega)}{\sum\limits_{\omega\in(A_p^K\cap\Omega_\Delta\cap H)}W(\omega)}. \tag{3.12}$$

Since all configurations $\omega\in A_p^K\cap\Omega_\Delta$, by maximality of the clusters, must assign $n_1(p)+n_2(p)=0$ $\forall p\in BK$ we have

$$\frac{\mu_\Delta(A_p^K)}{\mu_\Delta^0(A_p^K)}=\frac{\sum\limits_{\omega\in(\Omega_{\Delta\setminus K}\cap H^{K^*})}W(\omega)}{\sum\limits_{\omega\in(\Omega_{\Delta\setminus K}\cap H\cap H^{K^*})}W(\omega)}, \tag{3.13}$$

where $K^*=K\cup BK$,

$H^{K^*}=\{\omega\in\Omega_{\Delta\setminus K}:n_1(p)+n_2(p)=0,\ \forall p\in BK\}$.

Now we denote by $Z_{\Delta'}^{c,0}(Z_{\Delta'}^{0,0})$ the partition functions in $\Delta'=\Delta_0\setminus K^*$ with closed (free) boundary conditions on $\Delta\setminus\Delta_0$ and free b.c. on BK. Finally we have

$$\frac{\mu_\Delta(A_p^K)}{\mu_\Delta(A^K\cap H)}=\frac{(Z_{\Delta'}^{c,0})^2}{(Z_{\Delta'}^{0,0})^2}. \tag{3.14}$$

However, for $\mathfrak{H}$

$$\begin{aligned}\mathfrak{H}&=\frac{(Z_{\Delta_0}^0)^2(Z_{\Delta'}^{c,0})^2}{(Z_{\Delta_0}^c)^2(Z_{\Delta'}^{0,0})^2}\\&=\left[\frac{\left\langle\exp\beta\sum\limits_{l\in G}\sigma_l\right\rangle_{\Delta'}^{0,0}}{\left\langle\exp\beta\sum\limits_{l\in G}\sigma_l\right\rangle_{\Delta_0}^{0}}\right]^2,\end{aligned} \tag{3.15}$$

where G is the set of links which belongs to plaquettes of Δ_0 and plaquettes of $\Delta\setminus\Delta_0$. Since $\Delta'\subset\Delta_0$ for G.K.S. (Ref. 10) inequalities, we have

$$\mathfrak{H}\leqslant 1.$$

A by-product of the result of this section is the following

Proposition: There is no first-order phase transitions for the GIIM if $\mu(A_p^\infty)=0$.

4. CONCLUSION

It is interesting to point out the relation between the results for GIIM and analogous ones for the Ising model. This will show also the limitations of the inequality (3.8) as a possible criterion for phase transition in the GIIM.

For the Ising model we use the same notations as in the previous sections, but in this case one should read link for plaquette and site for link. Then we have the following relations:

$$((\langle\sigma_x\rangle^c)^2=\mu(A_x^\infty), \tag{4.1}$$

$$((\langle\sigma_x\sigma_y\rangle^c)^2-((\langle\sigma_x\sigma_y\rangle^0)^2\leqslant\mu(A_{xy}^\infty), \tag{4.2}$$

where $A_x^\infty(A_{xy}^\infty)$ is the event

$A_x^\infty(A_{xy}^\infty)=\{\omega\in\Omega:x(xy)$ belongs to an infinite cluster in $\omega\}$.

Formula (4.1) is in accord with Ref. 7, in which the long range order is identified as a percolative phenomenon, and characterizes the phase transition. Formula (4.2) is the analog of (3.8) and it is interesting to analyze the content of this inequality for the two-dimensional case, where the phase space structure is well known (see Ref. 6). In this case the left side of (4.2) is zero for all temperatures T, while the right one is zero for $T\geqslant T_c$, because

$$\mu(A_{xy}^\infty)\leqslant\mu(A_x^\infty)$$

and is greater than zero for $T<T_c$, as shown by

$$\mu(A_x^\infty)=\mu\Big(\bigcup_{y:|x-y|=1}A_{xy}^\infty\Big).$$

For the GIIM it is possible to prove that $\mu(A_p^\infty)=0$ for small β (see Appendix B), showing absence of phase transition as an application of (3.8). On the other hand (3.8) holds only as inequality and it could happen, as in Ising model, that the left side is zero (as it is expected beyond the self-dual point) while the right hand stays greater than zero. Clearly the most favorable situation for the application of (3.8) would be the case where $\mu(A_p^\infty)$ would remain equal to zero up to the singular point, where the phase transition in this case would be characterized by the fact that the left and the right side become simultaneously different from zero. In order to clarify the situation it would be very useful to obtain bounds from below for the left side of (3.8).

ACKNOWLEDGMENTS

I would thank Professor R. Esposito and Professor F. Guerra for very helpful discussions. I would also thank Professor M. Aizenman for useful comments and suggestions.

APPENDIX A

We now prove the lemma claimed in Sec. 2.

For each fixed configuration n on S^d and for any plaquette p_0 let $\tilde{n}$ and n_0 be two configurations such that

$$\begin{aligned} &\tilde{n} + n_0 = n, \\ &n_0(p_0) = n(p_0), \\ &\tilde{n}(p) = n(p) \quad \forall p \neq p_0. \end{aligned} \tag{A1}$$

It follows that for each $\Lambda \subset \mathbb{L}^d$,

$$(\partial n)^\Lambda = (\partial n)^{\Lambda \setminus p_0} \cup (\partial n)^{p_0} = (\partial \tilde{n})^{\Lambda \setminus p_0} \cup (\partial n)^{p_0}. \tag{A2}$$

Hence,

$$\sum_{n:(\partial n)^\Lambda = \Xi} \binom{m}{n} = \sum_{\tilde{n}:(\partial \tilde{n})^{\Lambda \setminus p_0} = \Xi \setminus p_0} \binom{m}{\tilde{n}} \sum_{n_0:(\partial n)^{p_0} = \Xi \cap p_0} \binom{m}{n_0}. \tag{A3}$$

The obvious identity

$$\sum_{K>0} \binom{m}{2K+1} = \sum_{K>0} \binom{m}{2K} \quad \forall m \neq 0 \tag{A4}$$

allows us to modify (A3) suitably. If $\tilde{n}$ is such that the sum on n_0 is a sum on odd (even) integers, we substitute for it the one on even (odd) integers; the above modification maps each configuration n in a configuration n' with the property

$$\sum_{p \supset l} n'(p) = 1 \ \ (\text{mode } 2), \quad \forall l \subset p_0, \ l \not\subset \Xi,$$

$$\sum_{p \supset l} n'(p) = 1 \ \ (\text{mode } 2), \quad \forall l \subset p_0, \ l \subset \Xi.$$

Therefore, denoting $\Xi' = \Xi \Delta p_0$ (Δ symmetric difference of sets), we get

$$\sum_{n:(\partial n)^\Lambda = \Xi} \binom{m}{n} = \sum_{n':(\partial n')^\Lambda = \Xi'} \binom{m}{n'}.$$

If there exists a cluster to which belongs a path for Ξ in m we can iterate this argument for any plaquette of the path and obtain

$$\sum_{n:(\partial n)^\Lambda = \Xi} \binom{m}{n} = \sum_{n:(\partial n)^\Lambda = \varnothing} \binom{m}{n}.$$

APPENDIX B

In order to prove that $\mu(A_p^\infty) = 0$ for small β, we consider, for any circuit Γ which contains a fixed link, l_0, the event A_Γ (see Sec. 3). For each positive integer K let B_K be the event

$$B_K = \bigcup_{\Gamma : A(\Gamma) = K} A_\Gamma,$$

where $A(\Gamma)$ is the area of the minimal surface for Γ.

The sequence $\{B_K\}$ is decreasing since a cluster for a given Γ such that $A(\Gamma) = K$ is a cluster for some Γ' such that $A(\Gamma') = K' \ \forall K' < K$.

Since $A_p^\infty \subset \cap_{K>0} B_K$ for a plaquette p such that $l_0 \in p$ we have

$$\mu(A_p^\infty) \leqslant \lim_{K \to \infty} \sum_{\Gamma : A(\Gamma) = K} \mu(A_\Gamma).$$

Then, because $\langle \sigma_\Gamma \rangle^c \leqslant a(\beta)^{A(\Gamma)}$ for β smaller than a suitable $\bar{\beta}$ [see for example Ref. 11] with $\lim_{\beta \to 0} a(\beta) = 0$, we obtain the result, for β sufficiently small, from the inequality

$$\mu(A_p^\infty) \leqslant \lim_{K \to \infty} (2(d-1)-1)^K a(\beta)^K,$$

where $(2(d-1)-1)^K$ is a bound on the number of circuits Γ such that $A(\Gamma) = K$, which contains a fixed link.

---

[1] M. Creutz, L. Jacobs, and C. Rebbi, Phys. Rev. Lett. **42**, 1390 (1979).

[2] F. Guerra, "Gauge Fields on a Lattice: Selected Topics II," in *Field Theoretical Methods in Particle Physics*, edited by W. Ruhl (Plenum, New York, 1980).

[3] R. Peierls, Proc. Cambridge Phil. Soc. **32**, 477 (1936).

[4] A. Coniglio, C. R. Nappi, F. Perugi, and L. Russo, Commun. Math. Phys. **51**, 315 (1976).

[5] L. Russo, Commun. Math. Phys. **67**, 251 (1979).

[6] M. Aizenman, Commun. Math. Phys. **73**, 83 (1980).

[7] M. Aizenman, Phys. Rev. Lett. **47**, 1 (1981).

[8] We note that $Z_\Lambda^+ = Z_\Lambda^-$ as a consequence of the gauge invariance.

[9] R. Griffiths, C. Hurst, and S. Sherman, J. Math. Phys. **11**, 790 (1970).

[10] R. Griffiths, J. Math. Phys. **8**, 478, 484 (1967); Commun. Math. Phys. **6**, 121 (1967); D. G. Kelly and S. Sherman, J. Math. Phys. **9**, 466 (1968).

[11] G. F. De Angelis, D. de Falco, F. Guerra, and R. Marra, Acta Physica Austriaca, Suppl. **XIX**, 205 (1978).

Orbit space and minima of Landau–Higgs potentials

Marko V. Jarić
Institut des Hautes Études Scientifiques, 35 route de Chartres, 91440–Bures-sur-Yvette, France

(Received 27 September 1982; accepted for publication 12 November 1982)

We establish a connection between recent propositions for minimization of Higgs potentials in an orbit space and related treatments in the Landau theory of phase transitions. Starting from isotropy groups we give a simple and systematic procedure for determining a stratification in an orbit space. As an illustration we treat the point groups O_h and T_d which occur in the Higgs mechanism for SO_7 and SU_4 adjoint representations.

PACS numbers: 05.70.Fh, 02.20. + b, 11.30.Qc

In the Landau theory of phase transitions one is faced with the problem of minimizing an R-invariant, at least quartic, potential $V(\psi)$. R is usually a finite matrix group $R < \mathrm{O}(n)$ which is a representation of a crystallographic (magnetic) space group. We will assume, without a loss of generality, that this is the case in what follows. Furthermore, even when $V(\psi)$ is a Higgs potential and the symmetry group is continuous, compact, it is often the case that the group relevant for spontaneous symmetry breaking is a finite group.[1]

When the field $\psi \in \mathbb{R}^n$ has a large number of components n, an explicit minimization becomes a difficult problem. Faced with such difficulties Gufan suggested, some ten years ago,[2] that $V(\psi)$ should be considered a function of an integrity basis $\theta = \Theta(\psi)$, $\theta \in \mathbb{R}^m$, for the ring of invariant polynomials on $\mathbb{R}^n$,

$$V(\psi) = \hat{V}(\theta)\,. \tag{1}$$

Since $\Theta(\psi)$ depends only on the group and not on V he realized that, regardless of the details of V, some components of the equations for extrema,

$$\partial V(\psi) = \delta\hat{V}(\theta)\partial\theta = 0\,, \tag{2}$$

will be identically satisfied at appropriate symmetry (hyper) planes.[3] In the above equation ∂ denotes differentiation in $\mathbb{R}^n$ with respect to ψ, whereas δ denotes differentiation in $\mathbb{R}^m$ with respect to θ.

It was realized later that $\hat{V}(\theta)$ could actually be minimized in orbit space, provided that the constraints determining the domain of θ, which we denote $\Theta(\mathbb{R}^n) \leqslant \mathbb{R}^m$, are taken into account. This idea was explicitly used to prove that no minima will be found on the generic stratum whenever at least one component of $\delta\hat{V}^*(\theta)$ is identically constant (a star indicates that constraints are included via Lagrange multipliers).[4] In fact, an observation that a vector δV is zero if and only if its norm is zero transfers Eq. (2) entirely to orbit space,

$$\partial V\,\partial\tilde{V} = \delta\widehat{V}\widehat{P}(\theta)\tilde{\delta}\widehat{V} = 0, \quad \theta \in \Theta(\mathbb{R}^n)\,. \tag{3}$$

In Eq. (3) a tilde denotes transposition and, using R-invariance of the norm, we introduced $\hat{P}(\theta) = \partial\theta\,\partial\tilde{\theta}$. $\hat{P}(\theta)$ is a linear, symmetric operator (a matrix) on $\mathbb{R}^m$ [by an analytical continuation from $\Theta(\mathbb{R}^n)$].

$\hat{P}(\theta)$ is precisely the matrix introduced by Abud and Sartori.[5] Their observations follow from Eq. (3) and the fact that ∂V is tangential to a stratum.[6] We only recall that $\delta\hat{V}$ must belong to the null space of $\hat{P}(\theta)$ and connected parts of an i-dimensional stratum (i.e., a stratum with an i-dimensional invariant slice) are entirely contained in connected parts of the following analytical stratification of $\mathbb{R}^m$:

$$\operatorname{rank}\hat{P}(\theta) = i, \quad \hat{P}(\theta) \geqslant 0\,. \tag{4}$$

We would like to add that Eq. (4) may be written in a more explicit form

$$\mathrm{Sp}_\alpha\hat{P}(\theta)\begin{cases} > 0: & \alpha \leqslant i\,, \\ = 0: & \alpha > i\,, \end{cases} \tag{5}$$

where Sp_α (spur) of a matrix is the sum of its principal minors of order α. In particular, $\mathrm{Sp}_1 P = \operatorname{tr} P$ and $\mathrm{Sp}_m P = \det P$. When considering Eq. (4) on $\mathbb{R}^n$ the second condition is trivially satisfied whereas the first condition gives exactly the same stratification as does the action of R. However, if we are looking for a stratification by R in $\mathbb{R}^m$ then Eq. (4) gives proper stratification on $\Theta(\mathbb{R}^n)$ but it also gives in general some extra pieces in $\mathbb{R}^m$. We will show later in the text that such extra pieces can easily be distinguished from $\Theta(\mathbb{R}^n)$.

Although a minimization of V is equivalent to a minimization of $\hat{V}$ with the constraints Eq. (5) included, the idea is not always practical. Namely, when R admits a large integrity basis we will be replacing a system of nonlinear (cubic) equations in ψ by a system of nonlinear equations of much higher degree in θ.[7] Even when a solution is found in $\mathbb{R}^m$ it remains unanswered in the above method what are the corresponding solutions in $\mathbb{R}^n$ and what are their isotropy groups.

It was also emphasized in Ref. 2 that the broken symmetry is determined by cosines of angles of ψ, $\hat{\psi} \equiv \psi/\|\psi\|$, independently of $\|\psi\|$ which is $\mathrm{O}(n)$ invariant. A similar idea was pursued by Kim[8] who worked on the Higgs problem, but whose results are equally applicable to the Landau theory. The greatest merit of this approach is its geometrical appeal. Nevertheless, when R admits more than four independent cubic and quartic invariants a reliance on geometrical intuition becomes questionable. Furthermore, this method does not offer a general procedure for nonparametric determination of the domain of "orbit parameters" which it introduces. In this respect there is, however, a trivial connection between Refs. 5 and 8. Orbit parameters are actually components of the integrity basis, evaluated at the unit

---

[a] Current address: Physics Department, Montana State University, Bozeman, Montana 59717.

0022-2488/83/040917-05$02.50

sphere $\theta_0 \equiv \|\psi\|^2 = 1$. Their domain is found from $\Theta(\mathbb{R}^n)$ by taking a section $\theta_0 = 1$.

We have developed elsewhere[9] a simple and systematic method for minimizing Landau–Higgs potentials on $\mathbb{R}^n$. The method is algebraic and derives from the fact that an action of R is linear on $\mathbb{R}^n$. In such a case the method utilizes all the symmetry contained in Eq. (2) and often provides explicit solutions in $\mathbb{R}^n$. However, if one is interested in general results or in the stability of the solutions found, the methods of Refs. 5 and 8 are a useful complement. Therefore, we give here a direct and simple procedure for determining $\Theta(\mathbb{R}^n)$ and, consequently, the domain of orbit parameters. By using the same projection technique as in Ref. 9 we will also establish a connection between the three methods.

Let us assume that the isotropy (little) groups of R are known. They may be determined using the chain criterion[10] (for countable and, in particular, for space groups[11] or branching rules and some ingenuity[12,13] (for continuous compact groups). In the latter case the chain criterion is a (selective) necessary condition for a group to be an isotropy group. For each isotropy group L we will define a subspace Fix L of R whose every vector is invariant under L (L is the centralizer of Fix L, and Fix L is identical to an invariant slice). Since the action of R is linear Fix L is a linear subspace. We will emphasize this by calling Fix L an $i(L)$-dimensional symmetry plane in $\mathbb{R}^n$ [$i(L)$, equal to the number of linearly independent invariant vectors/singlets of L in R, is called the subduction frequency]. The linear equation of Fix L,

$$[1 - P(L)] \cdot \psi = 0 , \tag{6}$$

is easily obtained by using a projector

$$P(L) = \frac{1}{|L|} \sum_{r\in L} r , \tag{7}$$

where $(1/|L|)\Sigma_{r\in L}$ is an average over the little group L, $|L|$ is the order of L, and r is an element (a matrix) from $L \leqslant R$. The last equation is readily generalizable to a compact group L,

$$P(L) = \frac{1}{\Omega(L)} \int_L d\mu(r) r , \tag{7'}$$

where $\Omega(L) \equiv \int_L d\mu(r)$ is the group volume, $\int_L d\mu(r)$ is a group integral on L with appropriate measure $d\mu(r)$.[14]

Following Eq. (6) we obtain an equation for the whole family of hyperplanes Fix L associated with a conjugacy class $[L]$ of isotropy subgroups in R:

$$F[L;\psi] \equiv \prod_{L\in[L]} \|[1 - P(L)] \cdot \psi\|^2 = 0 . \tag{8}$$

In deriving Eq. (8) we used the fact that a vector Eq. (6) is zero if and only if its norm is zero and the fact that a product is zero if and only if one of its factors is zero. We note that the order of a class $[L]$ is equal to the order of an orbit $[R{:}N(L)]$ of $N(L)$ in R, $N(L)$ being the normalizer of L in R. A geometrical reason for this is that a plane Fix L is left invariant by $N(L)$. This relationship may be used to rewrite a product over $[L]$ in Eq. (8) as a product over the orbit $r \in [R{:}N(L)]$ provided ψ is replaced by $r \cdot \psi$ [r may be considered a coset representative in a coset decomposition of R with respect to $N(L)$]. Furthermore, by exponentiating, we change a product into a sum so that $F[L;\psi]$ reads

$$F[L;\psi] = \exp\left\{ \sum_{r\in[R:N(L)]} \ln \|[1 - \mathrm{P(L)}]\mathrm{r} \cdot \psi\|^2 \right\} . \tag{9}$$

The sum in Eq. (9) may be replaced by a sum over the whole group with a factor $|N(L)|^{-1}$. Such a form of $F[L;\psi]$ is, at least formally, generalizable to the case of a compact group

$$F[L;\psi] = \exp\left\{ \frac{\Omega(R)}{\Omega(N(L))} \times \int_R d\mu(r) \ln \|[1 - P(L)] \cdot r \cdot \psi\|^2 \right\} . \tag{8'}$$

Results of Ref. 5 suggest that even in the case of a compact group, an equation $F[L;\psi] = 0$ should be equivalent to a polynomial equation in ψ.

Using $F[L;\psi]$ we can specify in $\mathbb{R}^n$ a stratum $\Sigma[L;\psi]$ associated with a class $[L]$,

$$\sum [L;\psi] = \{\psi \in \mathbb{R}^n, F[L;\psi] = 0, \; \forall [L'] > [L], F[L';\psi] > 0\} , \tag{10}$$

which is a semialgebraic manifold in $\mathbb{R}^n$. In Eq. (10) the condition $\forall [L'] > [L]$, (i.e., "for every $[L']$ greater than $[L]$") is satisfied generally whenever it is already satisfied for the "immediately" greater $[L']$'s, i.e., for such $[L']$ that there is no $[L'']$, $[L'] > [L''] > [L]$ (partial order among $[L]$'s is defined on the basis of a subgroup-supergroup order among L's).

Our goal is to determine images of Eqs. (8) and (10) in $\mathbb{R}^m$. From the construction it is clear that $F[L;\psi]$ is an R-

TABLE I. $\hat{F}[L;\theta]$ for O_h (SO_7 adjoint).

$[L]$	$\hat{F}[L;\theta]$
$[O_h]$, $[SO_7]$	θ_0
$[C_{4v}]$, $[SO_5 \times U_1]$	$\frac{1}{3}(\theta_0^3 - \theta_2)$
$[C_{2v}]$, $[SO_3 \times SU_2 \times U_1]$	$(\frac{1}{2})^7(19\theta_0^6 - 45\theta_0^4\theta_1 + 60\theta_0^3\theta_2 - 69\theta_0^2\theta_1^2 + 48\theta_0\theta_1\theta_2 + \theta_1^3 - 6\theta_2^2)$
$[C_{3v}]$, $[SU_3 \times U_1]$	$(\frac{2}{9})^2(-7\theta_0^4 + 26\theta_0^2\theta_1 - 16\theta_0\theta_2 + \theta_1^2)$
$[C_s]$, $[SU_2 \times U_1 \times U_1]$	$(\frac{1}{2})^7\frac{1}{3}(-\theta_0^6 + 9\theta_0^4\theta_1 - 8\theta_0^3\theta_2 - 21\theta_0^2\theta_1^2 + 36\theta_0\theta_1\theta_2 + 3\theta_1^3 - 18\theta_2^2)$
$[C_s']$, $[SO_3 \times U_1 \times U_1]$	$\frac{1}{6}(\theta_0^3 - 3\theta_0\theta_1 + 2\theta_2)$
$[C_1]$, $[U_1 \times U_1 \times U_1]$	0

TABLE II. Strata $\hat{\Sigma}[L;\theta]$ for O_h (SO_7 adjoint).[a]

$[L]$	$\hat{\Sigma}[L;\theta]$
$[O_h]$, $[SO_7]$	$\theta_0=0, \theta_1=0, \theta_2=0$
$[C_{4v}]$, $[SO_5\times U_1]$	$\theta_0>0, \theta_1=\theta_0^2, \theta_2=\theta_0^3$
$[C_{2v}]$, $[SO_3\times SU_2\times U_1]$	$\theta_0>0, \theta_1=\frac{1}{2}\theta_0^2, \theta_2=\frac{1}{4}\theta_0^3$
$[C_{3v}]$, $[SU_3\times U_1]$	$\theta_0>0, \theta_1=\frac{1}{3}\theta_0^2, \theta_2=\frac{1}{9}\theta_0^3$
$[C_s]$, $[SU_2\times U_1\times U_1]$	$\theta_0>0, \theta_1>\frac{1}{3}\theta_0^2, \theta_2>\frac{1}{2}(-\theta_0^3+3\theta_0\theta_1), \theta_2=\theta_2^{\pm}$
$[C_s']$, $[SO_3\times U_1\times U_1]$	$\theta_0>0, \theta_0^2>\theta_1>\frac{1}{2}\theta_0^2, \theta_2=\frac{1}{2}(-\theta_0^3+3\theta_0\theta_1)$
$[C_1]$, $[U_1\times U_1\times U_1]$	$\theta_0>0, \theta_0^2>\theta_1>\frac{1}{3}\theta_0^2, \theta_2>\frac{1}{2}(-\theta_0^3+3\theta_0\theta_1), \theta_2^-<\theta_2<\theta_2^+$

[a] $\theta_2^{\pm}\equiv-\frac{2}{9}\theta_0^3+\theta_0\theta_1\pm(\theta_1-\frac{1}{3}\theta_0^2)[\frac{1}{6}\theta_1-\frac{1}{18}\theta_0^2]^{1/2}$.

invariant polynomial of degree $2|R|/|N(L)|$. Therefore, $F[L;\psi]$ may be written as a polynomial in θ which is automatically continued from $\Theta(\mathbb{R}^n)$ to $\mathbb{R}^m$,

$$\hat{F}[L;\theta]\equiv F[L;\psi]. \tag{11}$$

$\hat{F}[L;\theta]$ is most easily found by writing it as a polynomial in θ of degree $2|R|/|N(L)|$ in ψ. Arbitrary coefficients of the monomials in $\hat{F}[L;\theta]$ are determined from linear equations obtained by choosing a sufficient number of different (generic) values of ψ in Eq. (11) (they should not lie on the same or conjugated axes).

A substitution of Eq. (11) into Eq. (10) does not give *only* the image of $\Sigma[L;\psi]$ in $\mathbb{R}^m$. Although it gives a correct image in $\Theta(\mathbb{R}^n)$, it may also give some extra pieces elsewhere in $\mathbb{R}^m$. This is well illustrated by $\hat{F}[L;\theta]\geqslant 0$ being trivially satisfied on $\Theta(\mathbb{R}^n)$ but which may also be satisfied elsewhere in $\mathbb{R}^m$ [something similar was already encountered with positivity of the matrix $\hat{P}(\theta)$]. In order to eliminate as many such spurious pieces as possible we observe that $\hat{F}[L;\theta]=0$ implies that on $\Theta(\mathbb{R}^n)$ for every $[L'']<[L]$, $\hat{F}[L'';\theta]=0$. If we extend this to the whole $\mathbb{R}^m$ we obtain

$$\hat{\sum}[L;\theta]=\{\theta\in\mathbb{R}^m, \forall[L'']\leqslant[L], F[L'';\theta]=0, \forall[L']\not\leqslant[L], F[L';\theta]>0\}. \tag{12}$$

This stratification gives often the image of Eq. (10) without spurious parts. A difficulty with spurious pieces occurs in some cases when the ring of invariant polynomials is nonregular, i.e., there are algebraic relations among elements of the integrity basis. Such relations should be added to Eq. (12). Furthermore, if there are spurious pieces $\hat{\Sigma}[L,\theta]$ will be often given as an intersection of Eq. (12) and $\Theta(\mathbb{R}^n)$, $\Theta(\mathbb{R}^n)$ being determined in the following fashion. We choose any generic point $\psi^*\in\mathbb{R}^n$ and calculate its image $\theta^*=\Theta(\psi^*)$. $\Theta(\mathbb{R}^n)$ is then a connected part of the generic stratum, cf. Eq. (12), which contains θ^*, plus its boundary in the union (over $[L]$) of the varieties given by Eq. (12). Taking the above into account we write

$$\Theta(\mathbb{R}^n)=\bigcup_{[L]}\hat{\sum}[L;\theta]. \tag{13}$$

Further useful information for eliminating spurious pieces in the stratification Eq. (12) is that for every component θ_α of θ whose degree in ψ is d_α there are two numbers m_α and M_α such that $m_\alpha\theta_0^{d_\alpha/2}\leqslant\theta_\alpha\leqslant M_\alpha\theta_0^{d_\alpha/2}$ on $\Theta(\mathbb{R}^n)$.

Once the stratification, Eqs. (12) and (13), is determined the domain of orbit parameters is easily found by setting $\theta_0=1$ and making appropriate projections, as described earlier.

As an explicit example we consider the cubic group O_h which is relevant to the study of ferroelectric phase transitions in perovskites. The same group occurs in the study of the Higgs problem for the SO_7 adjoint representation.[9] Its

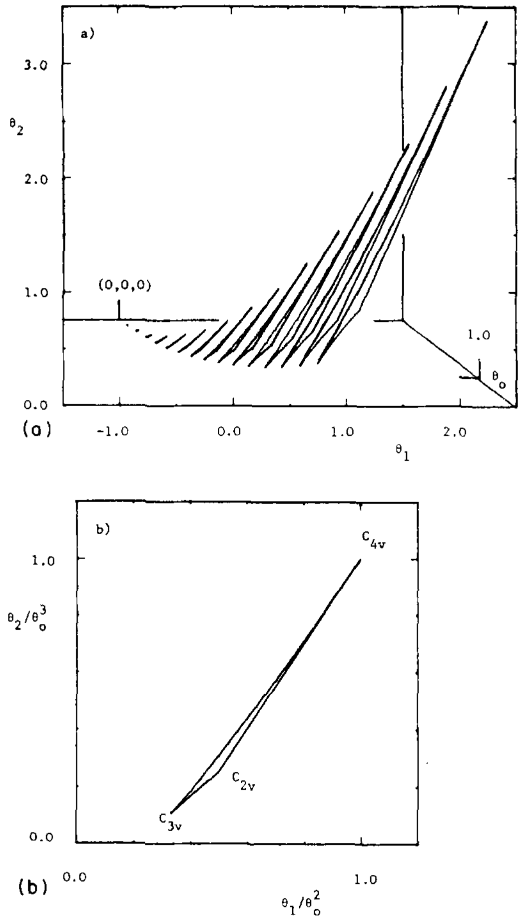


FIG. 1. Orbit space for O_h (SO_7 adjoint), cf. Table II: (a) sections θ_0 = const; (b) projective orbit space.

TABLE III. $\hat{F}[L;\theta]$ for T_d (SU_4 adjoint).

$[L]$	$\hat{F}[L;\theta]$
$[T_d]$, $[SU_4]$	θ_0
$[C_{2v}]$, $[SU_2\times SU_2\times U_1]$	$\frac{1}{2}[\theta_0^3-\theta_0\theta_2-2\theta_1^2]$
$[C_{3v}]$, $[SU_3\times U_1]$	$\frac{4}{81}[\theta_0^4+2\theta_0\theta_2-48\theta_0\theta_1^2+\theta_2^2]$
$[C_s]$, $[SU_2\times U_1\times U_1]$	$(\frac{1}{2})^8[-\theta_0^6+4\theta_0^4\theta_2+20\theta_0^3\theta_1^2-5\theta_0^2\theta_2^2-36\theta_0\theta_1^2\theta_2-108\theta_1^4+2\theta_2^3]$
$[C_1]$, $[U_1\times U_1\times U_1]$	0

isotropy groups and the integrity basis are[15] $[O_h]$, $[C_{4v}]$, $[C_{3v}]$, $[C_{2v}]$, $[C_s]$, $[C_s']$, and

$$\theta_\alpha=\sum_{i=1}^{3}\psi_i^{2\alpha+2},\quad \alpha=0,1,2\,. \tag{14}$$

Let us construct $\hat{F}[C_{4v};\theta]$. Choosing a fourfold rotation axis along ψ_3 direction we find, cf. Eq. (7),

$$P(C_{4v})=\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}. \tag{15}$$

Similar expressions for the three other axis lead, via Eq. (8), to

$$F[C_{4v};\psi]=(\psi_1^2+\psi_2^2)(\psi_2^2+\psi_3^2)(\psi_3^2+\psi_1^2)\,. \tag{16}$$

Since $F[C_{4v};\psi]$ is of 6th degree and O_h invariant it may be expressed as $\hat{F}[C_{4v};\theta]=a\theta_0^3+b\theta_0\theta_1+c\theta_2$. Coefficients a, b, and c are evaluated from a system of linear equations obtained by substituting three independent values of ψ in $\hat{F}[C_{4v};\theta]=F[C_{4v};\psi]$. This leads to $a=-c=\frac{1}{3}$, $b=0$, and

$$\hat{F}[C_{4v};\theta]=\tfrac{1}{3}(\theta_0^3-\theta_2)\,. \tag{17}$$

The results for other isotropy groups are collected in Table I. Using these results and Eq. (12) we find the strata which are listed in Table II[16] (no spurious pieces occur). The corresponding orbit space is shown in Fig. 1. A projection on the θ_1/θ_0^2 axis gives an "orbit space" for quartic potentials which clearly illustrates why a symmetry breaking may lead in this case only to isotropy groups $C_{4v}(SO_5\times U_1)$ or $C_{3v}(SU_3\times U_1)$ but not, e.g., to $C_{2v}(SO_3\times SU_2\times U_1)$.

The above results lead also to the results for the tetrahedral group T_d, Tables III, IV, and Fig. 2 (the integrity basis is $\theta_0=\Sigma\psi_i^2$, $\theta_1=\psi_1\psi_2\psi_3$ and $\theta_2=\Sigma\psi_i^4$). The same group occurs in the Higgs problem for the SU_4 adjoint representation for which the orbit space was previously determined in a heuristic fashion.[8]

In conclusion, we make several remarks.

The strata should be described by homogeneous polynomials in ψ, which is the case in the present method. It is not apparent that this is the case in Ref. 5. However, for $\theta_0>0$ the matrix $\hat{P}(\theta)$ may be multiplied on the right and left by the diagonal matrix diag $[\theta_0^{-1/2},\theta_0^{(1-d_1)/2},\theta_0^{(1-d_2)/2},\cdots]$. Since this matrix is positive definite the multiplication does not change the rank and positivity conditions (except rank $\hat{P}=0$) but replaces $\hat{P}(\theta)$ by a "dimensionless" matrix $\hat{P}(\lambda)$, $\lambda_\alpha\equiv\theta_\alpha/\theta_0^{d_\alpha/2}$. Therefore, we suggest that $\hat{P}(\theta)$ be replaced in Eqs. (4) and (5) by $\hat{P}(\lambda)$ and $\theta_0>0$ except for rank $\hat{P}=0\Rightarrow\theta=0$. Note that λ's are the "orbit parameters".[8]

Comparison of the conditions Eqs. (5) and (10) is somewhat simplified when R is a group generated by reflections. Then $m=n$ and $\mathrm{Sp}_nP=(\det\partial\theta)^2$ which is proportional to the square of a product of linear forms of the reflection hyperplanes in R.[17] Therefore, Sp_nP is proportional to $\Pi\,F[C_s;\psi]$ where the product is over all classes of isotropy groups $[C_s]$ consisting of the identity and a reflection. Consequently, $\mathrm{Sp}_n\hat{P}=0$ gives a boundary of the generic stratum [plus some extras outside $\theta(\mathbb{R}^n)$] just as a union of $\hat{F}[C_s;\theta]=0$ does. Otherwise, there is no obvious connection between Eqs. (5) and (10).

ACKNOWLEDGMENTS

I am grateful to Professor L. Michel and Professor R. T. Sharp for reading and commenting on the manuscript. I also

TABLE IV. Strata $\hat{\Sigma}[L;\theta]$ for T_d (SU_4 adjoint).[a]

$[L]$	$\hat{\Sigma}[L;\theta]$
$[T_d]$, $[SU_4]$	$\theta_0=0,\ \theta_1=0,\ \theta_2=0$
$[C_{2v}]$, $[SU_2\times SU_2\times U_1]$	$\theta_0>0,\ \theta_1=0,\ \theta_2=\theta_0^2$
$[C_{3v}]$, $[SU_3\times U_1]$	$\theta_0>0,\ \theta_1=(\frac{1}{3}\theta_0)^{3/2},\ \theta_2=\frac{1}{3}\theta_0^2$
$[C_s]$, $[SU_2\times U_1\times U_1]$	$\theta_0>0,\ [\theta_1=\theta_1^{\pm +},\ \theta_0^2>\theta_2>\frac{1}{3}\theta_0^2],\ [\theta_1=\theta_1^{\pm -},\ \frac{1}{2}\theta_0^2\geqslant\theta_2>\frac{1}{3}\theta_0^2]$
$[C_1]$, $[U_1\times U_1\times U_1]$	$\theta_0>0,\ [\theta_1^{++}>\theta_1>\theta_1^{+-},\ \theta_1^{--}>\theta_1>\theta_1^{-+},\ \frac{1}{2}\theta_0^2>\theta_2>\frac{1}{3}\theta_0^2]$, $[\theta_1^{++}>\theta_1>\theta_1^{-+},\ \theta_0^2>\theta_2\geqslant\frac{1}{2}\theta_0^2]$

[a] $\theta_1^{\pm\pm}\equiv\pm[\frac{5}{54}\theta_0^3-\frac{1}{6}\theta_2\theta_0\pm(\frac{1}{3}\theta^2-\frac{1}{9}\theta_0^2)(\frac{1}{6}\theta_2-\frac{1}{18}\theta_0^2)^{1/2}]^{1/2}$.

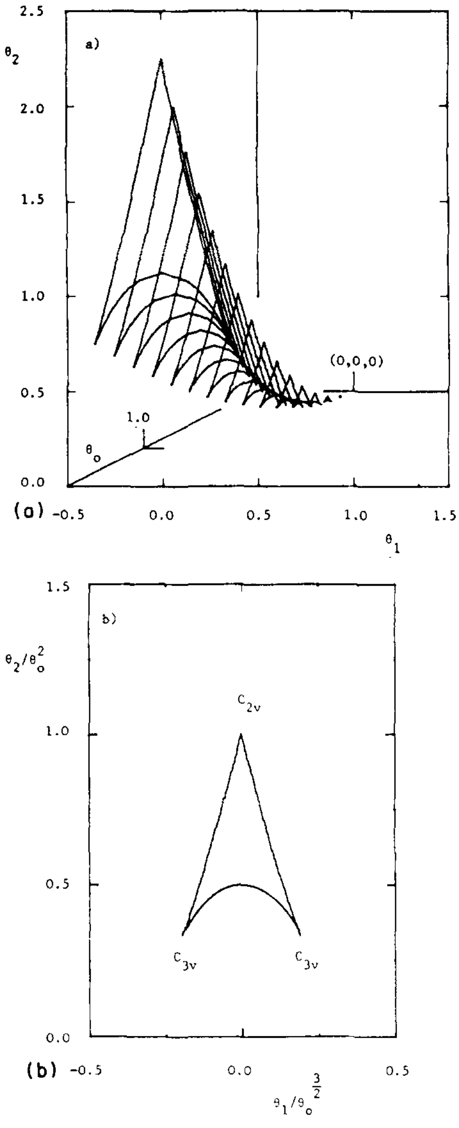


FIG. 2. Orbit space for T_d (SU_4 adjoint), cf. Table IV: (a) sections $\theta_0 = \text{const}$; (b) projective orbit space.

acknowledge stimulating discussions with Professor E. Bierstone, and I acknowledge an Alexander von Humboldt research fellowship.

---

[1] L. Michel, Primorsko Summer School, 1981 (Lecture Notes in Mathematics).

[2] Yu. M. Gufan, Fiz. Tverd. Tela **17**, 225 (1971) [Sov. Phys. Solid State **13**, 175 (1971)].

[3] By a plane we mean an i-dimensional, $0 \leqslant i \leqslant n$, linear subspace of $\mathbf{R}^n$. A symmetry plane is a plane left fixed by a subgroup L of R such that vectors perpendicular to it are transformed nontrivially under L. An axis of rotation and a reflection hyperplane are examples of 1- and $(n-1)$-dimensional symmetry planes, respectively.

[4] M. V. Jarić, *Lecture Notes in Physics*, Vol. 135 (Springer, New York, 1980); for a notion of orbit space see also L. Michel and L. A. Radicati, *Symmetry Principles at High Energy*, edited by Kursunoglu *et al.* (Benjamin, New York, 1968), p. 19 and L. Michel and L. A. Radicati, Ann. Phys. **66**, 758 (1971).

[5] M. Abud and G. Sartori, Phys. Lett. B **104**, 147 (1981).

[6] L. Michel, *Lecture Notes in Physics*, Vol. 6 (Springer, New York, 1970).

[7] For example, if the degrees of θ are d_α, $\alpha = 0,\dots, m-1$, then $\mathrm{Sp}_m\hat{P}$ is a homogeneous polynomial (in ψ) of degree $\Sigma_\alpha(2d_\alpha - 2)$. Consequently $\mathrm{Sp}_m\hat{P}$ is of a degree $\Sigma_\alpha(d_\alpha - 1)$ in $\theta_0 \equiv \|\psi\|^2$.

[8] J. S. Kim, Nucl. Phys. B **196**, 285 (1982); S. Frautschi and J. S. Kim, Nucl. Phys. B **196**, 301 (1982); J. S. Kim, Nucl. Phys. B **197**, 174 (1982).

[9] M. V. Jarić, Phys. Rev. Lett. **48**, 164 (1982).

[10] M. V. Jarić, Phys. Rev. B **23**, 3460 (1981).

[11] M. V. Jarić, Preprint IHES/P/82/17.

[12] R. C. King, J. Patera, and R. T. Sharp, J. Phys. A **15**, 1143 (1982).

[13] The approach of Ref. 5 may, in principle, be used to determine isotropy groups. However, it is not clear at present whether this is indeed feasible.

[14] Projecting need not be done by a "brute force" method, Eqs. (7) and (7'); see for example, R. Slansky, Phys. Rep. **79**, 1 (1981) and W. McKay and J. Patera, *Tables of Dimensions, Second and Fourth Indices and Branching Rules of Simple Algebras* (Dekker, New York, 1980).

[15] L. Michel and J. Mozrzymas, *Lecture Notes in Physics*, Vol. 79 (Springer, New York, 1978).

[16] The result, Table II, essentially agrees with Ref. 15 except for few corrections given here.

[17] L. Michel, *Proceedings of the 5th International Colloquium on Group Theory in Physics* (Academic, New York, 1977).

On the two-point functions of some integrable relativistic quantum field theories[a)]

S. N. M. Ruijsenaars[b)]
Department of Mathematics, Texas A&M University, College Station, Texas 77843

(Received 4 September 1981; accepted for publication 25 January 1982)

Two-point functions associated with the Federbush, massless Thirring, and continuum Ising models and their boson analogs are studied. In the Thirring case it is shown that the fields do not define operator-valued distributions, while temperedness of the two-point Wightman function is proved in the Ising case and in the Federbush case for a certain range of coupling constants. By relating the short-distance singularity of the Schwinger functions to the high-energy behavior of the spectral measures it is shown the fields cannot be made to satisfy the CCR/CAR by a rescaling. In the fermionic Federbush case this breakdown of the CAR occurs in spite of the fact that the fields correspond to a local Lagrangian.

PACS numbers: 11.10.Cd

I. INTRODUCTION

This paper is a continuation of previous work of the author on the Federbush, massless Thirring, and continuum Ising models and their boson analogs. In Ref. 1 we showed that the quantum fields of these models are normal ordered quadratic forms that are closely related or equal to the forms implementing improper Bogoliubov transformations generated by local and covariant classical field operators. We also studied the equations of motion and various other aspects, and discussed relations with work of other authors. In Ref. 2 we considered the scattering theory of these models at the classical and at the unphysical and physical quantum levels. The present work, some of whose results were announced in Ref. 1, deals with two-point Schwinger and Wightman functions arising in these models. The main issue we consider is whether the functional $\int dx\, F(x)\phi(x)\Omega$ [where Ω is the vacuum, $\phi(x)$ the quantum field, and F a test function in the Schwartz space $S(\mathbb{R}^2)$] corresponds to a vector in Fock space. In the case of the Federbush and Ising models and their boson analogs (studied in Secs. II and IV resp.) this is obvious if the Fourier transform $\widetilde{F}$ has compact support, but in the case of the Thirring model and its boson analog (Sec. III) we prove that for any $F \in S(\mathbb{R}^2)$ the functional is either zero or does not correspond to a vector in Fock space. In the former case the main problem is therefore to establish whether the corresponding two-point Wightman distribution in $\mathscr{D}'(\mathbb{R}^4)$ extends to a tempered distribution. Settling this question is a first step towards verification of the Wightman axioms[3] for these models.

As will be seen, the usual covariance and spectral properties are obvious, so that temperedness of the two-point Wightman function is equivalent to a polynomial short-distance singularity for the two-point Schwinger function. This follows from the theory of Laplace transforms,[3] but in the case at hand this can also be seen in a direct and illuminating way (cf. below). Thus, a large part of the work in Secs. II and IV is concerned with finding the dominant short-distance singularity of the various Schwinger functions. (For the fermionic Ising model of Sec. IV A we refer to the work of McCoy *et al.*[4]) A very useful tool in this study is a transformation to center of mass variables [cf. (2.22) and its generalization (2.54)], which is also tailor-made to study the measures in the Källen–Lehmann representations of the various two-point functions. [We found a similar transformation some time ago in a rather different context (cf. Ref. 5, pp. 418–9), but the advantage of the present transformation is that its Jacobian equals one.] By relating the short-distance singularity of the Schwinger function to the high-energy behavior of the spectral measure [through formulas like (2.37) below], we study existence of time-zero restrictions, and we find that for all interacting fields below the integral $\int d\rho(m)$ diverges, implying that none of these fields can be made to satisfy canonical (anti-) commutation relations by a rescaling. In the case of the fermionic Federbush model this holds true, although the fields are derived from a local Lagrangian. To our knowledge this is the first explicit example of a breakdown of the CAR for a Lagrangian field theory describing massive particles. Also, it follows that the field strength renormalization constant $Z \equiv [\int d\rho(m)]^{-1}$ vanishes, which indicates nonexistence of time-zero fields according to conventional wisdom. However, Challifour[6] has shown that (an infinite resummation of) the time-zero field considered in Sec. IIA exists in the usual axiomatic sense if the coupling constant is small enough. Thus, this constitutes a counterexample to this lore (provided the resummation involved does not change the fields).

Since we consider here the action of the fields on the vacuum, only the pure creation part of the fields is needed. Therefore, we refrain from giving complete definitions of the fields below, and refer instead to Ref. 1, where more references and background information can also be found. The fields of Secs. II and III act on a Fock space $\mathscr{F}_\epsilon(\mathscr{H}_1 \oplus \mathscr{H}_{-1})$, where $\epsilon =$ a (s) stands for antisymmetric (symmetric) in the fermion (boson) case and where $\mathscr{H}_1$ and $\mathscr{H}_{-1}$ are copies of a space $\mathscr{H} \equiv L^2(\mathbb{R}, d\theta)^2$. Thus we can write $\mathscr{H} = \mathscr{H}_+ \oplus \mathscr{H}_-$, where $\mathscr{H}_+$ and $\mathscr{H}_-$ are copies of $L^2(\mathbb{R}, d\theta)$. The spaces $\mathscr{H}_+$ and $\mathscr{H}_-$ are physically interpreted as state spaces of one-dimensional relativistic particles of positive and negative charge resp., described by rapidity wavefunctions, while the spaces $\mathscr{H}_1$ and $\mathscr{H}_{-1}$ correspond to particles of both charges but of different species ("species"

[a)] Work supported in part by NSF Grant PHY 78 23952.

[b)] Present address: Department of Mathematics, Tübingen University, 7400 Tübingen 1, West Germany.

meaning left- or right-moving in the Thirring case). In Sec. IV the relevant Fock spaces are $\mathscr{F}_\epsilon(\mathscr{H}_+)$, $\epsilon=$ a,s, corresponding to neutral particles of only one species. Throughout the paper the notation Kc^*c^* is used as shorthand for the Wick monomial $\int d\theta_1 d\theta_2\, K(\theta_1,\theta_2)c^*(\theta_1)c^*(\theta_2)$.

II. THE FEDERBUSH CASE

A. Fields on $\mathscr{F}_{\rm a}(\mathscr{H})$

We define for any $\lambda\in\mathbb{R}$

$$\phi^{\rm F}_\lambda(x)\Omega\equiv\exp(K^{{\rm F},x}_\lambda c^*_1 c^*_{-1})\Omega, \qquad (2.1)$$

where

$$K^{{\rm F},x}_\lambda(\theta_1,\theta_2)\equiv\exp(ix\cdot[\,p(\theta_1)+p(\theta_2)]K^{\rm F}_\lambda(\theta_1-\theta_2). \qquad (2.2)$$

Here

$$x\equiv(t,x^1), \qquad (2.3)$$

$$p(\theta)\equiv(\cosh\theta,\sinh\theta), \qquad (2.4)$$

the dot denotes the Lorentz inner product, and

$$K^{\rm F}_\lambda(\theta)\equiv i\,\frac{\sin\pi\lambda}{2\pi}\,e^{\lambda\theta}\,{\rm sech}\tfrac{1}{2}\theta. \qquad (2.5)$$

Theorem 2A.1: For any $\lambda\in\mathbb{R}$ and $t>0$ the Schwinger function

$$S^{\rm F}_\lambda(t)\equiv\|\phi^{\rm F}_\lambda(\tfrac{1}{2}it,0)\Omega\,\|^2 \qquad (2.6)$$

is finite-valued and satisfies

$$S^{\rm F}_\lambda(t)=\det[1+\Lambda^{\rm F}_\lambda(t)^*\Lambda^{\rm F}_\lambda(t)], \qquad (2.7)$$

where $\Lambda^{\rm F}(t)$ is the integral operator on $L^2(\mathbb{R})$ with kernel

$$\Lambda^{\rm F}_\lambda(t,\theta_1,\theta_2)=\frac{\sin\pi\lambda}{2\pi}$$
$$\times\exp[-\tfrac{1}{2}t(\cosh\theta_1+\cosh\theta_2)+\lambda(\theta_1-\theta_2)]$$
$$\times\ {\rm sech}[\tfrac{1}{2}(\theta_1-\theta_2)]. \qquad (2.8)$$

For any $\lambda\in(-\tfrac{1}{2},\tfrac{1}{2})$ there is an $N_\lambda>0$ so that

$$\lim_{t\to0} t^{N_\lambda}S^{\rm F}_\lambda(t)=0. \qquad (2.9)$$

Furthermore, for any $\lambda\in\mathbb{R}$,

$$S^{\rm F}_\lambda(t)=1+O(e^{-2t}),\quad t\to\infty, \qquad (2.10)$$

and, for any $\lambda\in[-\tfrac{1}{4},\tfrac{1}{4}]$,

$$\lim_{t\to0}\frac{\ln[S^{\rm F}_\lambda(t)]}{\ln(1/t)}=2\lambda^2. \qquad (2.11)$$

Finally, for any $\lambda\in\mathbb{R}$, $S^{\rm F}_\lambda(t)$ admits a Källen–Lehmann representation

$$S^{\rm F}_\lambda(t)=1+\int_2^\infty d\rho^{\rm F}_\lambda(m)\,K_0(mt), \qquad (2.12)$$

where K_0 is the modified Bessel function

$$K_0(t)\equiv\int_0^\infty d\theta\exp[-t\cosh\theta],\quad t>0. \qquad (2.13)$$

Here, the measure satisfies for any $\lambda\in[-\tfrac{1}{4},\tfrac{1}{4}]$

$$\int_2^\infty\frac{d\rho^{\rm F}_\lambda(m)}{m^\gamma}\begin{cases}<\infty, & \forall\gamma>2\lambda^2,\\ =\infty, & \forall\gamma<2\lambda^2.\end{cases} \qquad (2.14)$$

Theorem 2A.2: For any $\lambda\in(-\tfrac{1}{2},\tfrac{1}{2})$ the Wightman function

$$W^{\rm F}_\lambda(\widetilde{F},F)\equiv\|\phi^{\rm F}_\lambda(F)\Omega\,\|^2,\quad \widetilde{F}\in C^\infty_0(\mathbb{R}^2), \qquad (2.15)$$

extends to a tempered distribution. For any $\lambda\in[-\tfrac{1}{4},\tfrac{1}{4}]$ its time-zero restriction exists and defines a tempered distribution.

Conjecture 2A.3: The relations (2.11) and (2.14) hold for any $\lambda\in[-\tfrac{1}{2},\tfrac{1}{2}]$. For any noninteger λ with $|\lambda|>\tfrac{1}{2}$ the Schwinger function increases faster than t^{-N} for any $N>0$ as $t\to0$; equivalently, the Wightman function (2.15) does not extend to a tempered distribution.

Proof of Theorem 2A.1: A calculation shows that

$$S^{\rm F}_\lambda(t)=\sum_{n=0}^\infty\frac{1}{n!}\int d\theta_1\cdots d\theta_{2n}\sum_{\sigma\in S_n}(-)^\sigma$$
$$\times\prod_{i=1}^n\Lambda^{\rm F}_\lambda(t,\theta_i,\theta_{i+n})\Lambda^{\rm F}_\lambda(t,\theta_i,\theta_{\sigma(i)+n}), \qquad (2.16)$$

where S_n is the symmetric group. Since $\Lambda^{\rm F}_\lambda(t)$ is clearly Hilbert–Schmidt for any $t>0$ and $\lambda\in\mathbb{R}$, convergence of the series and (2.7) follow from Sec. 3 of Ref. 7. The bound (2.10) follows from the estimates

$$\det(1+T)\leqslant\exp(\|T\|_1),\quad T\geqslant0, \qquad (2.17)$$

and

$$\|\Lambda^{\rm F}_\lambda(t)\|_2^2=O(e^{-2t}),\quad t\to\infty, \qquad (2.18)$$

whose proof is obvious. The assertion (2.9) is a consequence of (2.17) and the fact that the lhs of (2.18) is bounded above by $C_\lambda\ \ln(1/t)$ for $|\lambda|<\tfrac{1}{2}$, as is readily verified.

To prove (2.11), we recall the well-known fact that

$$\ln\det(1+T)=-\sum_{n=1}^\infty\frac{(-)^n}{n}{\rm Tr}T^n, \qquad (2.19)$$

provided the rhs converges. To apply this to the case at hand, we shall first derive an upper bound to the function

$$a^{\rm F}_{n,\lambda}(t)\equiv{\rm Tr}[\Lambda^{\rm F}_\lambda(t)^*\Lambda^{\rm F}_\lambda(t)]^n$$
$$=(\sin\pi\lambda)^{2n}\int d\boldsymbol{\theta}\exp\Big(-t\sum_{j=1}^{2n}\cosh\theta_j\Big)$$
$$\times\prod_{i=1}^n h_\lambda(\theta_{2i-1}-\theta_{2i})h_\lambda(\theta_{2i+1}-\theta_{2i}), \qquad (2.20)$$

where $\theta_{2n+1}\equiv\theta_1$ and

$$h_\lambda(\theta)\equiv(1/2\pi)e^{\lambda\theta}\,{\rm sech}\tfrac{1}{2}\theta. \qquad (2.21)$$

To this end, we introduce a transformation to center of mass variables,

$$y_0=\ln\Big[\Big(\sum_{i=1}^k e^{\theta_i}\Big)\Big/M_k(\boldsymbol{\theta})\Big] \qquad (2.22)$$

$$y_j=\theta_{j+1}-\theta_j,\quad j=1,\ldots,k-1,$$

where M_k is the invariant mass,

$$M_k(\boldsymbol{\theta})\equiv\Big[\sum_{i,j=1}^k\cosh(\theta_i-\theta_j)\Big]^{1/2}. \qquad (2.23)$$

It is readily verified the Jacobian of this transformation is 1. Hence, setting

$$\widetilde{M}_k(\mathbf{y})\equiv M_k(\boldsymbol{\theta}(y_0,\mathbf{y}))$$
$$=\Big[k+2\sum_{\substack{i,j=0\\ i>j}}^{k-1}\cosh\Big(\sum_{l=j+1}^i y_l\Big)\Big]^{1/2},\quad \mathbf{y}\in\mathbb{R}^{k-1}, \qquad (2.24)$$

and using (2.13), it follows that

$$a^{\mathrm{F}}_{n,\lambda}(t) = 2(\sin\pi\lambda)^{2n}\int d\mathbf{y}\, K_0(t\widetilde{M}_{2n}(\mathbf{y}))$$
$$\times\prod_{i=1}^{n-1} h_\lambda(-y_{2i-1})h_\lambda(y_{2i})h_\lambda(-y_{2n-1})$$
$$\times h_\lambda\left(-\sum_{j=1}^{2n-1} y_j\right). \tag{2.25}$$

Since

$$\widetilde{M}_k(\mathbf{y})\geqslant k, \quad \forall\ \mathbf{y}\in\mathbb{R}^{k-1}, \tag{2.26}$$

we now get the upper bound

$$a^{\mathrm{F}}_{n,\lambda}(t)\leqslant 2(\sin\pi\lambda)^{2n}I_{n,\lambda}K_0(2nt), \tag{2.27}$$

where $I_{n,\lambda}$ is the cycle integral

$$I_{n,\lambda}\equiv\int d\phi_1\cdots d\phi_{2n-1}h_\lambda(-\phi_1)$$
$$\times\prod_{i=1}^{n-1} h_\lambda(\phi_{2i}-\phi_{2i-1})h_\lambda(\phi_{2i}-\phi_{2i+1})h_\lambda(-\phi_{2n-1}), \tag{2.28}$$

obtained from (2.25) by the transformation

$$\phi_k = \sum_{i=1}^{k} y_i, \quad k = 1,\ldots,2n-1, \tag{2.29}$$

which renders its convolution structure more transparent. Setting

$$\hat{h}_\lambda(x)\equiv\int d\theta \exp(i\theta x)h_\lambda(\theta), \quad |\lambda|<\tfrac{1}{2}, \tag{2.30}$$

one verifies by a contour integration that

$$\hat{h}_\lambda(x) = \operatorname{sech}[\pi(x-i\lambda)], \quad |\lambda|<\tfrac{1}{2}, \tag{2.31}$$

so that

$$I_{n,\lambda} = \frac{1}{2\pi}\int dx[\hat{h}_\lambda(x)\hat{h}_\lambda(-x)]^n$$
$$= \pi^{-2}2^{n-1}\int_0^\infty dx\,(\cosh x+\cos 2\pi\lambda)^{-n}, \quad |\lambda|<\tfrac{1}{2}. \tag{2.32}$$

As a result, the terms of the series $\Sigma(1/n)(-\sin^2\pi\lambda)^n I_{n,\lambda}$ diverge for $n\to\infty$ if $|\lambda|>\frac{1}{4}$, so that the series diverges for $|\lambda|>\frac{1}{4}$. However, using the relation

$$\ln(1-x) = -\sum_{n=1}^{\infty}\frac{1}{n}x^n, \quad |x|<1, \tag{2.33}$$

one infers that the series converges absolutely for $|\lambda|\leqslant\frac{1}{4}$ and that

$$-\sum_{n=1}^{\infty}\frac{1}{n}(-\sin^2\pi\lambda)^n I_{n,\lambda}$$
$$= \frac{1}{2\pi^2}\int_0^\infty dx\,\ln[(\cosh x+1)(\cosh x+\cos 2\pi\lambda)^{-1}]$$
$$=\lambda^2, \quad |\lambda|\leqslant\tfrac{1}{4}. \tag{2.34}$$

(Here we used the integral

$$\int_0^\infty dx\,\ln[(\cosh x+\cos\alpha)(\cosh x+\cos\beta)^{-1}]$$
$$= \int_\alpha^\beta d\phi\,\sin\phi\int_0^\infty dx\,\frac{1}{\cosh x+\cos\phi}$$
$$=\tfrac{1}{2}(\beta^2-\alpha^2), \quad |\alpha|,|\beta|\leqslant\pi, \tag{2.35}$$

where the last step can be verified by contour integration.) As a consequence we may write

$$\ln[S^{\mathrm{F}}_\lambda(t)] = -\sum_{n=1}^{\infty}\frac{1}{n}(-)^n a^{\mathrm{F}}_{n,\lambda}(t),$$
$$\forall\ t>0,\forall\ \lambda\in[-\tfrac{1}{4},\tfrac{1}{4}]. \tag{2.36}$$

We finally note that the function $K_0(mt)/\ln(1/t)$ is bounded for $t>0$ and $m\geqslant 2$ and has limit 1 for $t\to 0$ and any $m\geqslant 2$. The relation (2.11) therefore follows from (2.25), (2.34), and (2.36) by virtue of dominated convergence.

To prove the remaining claims, we first note that (2.12) can be obtained from (2.16) by making the change of variables (2.22) in each term of the series; each term contributes a measure $dF_{2n}(m)$, where $F_{2n}(m)$ is the result of omitting the y_0-dependent exponential factor and then integrating the internal variables over the region $\widetilde{M}_{2n}(\mathbf{y})\leqslant m$ in $\mathbb{R}^{2n-1}$ [note that $F_{2n}(m)=0$ for $m\leqslant 2n$]. Using (2.12), we can now connect the short-distance singularity of the Schwinger function and the high-energy behavior of the spectral measure by observing that we may write

$$\int_0^\infty dt\, t^{\delta}[S^{\mathrm{F}}_\lambda(t)-1] = \int_2^\infty\frac{d\rho^{\mathrm{F}}_\lambda(m)}{m^{\delta+1}}\int_0^\infty dx\, x^\delta K_0(x). \tag{2.37}$$

Assuming from now on that $|\lambda|\leqslant\frac{1}{4}$, it follows from (2.11) that the integral at the lhs converges/diverges for δ greater/smaller than $2\lambda^2-1$. Since the second factor on the rhs is finite for $\delta>-1$, (2.14) results.□

Proof of Theorem 2A.2: The first statement follows from (2.9) and general results on Laplace transforms,[3] but it is more illuminating to observe that for $|\lambda|<\frac{1}{2}$ one has

$$W^{\mathrm{F}}_\lambda(\bar{F},F)-|\tilde{F}(0)|^2$$
$$=\sum_{n=1}^{\infty}(n!)^{-2}\int d\theta_1\cdots d\theta_{2n}\left|\tilde{F}\left(\sum_{j=1}^{2n}p(\theta_j)\right)\sum_{\sigma\in S_n}(-)^\sigma\right.$$
$$\left.\times\sum_{i=1}^{n}K^{\mathrm{F}}_\lambda(\theta_i-\theta_{\sigma(i)+n})\right|^2$$
$$\leqslant\|F\|^2_\alpha\sum_{n=1}^{\infty}\frac{1}{n!}\int d\theta_1\cdots d\theta_{2n}\left(\sum_{j=1}^{2n}\cosh\theta_j\right)^{-\alpha}\sum_{\sigma\in S_n}(-)^\sigma$$
$$\times\prod_{i=1}^{n}\overline{K^{\mathrm{F}}_\lambda}(\theta_i-\theta_{i+n})K^{\mathrm{F}}_\lambda(\theta_i-\theta_{\sigma(i)+n})$$
$$=\|F\|^2_\alpha\Gamma(\alpha)^{-1}\int_0^\infty dt\, t^{\alpha-1}[S^{\mathrm{F}}_\lambda(t)-1]. \tag{2.38}$$

Here, $\|\cdot\|_\alpha$ is a Schwartz norm, and the last step follows from (2.16). In view of (2.9) and (2.10) the integral on the rhs converges for $|\lambda|<\frac{1}{2}$, provided $\alpha>N_\lambda$. This clearly implies the assertion.

Finally, we note that

$$W^{\mathrm{F}}_\lambda((0,\bar{f}),(0,f))-|\hat{f}(0)|^2$$
$$=\frac{1}{2}\int_2^\infty d\rho^{\mathrm{F}}_\lambda(m)\int d\theta\,|\hat{f}|^2(m\sinh\theta)$$
$$\leqslant\frac{1}{2}\int_2^\infty[d\rho^{\mathrm{F}}_\lambda(m)/m]\int dp\,|\hat{f}|^2(p), \tag{2.39}$$

where $\hat{f}$ is the Fourier transform of $f\in S(\mathbb{R})$. By virtue of (2.19) the first factor is finite for $|\lambda|\leqslant\frac{1}{4}$, so that the remaining statement follows. □

B. Fields on $\mathscr{F}_a(\mathscr{H}_1\otimes\mathscr{H}_{-1})$

We define for any $\lambda\in\mathbb{R}$

$$\psi_{\lambda,1}(x)\Omega\equiv\left(\frac{m(1)}{4\pi}\right)^{1/2}\int d\theta\, c^*_{1,-1}(\theta)$$
$$\times\begin{pmatrix}ie^{\theta/2}\\ -ie^{-\theta/2}\end{pmatrix}e^{im(1)x\cdot p(\theta)}\phi^{\mathrm{F}}_{\lambda}(m(-1)x)\Omega. \tag{2.40}$$

Here, $\phi^{\mathrm{F}}_{\lambda}\Omega$ is given by (2.1) with c^*_δ replaced by $c^*_{-1,\delta}$, and the masses $m(s)$ of the two different species $s=\pm1$ are strictly positive.

Theorem 2B.1: For any $\lambda\in\mathbb{R}$ and $t>0$ the Schwinger function

$$\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)\equiv\left\|\psi_{\lambda,1}\left(\frac{1}{2}it,0\right)\Omega\right\|^2 \tag{2.41}$$

is finite-valued and satisfies

$$\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)=m(1)S_{\mathrm{D}}(m(1)t)S^{\mathrm{F}}_{\lambda}(m(-1)t). \tag{2.42}$$

Here, S^{F}_{λ} is given by (2.7) and S_{D} is the Schwinger function of the free Dirac field of mass 1,

$$S_{\mathrm{D}}(t)\equiv\frac{1}{4\pi}\int d\theta\exp(-t\cosh\theta)\begin{pmatrix}e^{\theta}&-1\\-1&e^{-\theta}\end{pmatrix}. \tag{2.43}$$

For any $\lambda\in(-\frac{1}{2},\frac{1}{2})$ there is an $N'_\lambda>0$ such that

$$\lim_{t\to0}t^{N'_\lambda}\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)=0. \tag{2.44}$$

Furthermore, for any $\lambda\in\mathbb{R}$,

$$\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)=m(1)S_{\mathrm{D}}(m(1)t)+O(\exp(-[m(1)+2m(-1)]t)),$$
$$t\to\infty, \tag{2.45}$$

and, for any $\lambda\in[-\frac{1}{4},\frac{1}{4}]$,

$$\lim_{t\to0}\frac{\ln[\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)]}{\ln(1/t)}=\begin{pmatrix}1+2\lambda^2&2\lambda^2\\2\lambda^2&1+2\lambda^2\end{pmatrix}. \tag{2.46}$$

Finally, for any $\lambda\in\mathbb{R}$, $\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)$ has a spectral representation

$$\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)=\frac{m(1)}{2\pi}\begin{pmatrix}K_1(m(1)t)&-K_0(m(1)t)\\-K_0(m(1)t)&K_1(m(1)t)\end{pmatrix}$$
$$+\int_{m(1)+2m(-1)}^{\infty}m\,d\rho^{\mathrm{F}}_{1,\lambda,1}(m)\begin{pmatrix}K_1(mt)&0\\0&K_1(mt)\end{pmatrix}$$
$$-d\rho^{\mathrm{F}}_{2,\lambda,1}(m)\begin{pmatrix}0&K_0(mt)\\K_0(mt)&0\end{pmatrix}, \tag{2.47}$$

where

$$K_1(t)\equiv-K'_0(t)=\int_0^\infty d\theta\cosh\theta\exp(-t\cosh\theta),\quad t>0. \tag{2.48}$$

Here, the measures are positive and satisfy the inequalities

$$m(1)d\rho_1(m)<d\rho_2(m) \tag{2.49}$$

and

$$m^2d\rho_1(m)>m(1)d\rho_2(m). \tag{2.50}$$

Moreover, for $\lambda\in[-\frac{1}{4},\frac{1}{4}]$ and $j=1,2$,

$$\int_{m(1)+2m(-1)}^{\infty}\frac{d\rho^{\mathrm{F}}_{j,\lambda,1}(m)}{m^\gamma}\begin{cases}<\infty,&\forall\gamma>2\lambda^2,\\=\infty,&\forall\gamma<2\lambda^2.\end{cases} \tag{2.51}$$

Theorem 2B.2: For any $\lambda\in(-\frac{1}{2},\frac{1}{2})$ the Wightman function

$$\mathscr{W}^{F}_{\lambda,1}(\bar{F},F)\equiv\|\psi_{\lambda,1}(F)\Omega\|^2,\quad \tilde{F}\in C_0^\infty(\mathbb{R}^2)^2, \tag{2.52}$$

extends to a tempered distribution. For any $\lambda\in[-\frac{1}{4},\frac{1}{4}]$ its off-diagonal elements have a tempered time-zero restriction. Its diagonal elements do not admit a time-zero restriction for any noninteger $\lambda\in\mathbb{R}$.

Remark 2B.3: As we have shown in Ref. 1, the fields $\psi_{\lambda,1}$ and $\psi_{\lambda,-1}$ fail to satisfy the Federbush equation of motion, but are presumably local in the usual axiomatic sense for $|\lambda|\leqslant\frac{1}{2}$. By omitting the kernel $e^{\lambda\theta}$ in $\phi^{\mathrm{F}}_{\lambda}$ one obtains fields that solve it, but these fields are most likely nonlocal. The results of Sec. IIA and IIB can be easily extended to this case, and lead to similar qualitative properties. We leave the details to the interested reader.

Proof of Theorem 2B.1: The only assertions that do not immediately follow from Theorem 2A.1 concern the spectral representation and the measures occurring in it. To prove that they hold true, we note first that we may write $\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t)$ as

$$\sum_{n=0}^{\infty}\int d\theta_1\cdots d\theta_{2n+1}\begin{pmatrix}e^{\pm\theta_1}&-1\\-1&e^{\pm\theta_1}\end{pmatrix}$$
$$\times\exp\left(-t\left[m(1)\cosh\theta_1+m(-1)\sum_{j=2}^{2n+1}\cosh\theta_j\right]\right)$$
$$\times H_{n,\lambda}(\theta_2,\ldots,\theta_{2n+1}). \tag{2.53}$$

Next, we introduce a generalization of the transformation (2.22) to the case where the masses m_i corresponding to the rapidities θ_i are not necessarily equal to 1:

$$y_0=\ln\left[\left(\sum_{i=1}^{k}\right)m_i\,e^{\theta_i}\Big/M_k(\mathbf{m},\boldsymbol{\theta})\right]$$
$$y_j=\theta_{j+1}-\theta_j,\quad j=1,\ldots,k-1, \tag{2.54}$$

where

$$M_k(\mathbf{m},\boldsymbol{\theta})\equiv\left[\sum_{i,j=1}^{k}m_i\,m_j\cosh(\theta_i-\theta_j)\right]^{1/2}. \tag{2.55}$$

Since the Jacobian is still equal to 1, the representation (2.47) now follows as in the proof of Theorem 2A.1, $M_{2n+1}((m(1),m(-1),\ldots,m(-1)),\boldsymbol{\theta}(y_0,\mathbf{y}))$ playing the role of $\mathbf{M}_{2n}(\mathbf{y})$. The inequalities (2.49) and (2.50) are easily seen to hold after one substitutes center of mass variables in $(2.53)_+$ and $(2.53)_-$, resp. Finally, (2.51) results from (2.46) by virtue of a formula analogous to (2.37).□

Proof of Theorem 2B.2: The first claim follows from the relation

$$\frac{m(1)}{4\pi}\sum_{n=0}^{\infty}\frac{1}{n!}\int d\theta_0\cdots d\theta_{2n}$$
$$\times\left(m(1)\cosh\theta_0+m(-1)\sum_{j=1}^{2n}\cosh\theta_j\right)^{-\alpha}$$
$$\times\begin{pmatrix}e^{\theta_0}&-1\\-1&e^{\theta_0}\end{pmatrix}\sum_{\sigma\in S_n}(-)^{\sigma}\prod_{i=1}^{n}\overline{K^{\mathrm{F}}_{\lambda}}(\theta_i-\theta_{i+n})$$
$$\times K^{\mathrm{F}}_{\lambda}(\theta_i-\theta_{\sigma(i)+n})$$
$$=\Gamma(\alpha)^{-1}\int_0^\infty dt\,t^{\alpha-1}\mathscr{S}^{\mathrm{F}}_{\lambda,1}(t) \tag{2.56}$$

as in the proof of Theorem 2A.2. The second assertion also follows as in that proof from (2.51) with $j=2$. For $|\lambda|\leqslant\frac{1}{4}$ the last statement is a consequence of the fact that the integral $\int d\rho_{1,\lambda,1}^{\rm F}(m)$ diverges. To prove it for any noninteger λ, we note that if we smear, e.g., the upper component of $\psi_{\lambda,1}(0,x^1)\Omega$ with some $f\in S(\mathbb{R})$, then the squared L^2-norm of the three-body component is proportional to

$$\int d\theta_0\, d\theta_1\, d\theta_2\, e^{\theta_0}\,|\hat{f}(m(1)\sinh\theta_0+m(-1)$$

$$\times(\sinh\theta_1+\sinh\theta_2))|^2 e^{2\lambda(\theta_1-\theta_2)}\operatorname{sech}^2[\tfrac{1}{2}(\theta_1-\theta_2)]$$

$$\geqslant\frac{2}{m(1)}\int d\theta\, e^{4\lambda\theta}\operatorname{sech}^2\theta\int d\phi\int_{2m(-1)\sinh\phi\cosh\theta}^{\infty} dp|\hat{f}(p)|^2. \tag{2.57}$$

Thus, if $f\neq 0$, the ϕ-integral diverges at $-\infty$ for any fixed θ, and therefore the L^2-norm is infinite by Fubini's theorem.□

C. Fields on $\mathscr{F}_{\rm s}(\mathscr{H})$

We define for any $\lambda\in\mathbb{R}$,

$$\phi_\lambda^{\rm B}(x)\Omega\equiv\exp(K_\lambda^{{\rm B},x}c_1^* c_{-1}^*)\Omega, \tag{2.58}$$

where

$$K_\lambda^{{\rm B},x}(\theta_1,\theta_2)\equiv\exp\{ix\cdot[p(\theta_1)+p(\theta_2)]\}\,K_\lambda^{\rm B}(\theta_1-\theta_2). \tag{2.59}$$

Here we have used the same notation as in (2.2), and

$$K_\lambda^{\rm B}(\theta)\equiv\frac{\sin\pi\lambda}{2\pi}e^{(\lambda-1/2)\theta}\operatorname{sech}\tfrac{1}{2}\theta. \tag{2.60}$$

Theorem 2C.1: For any $\lambda\in[0,1]$ and $t>0$ the Schwinger function

$$S_\lambda^{\rm B}(t)\equiv\|\phi_\lambda^{\rm B}(\tfrac{1}{2}it,0)\Omega\|^2 \tag{2.61}$$

is finite-valued and satisfies

$$S_\lambda^{\rm B}(t)=\{\det[1-\Lambda_\lambda^{\rm B}(t)^*\Lambda_\lambda^{\rm B}(t)]\}^{-1}, \tag{2.62}$$

where $\Lambda_\lambda^{\rm B}(t)$ is the integral operator on $L^2(\mathbb{R})$ with kernel

$$\Lambda_\lambda^{\rm B}(t,\theta_1,\theta_2)=\frac{\sin\pi\lambda}{2\pi}$$

$$\times\exp[-\tfrac{1}{2}t(\cosh\theta_1+\cosh\theta_2)+(\lambda-\tfrac{1}{2})(\theta_1-\theta_2)]$$

$$\times\operatorname{sech}[\tfrac{1}{2}(\theta_1-\theta_2)]. \tag{2.63}$$

For any $\lambda\in\mathbb{R}$,

$$S_\lambda^{\rm B}(t)=1+O(e^{-2t}),\quad t\to\infty, \tag{2.64}$$

and, for any $\lambda\in[0,1]$,

$$\lim_{t\to 0}\frac{\ln[S_\lambda^{\rm B}(t)]}{\ln(1/{\rm t})}=2(\lambda-\lambda^2). \tag{2.65}$$

Moreover, for any noninteger λ not in $(0,1)$ there is a $C_\lambda>0$ such that

$$S_\lambda^{\rm B}(t)=\infty,\quad\forall\, t\in(0,C_\lambda). \tag{2.66}$$

Finally, for any $\lambda\in\mathbb{R}$, $S_\lambda^{\rm B}(t)$ admits a spectral representation

$$S_\lambda^{\rm B}(t)=1+\int_2^\infty d\rho_\lambda^{\rm B}(m)\,K_0(mt), \tag{2.67}$$

where the measure satisfies

$$\int_2^\infty\frac{d\rho_\lambda^{\rm B}(m)}{m^\gamma}\begin{cases}<\infty, & \forall\,\gamma>2(\lambda-\lambda^2),\\ =\infty, & \forall\,\gamma<2(\lambda-\lambda^2).\end{cases} \tag{2.68}$$

for any $\lambda\in(0,1)$.

Theorem 2C.2: For any $\lambda\in(0,1)$ the Wightman function

$$W_\lambda^{\rm B}(\overline{F},F)\equiv\|\phi_\lambda^{\rm B}(F)\Omega\|^2,\quad\widetilde{F}\in C_0^\infty(\mathbb{R}^2), \tag{2.69}$$

extends to a tempered distribution. Moreover, its time-zero restriction exists and defines a tempered distribution. Finally, if λ is not in $(0,1)$ and not in $\mathbb{Z}$, $W_\lambda^{\rm B}$ does not extend to a tempered distribution.

Proof of Theorem 2C.1: But for the assertion (2.66), the proof proceeds along similar lines as the proof of Theorem 2A.1. The analog of (2.16) is

$$S_\lambda^{\rm B}(t)=\sum_{n=0}^\infty\frac{1}{n!}\int d\theta_1\cdots d\theta_{2n}$$

$$\times\sum_{\sigma\in S_n}\prod_{i=1}^n\Lambda_\lambda^{\rm B}(t,\theta_i,\theta_{i+n})\Lambda_\lambda^{\rm B}(t,\theta_i,\theta_{\sigma(i)+n}). \tag{2.70}$$

Again, the Hilbert–Schmidt property of $\Lambda_\lambda^{\rm B}(t)$ for any $t>0$ and $\lambda\in\mathbb{R}$ is obvious, but in this case convergence of the series and (2.62) only follow from Sec. 3 of Ref. 7 provided

$$\|\Lambda_\lambda^{\rm B}(t)\|<1. \tag{2.71}$$

Assuming from now on that $\lambda\in(0,1)$, we may write

$$\Lambda_\lambda^{\rm B}(t)=e^{-tH_0/2}K_\lambda^{\rm B}e^{-tH_0/2}, \tag{2.72}$$

where H_0 is multiplication by $\cosh\theta$ and $K_\lambda^{\rm B}$ the convolution operator with kernel $K_\lambda^{\rm B}(\theta)$, whose norm is 1, since it turns into multiplication by $-i\sin\pi\lambda\operatorname{csch}[\pi(x-i\lambda)]$ upon Fourier transformation [cf. (2.21) and (2.31)]. Hence, (2.71) is satisfied for any $t>0$, implying the first statement. The bound (2.64) follows in the same way as (2.10), using instead of (2.17) the estimate

$$[\det(1-T)]^{-1}\leqslant\exp(c^{-1}\|T\|_1),\quad 0\leqslant T\leqslant 1-c<1, \tag{2.73}$$

whose proof is easy. [Note that (2.71) holds for any $\lambda\in\mathbb{R}$ if one chooses t sufficiently large, since $\|\Lambda_\lambda^{\rm B}(t)\|_2\to 0$ for $t\to\infty$ and $\|\cdot\|\leqslant\|\cdot\|_2$.]

We proceed to prove (2.65), using the relation

$$\ln[\det(1-T)]^{-1}=\sum_{n=1}^\infty\frac{1}{n}{\rm Tr}T^n,\quad 0\leqslant T<1. \tag{2.74}$$

Introducing

$$a_{n,\lambda}^{\rm B}(t)\equiv{\rm Tr}[\Lambda_\lambda^{\rm B}(t)^*\Lambda_\lambda^{\rm B}(t)]^n, \tag{2.75}$$

one obtains as the analog of (2.27) the bound

$$a_{n,\lambda}^{\rm B}(t)\leqslant 2(\sin\pi\lambda)^{2n}I_{n,\lambda-1/2}K_0(2nt). \tag{2.76}$$

Also, using the method leading to (2.34) and the integral (2.35), one obtains

$$\sum_{n=1}^\infty\frac{1}{n}(\sin\pi\lambda)^{2n}I_{n,\lambda-1/2}=\lambda-\lambda^2. \tag{2.77}$$

From this, (2.65) follows by the same arguments as those used in the proof of (2.11).

To prove the assertion (2.66), it suffices to show that the bound (2.71) is violated when λ is outside $(0,1)$ and not an integer, and when t is taken small enough. [Indeed, from Sec. 3 in Ref. 7 it is readily seen that this condition is not only sufficient, but also necessary for the convergence of the series at the rhs of (2.70).] To prove this, let us assume that λ is greater than 1 and not an integer (the proof for the other case is similar). Denote the characteristic functions of the intervals $(\ln(1/t),\ln(1/t)+1)$ and $(-\ln(1/t)-1,-\ln(1/t))$ by χ_t^+ and χ_t^-, resp. Then one has

$$\begin{aligned}|(\chi_t^+,\mathcal{A}_\lambda^{\mathrm{B}}(t)\chi_t^-)| &= \frac{|\sin\pi\lambda|}{2\pi}\int_{\ln(1/t)}^{\ln(1/t)+1} d\theta_1 d\theta_2 \\ &\quad\times\exp[-\tfrac{1}{2}t(\cosh\theta_1+\cosh\theta_2) \\ &\quad+(\lambda-\tfrac{1}{2})(\theta_1+\theta_2)]\operatorname{sech}[\tfrac{1}{2}(\theta_1+\theta_2)] \\ &> \frac{|\sin\pi\lambda|}{2\pi}\cdot e^{-e}\cdot\left(\frac{1}{t}\right)^{2(\lambda-1)}.\end{aligned} \tag{2.78}$$

Since $\|\chi_t^\pm\|=1$ and the rhs diverges for $t\to 0$, it follows that (2.71) is false for t small enough.

The remaining statements follow as in the proof of Theorem 2A.1, so that the theorem is proven. □

Proof of Theorem 2C.2: The first claim follows from (2.64) and (2.65) and the relation

$$\begin{aligned}&\sum_{n=1}^{\infty}\frac{1}{n!}\int d\theta_1\cdots d\theta_{2n}\left(\sum_{j=1}^{2n}\cosh\theta_j\right)^{-\alpha} \\ &\quad\times\sum_{\sigma\in S_n}\prod_{i=1}^{n}K_\lambda^{\mathrm{B}}(\theta_i-\theta_{i+n})K_\lambda^{\mathrm{B}}(\theta_i-\theta_{\sigma(i)+n}) \\ &\quad=\Gamma(\alpha)^{-1}\int_0^\infty dt\, t^{\alpha-1}[S_\lambda^{\mathrm{B}}(t)-1]\end{aligned} \tag{2.79}$$

[cf. (2.38)]. The assertion concerning the time-zero function follows as in Theorem 2A.2 from (2.65). The validity of the last statement can be seen as follows: By virtue of the uniform boundedness principle, convergence of the series on the lhs of (2.79) for sufficiently large α is not only sufficient, but also necessary for the temperedness of W_λ^{B}. However, in view of (2.66), the series diverges for any α. Stronger yet, it is readiy seen that the terms corresponding to the identity permutation already diverge if $n>(4\lambda-4)^{-1}\alpha$ (assuming $\lambda>1$, e.g.). Indeed, one has

$$\begin{aligned}&\int d\theta_1\cdots d\theta_{2n}\left(\sum_{j=1}^{2n}\cosh\theta_j\right)^{-\alpha} \\ &\quad\times\prod_{i=1}^{n}\exp[(2\lambda-1)(\theta_i-\theta_{i+n})]\operatorname{sech}^2[\tfrac{1}{2}(\theta_i-\theta_{i+n})] \\ &\quad\geqslant\int_0^\infty d\theta_1\cdots d\theta_{2n}\left(\sum_{j=1}^{2n}e^{\theta_j}\right)^{-\alpha}\exp\left[(2\lambda-2)\sum_{i=1}^{2n}\theta_i\right] \\ &\quad\geqslant\frac{(2n)!}{(2n)^\alpha}\int_0^\infty d\theta_1\, e^{(2\lambda-2)\theta_1}\cdots\int_{\theta_{2n-1}}^\infty d\theta_{2n}\, e^{(2\lambda-2)\theta_{2n}-\alpha\theta_{2n}} \\ &\quad=\infty,\quad n>(4\lambda-4)^{-1}\alpha,\ \lambda>1,\end{aligned} \tag{2.80}$$

which proves the assertion. □

D. Fields on $\mathscr{F}_{\mathrm{s}}(\mathscr{H}_1\otimes\mathscr{H}_{-1})$

We define for any $\lambda\in\mathbb{R}$

$$\begin{aligned}\phi_{\lambda,1}(x)\Omega&\equiv-(4\pi)^{-1/2} \\ &\quad\times\int d\theta\, c_{1,-1}^*(\theta)e^{im(1)x\cdot p(\theta)}\phi_\lambda^{\mathrm{B}}(m(-1)x)\Omega,\end{aligned} \tag{2.81}$$

where $\phi_\lambda^{\mathrm{B}}\Omega$ is given by (2.58) with c_δ^* replaced by $c_{-1,\delta}^*$.

Theorem 2D.1: For any $\lambda\in[0,1]$ and $t>0$ the Schwinger function

$$\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)\equiv\|\phi_{\lambda,1}(\tfrac{1}{2}it,0)\Omega\|^2 \tag{2.82}$$

is finite-valued and satisfies

$$\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)=S_{\mathrm{KG}}(m(1)t)S_\lambda^{\mathrm{B}}(m(-1)t). \tag{2.83}$$

Here, S_λ^{B} is given by (2.62) and S_{KG} is the Schwinger function of the free Klein–Gordon field of mass 1,

$$S_{\mathrm{KG}}(t)=\frac{1}{4\pi}\int d\theta\exp(-t\cosh\theta). \tag{2.84}$$

For any $\lambda\in\mathbb{R}$,

$$\begin{aligned}\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)&=S_{\mathrm{KG}}(m(1)t)+O(\exp(-[m(1) \\ &\quad+2m(-1)]t)),\quad t\to\infty,\end{aligned} \tag{2.85}$$

and, for any $\lambda\in[0,1]$,

$$\lim_{t\to 0}\frac{\ln[\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)]}{\ln(1/t)}=2(\lambda-\lambda^2). \tag{2.86}$$

Moreover, for any noninteger λ not in (0,1) there is a $C_\lambda'>0$ such that

$$\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)=\infty,\quad\forall\ t\in(0,C_\lambda'). \tag{2.87}$$

Finally, for any $\lambda\in\mathbb{R}$, $\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)$ admits a spectral representation

$$\begin{aligned}\mathscr{S}_{\lambda,1}^{\mathrm{B}}(t)&=\frac{1}{2\pi}K_0(m(1)t) \\ &\quad+\int_{m(1)+2m(-1)}^\infty d\rho_{\lambda,1}^{\mathrm{B}}(m)\,K_0(mt),\end{aligned} \tag{2.88}$$

where the measure satisfies (the analog of) (2.68) for any $\lambda\in(0,1)$.

Proof: It is readily seen that the assertions follow from Theorem 2C.1 and the behavior of $K_0(t)$ for $t\to 0$ and $t\to\infty$. □

Theorem 2D.2: For any $\lambda\in(0,1)$ the Wightman function

$$\mathscr{W}_{\lambda,1}^{\mathrm{B}}(\overline{F},F)\equiv\|\phi_{\lambda,1}(F)\Omega\|^2,\quad\widetilde{F}\in C_0^\infty(\mathbb{R}^2), \tag{2.89}$$

extends to a tempered distribution, which has a tempered time-zero restriction. For noninteger λ outside (0,1), $\mathscr{W}_{\lambda,1}^{\mathrm{B}}$ does not extend to a tempered distribution.

Proof: Using the analog of (2.56), the proof proceeds along the same lines as the proof of Theorem 2C.2.

III. THE THIRRING CASE

A. Fields on $\mathscr{F}_\epsilon(\mathscr{H}_s)$, $\epsilon=$ a,s

We define for any $\lambda\in\mathbb{R}$ and $s=\pm 1$,

$$\phi_{\lambda,s}^{A_0}(x)\Omega\equiv\exp(K_{\lambda,s}^{A_0,x}c_{s,1}^*c_{s,-1}^*)\Omega,\quad A=\mathrm{B,F}, \tag{3.1}$$

where

$$K_{\lambda,s}^{A_0,x}(\theta_1,\theta_2)\equiv\exp(ix\cdot[p_s(\theta_1)+p_s(\theta_2)])K_{-\lambda s}^{A}(\theta_1-\theta_2). \tag{3.2}$$

Here, the function $K_\lambda^A(\theta)$ is defined by (2.5) and (2.60) for $A=\mathrm{F}$ and $A=\mathrm{B}$, resp., and

$$p_s(\theta)\equiv(e^{s\theta},se^{s\theta}). \tag{3.3}$$

We also define the light rays

$$R_s\equiv\{(p^0,p^1)\in\mathbb{R}^2|sp^1=p^0\geqslant 0\}. \tag{3.4}$$

Theorem 3A.1: Let $F \in S(\mathbb{R}^2)$. If $\widetilde{F}$ vanishes on R_s, one has

$$\|\phi^{A_0}_{\lambda,s}(F)\Omega\| = 0, \quad A = \mathrm{F,B}. \tag{3.5}$$

If $\widetilde{F}(\bar{p}_s) \neq 0$ for some $\bar{p}_s \in R_s$, one has

$$\|\phi^{A_0}_{\lambda,s}(F)\Omega\| = \infty, \quad \forall\ \lambda \notin \mathbb{Z}, \quad A = \mathrm{F,B}. \tag{3.6}$$

Proof: The first statement is clear. To prove (3.6) for $A = \mathrm{F}$, we first note that the squared L^2-norm of the two-body component of $\phi^{\mathrm{F}_0}_{\lambda,s}(F)\Omega$ is proportional to

$$\begin{aligned}&\int d\theta_1 d\theta_2 |\widetilde{F}(p_s(\theta_1) + p_s(\theta_2))|^2 \\ &\quad \times \exp[-2\lambda s(\theta_1 - \theta_2)]\,\mathrm{sech}^2[\tfrac{1}{2}(\theta_1 - \theta_2)] \\ &\equiv \int dx dy\, f(2e^x \cosh y) e^{4|\lambda|y}\,\mathrm{sech}^2 y \\ &= \int du dy\, f(e^u + e^{u-2y}) e^{4|\lambda|y}\,\mathrm{sech}^2 y,\end{aligned} \tag{3.7}$$

where $f \geqslant 0$ and $f(\bar{p}^0_s) > 0$. Thus, if $|\lambda| \geqslant \frac{1}{2}$, the y-integral diverges at ∞ for any u in a neighborhood of $u = \ln \bar{p}^0_s$, so that the integral diverges by Fubini's theorem. We may therefore assume that $|\lambda| < \frac{1}{2}$. In this case the integral (3.7) may be convergent, but we claim that then the four-body component is not square integrable. To show this, we note its L^2-norm squared is proportional to

$$\begin{aligned}&\int d\theta_1 \cdots d\theta_4 f(e^{\theta_1} + \cdots + e^{\theta_4}) \\ &\quad \times |h_\lambda(\theta_1 - \theta_2)h_\lambda(\theta_3 - \theta_4) - h_\lambda(\theta_1 - \theta_4)h_\lambda(\theta_3 - \theta_2)|^2,\end{aligned} \tag{3.8}$$

where h_λ is defined by (2.21). One obtains a lower bound to this integral if one replaces f by $g_\epsilon f$, where g_ϵ is continuous, $0 \leqslant g_\epsilon \leqslant 1$, $g_\epsilon(x) = 1$ for $x > 2\epsilon$ and $g_\epsilon(x) = 0$ for $x < \epsilon$. Hence, to prove it diverges, we may as well assume that $f(x)$ vanishes for $x < \epsilon$ and that $\bar{p}^0_s > 2\epsilon$. We now write the integral as

$$\begin{aligned}&2\int d\theta_1 \cdots d\theta_4 f(e^{\theta_1} + \cdots + e^{\theta_4})[h^2_\lambda(\theta_1 - \theta_2)h^2_\lambda(\theta_3 - \theta_4) \\ &\quad - h_\lambda(\theta_1 - \theta_2)h_\lambda(\theta_3 - \theta_2)h_\lambda(\theta_3 - \theta_4)h_\lambda(\theta_1 - \theta_4)] \\ &\quad = c\int dx_1 dx_2 dy_1 dy_2 f(2e^{x_1}\cosh y_1 + 2e^{x_2}\cosh y_2) \\ &\qquad \times \exp[4|\lambda|(y_1 + y_2)]\,\mathrm{sech}^2 y_1\,\mathrm{sech}^2 y_2 \\ &\qquad - \int dy_0 \cdots dy_3 f(\widetilde{M}_4(\mathbf{y})e^{y_0})h_\lambda(-y_1)h_\lambda(y_2) \\ &\qquad \times h_\lambda(-y_3)h_\lambda(-y_1 - y_2 - y_3) \\ &\quad \equiv I_1 - I_2,\end{aligned} \tag{3.9}$$

where we have changed to the center of mass variables (2.22) in I_2. If x_2, y_2 in I_1 are chosen such that $2e^{x_2}\cosh y_2$ is in a neighborhood of $\bar{p}^0_s$, the x_1-integration diverges at $-\infty$ for any y_1. Hence, $I_1 = \infty$ by Fubini's theorem. Thus, it suffices to show I_2 is convergent. But by assumption $\operatorname{supp} f \subset [\epsilon, \infty)$, so that

$$\begin{aligned}I_2 &\leqslant C_\lambda \left[\int dy h_\lambda(y)\right]^3 \int_\epsilon^\infty \frac{dx}{x} f(x) \\ &< \infty,\end{aligned} \tag{3.10}$$

since $|\lambda| < \frac{1}{2}$ by assumption. This proves the theorem for $A = \mathrm{F}$. The proof for $A = \mathrm{B}$ can be simplified by noting that no minus sign occurs in the analog of (3.8), so that its divergence for any λ is a direct consequence of the divergence of (the analog of) I_1. □

B. Fields on $\mathscr{F}_\epsilon(\mathscr{H}_1 \otimes \mathscr{H}_{-1})$, $\epsilon = \mathrm{a,s}$

We define for any $\lambda \in \mathbb{R}$

$$\begin{aligned}&\psi^0_{\lambda,s}(x)\Omega \\ &\quad \equiv is(2\pi)^{-1/2}\int d\theta\, e^{s\theta/2} c^*_{s,-1}(\theta) e^{ix\cdot p_s(\theta)} \phi^{\mathrm{F}_0}_{\lambda,-s}(x)\Omega,\end{aligned} \tag{3.11}$$

which holds on $\mathscr{F}_\mathrm{a}$, and

$$\phi^0_{\lambda,s}(x)\Omega \equiv -(2\pi)^{-1/2}\int d\theta\, c^*_{s,-1}(\theta) e^{ix\cdot p_s(\theta)} \phi^{\mathrm{B}_0}_{\lambda,-s}(x)\Omega, \tag{3.12}$$

which holds on $\mathscr{F}_\mathrm{s}$. Here, $\phi^{A_0}_{\lambda,s}(x)\Omega$, $A = \mathrm{F,B}$, is given by (3.1). We also define the light cone

$$V_+ \equiv \{(p^0, p^1) \in R^2 | p^0 \geqslant |p_1|\}, \tag{3.13}$$

and denote the interior of V_+ by V^0_+.

Theorem 3B.1: Let $F \in S(\mathbb{R}^2)$. If $\tilde{F}$ vanishes on V_+, one has

$$\|\psi^0_{\lambda,s}(F)\Omega\| = \|\phi^0_{\lambda,s}(F)\Omega\| = 0. \tag{3.14}$$

If $\tilde{F}(\bar{p}) \neq 0$ for some $\bar{p} \in V^0_+$, one has

$$\|\psi^0_{\lambda,s}(F)\Omega\| = \|\phi^0_{\lambda,s}(F)\Omega\| = \infty, \quad \forall\ \lambda \notin \mathbb{Z}. \tag{3.15}$$

Remark 3B.2: For $\lambda \in [0, \frac{1}{2}]$ and $s = 1$ the fields considered here and in the preceding section coincide with the fields of Ref. 1. Analytic continuation in λ and the analogous definition for $s = -1$ lead to the above fields, which are, however, slightly different from the fields obtained from a consideration of the underlying Bogoliubov transformations. We introduced these fields to ease the notation and because it is natural to consider the analytically continued fields (as we also did in Sec. II). The fields of Ref. 1 and the fields obtained by omitting the kernel $e^{\lambda\theta}$ lead to the same results, as is easily verified.

Proof: We argue in a similar way as in the proof of Theorem 3A.1. The L^2-norm squared of the three-body component of $\psi^0_{\lambda,s}(x)\Omega$ is proportional to

$$\begin{aligned}&\int d\theta du dy\, G(e^u + e^{u-2y} + e^\theta, e^u \\ &\quad + e^{u-2y} - e^\theta) e^{\theta + 4|\lambda|y}\,\mathrm{sech}^2 y,\end{aligned} \tag{3.16}$$

where $G \geqslant 0$ and $G(\bar{p}^0, -s\bar{p}^1) > 0$. Hence, for $|\lambda| \geqslant \frac{1}{2}$ the y integral diverges at ∞ if one chooses u and θ such that $(e^u + e^\theta, e^u - e^\theta)$ is in a neighborhood of $(\bar{p}^0, -s\bar{p}^1)$, implying divergence of (3.16). For $|\lambda| < \frac{1}{2}$, the L^2-norm squared of the five-body component is bounded below by

$$\begin{aligned}&c\int_B d\boldsymbol{\theta}\, e^{s\theta_0}\left|\tilde{F}\Big(p_s(\theta_0) + \sum_{i=1}^4 p_{-s}(\theta_i)\Big)\right|^2 |h_{\lambda s}(\theta_1 - \theta_2)h_{\lambda s} \\ &\quad \times (\theta_3 - \theta_4) - h_{\lambda s}(\theta_1 - \theta_4)h_{\lambda s}(\theta_3 - \theta_2)|^2,\end{aligned} \tag{3.17}$$

where $\int_B d\boldsymbol{\theta}$ stands for the integral over the set of $\boldsymbol{\theta}$ for which the argument of $\tilde{F}$ belongs to a closed ball $B \subset V^0_+$ around $\bar{p}$. As this implies θ_0 ranges over a bounded interval, divergence of (3.17) follows in the same way as the divergence of (3.8), proving the theorem in the fermion case. Again, the proof is somewhat shorter in the boson case, since no minus sign occurs in the analog of (3.17), which immediately results in its divergence for any $\lambda \in \mathbb{R}$.□

IV. THE ISING CASE

A. Fields on $\mathscr{F}_a(\mathscr{H}_+)$

We define

$$\phi^{\mathrm{F}}_{-}(x)\Omega \equiv \exp(\tfrac{1}{2} K^{\mathrm{F},x} c^* c^*)\Omega , \tag{4.1}$$

where

$$K^{\mathrm{F},x}(\theta_1,\theta_2) \equiv \exp\{ix\cdot[\,p(\theta_1) + p(\theta_2)]\}K^{\mathrm{F}}(\theta_1 - \theta_2)\,. \tag{4.2}$$

Here, the first factor on the rhs is the same as in (2.2), and

$$K^{\mathrm{F}}(\theta) \equiv (i/2\pi)\tanh\tfrac{1}{2}\theta\,. \tag{4.3}$$

We also define

$$\phi^{\mathrm{F}}_{+}(x)\Omega \equiv i(4\pi)^{-1/2}\int d\theta\, c^*(\theta)e^{ix\cdot p(\theta)}\phi^{\mathrm{F}}_{-}(x)\Omega\,. \tag{4.4}$$

Theorem 4A.1: For any $t > 0$ the Schwinger function

$$S^{\mathrm{F}}_{-}(t) \equiv \|\phi^{\mathrm{F}}_{-}(\tfrac{1}{2}it,0)\Omega\|^2 \tag{4.5}$$

is finite-valued and satisfies

$$S^{\mathrm{F}}_{-}(t) = (\det[1 + \Lambda^{\mathrm{F}}(t)^*\Lambda^{F}(t)])^{1/2}\,, \tag{4.6}$$

where $\Lambda^{\mathrm{F}}(t)$ is the integral operator on $L^2(\mathbb{R})$ with kernel

$$\begin{aligned}\Lambda^{\mathrm{F}}(t,\theta_1,\theta_2) = (1/2\pi)\exp[-\tfrac{1}{2}t(\cosh\theta_1 + \cosh\theta_2)]\\ \times\tanh[\tfrac{1}{2}(\theta_1 - \theta_2)]\,.\end{aligned} \tag{4.7}$$

Moreover,

$$S^{\mathrm{F}}_{-}(t) = 1 + O(e^{-2t}), \quad t\rightarrow\infty\,, \tag{4.8}$$

and[4]

$$\lim_{t\rightarrow 0}\frac{\ln[S^{\mathrm{F}}_{-}(t)]}{\ln(1/t)} = \frac{1}{4}\,. \tag{4.9}$$

Finally, $S^{\mathrm{F}}_{-}(t)$ admits a spectral representation

$$S^{\mathrm{F}}_{-}(t) = 1 + \int_2^\infty d\rho^{\mathrm{F}}_{-}(m)\,K_0(mt)\,, \tag{4.10}$$

where the measure satisfies

$$\int_2^\infty \frac{d\rho^{\mathrm{F}}_{-}(m)}{m^\gamma}\begin{cases} < \infty, & \forall\,\gamma > \frac{1}{4},\\ = \infty, & \forall\,\gamma < \frac{1}{4}\,.\end{cases} \tag{4.11}$$

Theorem 4A.2: For any $t > 0$ the Schwinger function

$$S^{\mathrm{F}}_{+}(t) \equiv \|\phi^{\mathrm{F}}_{+}(\tfrac{1}{2}it,0)\Omega\|^2 \tag{4.12}$$

satisfies the bound

$$S^{F}_{+}(t) \leqslant (1/2\pi)K_0(t)S^{\mathrm{F}}_{-}(t)\,. \tag{4.13}$$

For t sufficiently large it admits a representation

$$S^{\mathrm{F}}_{+}(t) = \tfrac{1}{2}G(t)S^{\mathrm{F}}_{-}(t)\,, \tag{4.14}$$

where

$$\begin{aligned}G(t) \equiv \sum_{n=0}^{\infty}\frac{1}{(2\pi)^{2n+1}}\int d\theta_1\cdots d\theta_{2n+1}\\ \times\exp\left[-t\left(\sum_{j=1}^{2n+1}\cosh\theta_j\right)\right]\prod_{i=1}^{2n}\tanh[\tfrac{1}{2}(\theta_{i+1} - \theta_i)]\,.\end{aligned} \tag{4.15}$$

Moreover,

$$S^{\mathrm{F}}_{+}(t) = S_{\mathrm{KG}}(t) + O(e^{-3t}), \quad t\rightarrow\infty\,, \tag{4.16}$$

and[4]

$$\lim_{t\rightarrow 0}\frac{\ln[S^{\mathrm{F}}_{+}(t)]}{\ln(1/t)} = \frac{1}{4}\,. \tag{4.17}$$

Finally, (4.10) and (4.11) hold true with $-$ replaced by $+$.

Theorem 4A.3: The Wightman functions

$$W_{\pm}(\bar{F},F) \equiv \|\phi^{\mathrm{F}}_{\pm}(F)\Omega\|^2, \quad \tilde{F}\in C^\infty_0(\mathbb{R}^2), \tag{4.18}$$

extend to tempered distributions that admit tempered time-zero restrictions.

Proof of Theorem 4A.1: Since $\|\Lambda^{\mathrm{F}}(t)\|_2 < \infty$ for any $t > 0$ and, moreover, $\Lambda^{\mathrm{F}}(t,\theta_2,\theta_1) = -\Lambda^{\mathrm{F}}(t,\theta_1,\theta_2)$, (4.6) follows from Sec. 4 in Ref. 8. The work of McCoy *et al.*[4] implies (4.9), and the remaining statements then follow as in the proof of Theorem 2A.1.□

Proof of Theorem 4A.2: To prove (4.13), we observe that

$$S_{+}(t) = \|c^*(f_t)\,\phi^{\mathrm{F}}_{-}(\tfrac{1}{2}it,0)\Omega\|^2, \tag{4.19}$$

where

$$f_t(\theta) \equiv (4\pi)^{-1/2}\exp(-\tfrac{1}{2}t\cosh\theta). \tag{4.20}$$

Since $\|c^*(g)\| = \|g\|$ for $g\in L^2(\mathbb{R})$, (4.13) follows. Next, we note that if Λ is a Hilbert–Schmidt operator for which $\Lambda(\theta_2,\theta_1) = -\Lambda(\theta_1,\theta_2)$, then we may write, using the CAR,

$$\begin{aligned}&\|c^*(g)\exp(\tfrac{1}{2}\Lambda c^*c^*)\Omega\|^2\\ &= (g,g)\|\exp\cdots\Omega\|^2 - \|\exp(\cdots)c^*(\Lambda g)\Omega\|^2\\ &= \cdots = \sum_{n=0}^{N}(-)^n\|\Lambda^n g\|^2\|\exp\cdots\Omega\|^2\\ &+ (-)^{N+1}\|c^*(\Lambda^{N+1}g)\exp\cdots\Omega\|^2\\ &= \sum_{n=0}^{\infty}(-)^n\|\Lambda^n g\|^2\|\exp\cdots\Omega\|^2,\end{aligned} \tag{4.21}$$

where the last step holds true if in addition $\|\Lambda^N g\|\rightarrow 0$ for $N\rightarrow\infty$. Consequently, (4.14) follows from the fact that $\|\Lambda^{\mathrm{F}}(t)\|_2\rightarrow 0$ for $t\rightarrow\infty$. Finally, (4.17) follows from Ref. 4, and the remaining assertions then follow as before.□

Proof of Theorem 4A.3: The proof is similar to that of the analogous claims in Theorems 2A.2 and 2D.2 and will therefore be omitted.□

B. Fields on $\mathscr{F}_s(\mathscr{H}_+)$

We define

$$\phi^{\mathrm{B}}(x)\Omega \equiv \exp(\tfrac{1}{2}K^{\mathrm{B},x}c^*c^*)\Omega, \tag{4.22}$$

where $K^{\mathrm{B},x}$ is given by the rhs of (4.2) with F→B, and

$$K^{\mathrm{B}}(\theta) \equiv (1/2\pi)\operatorname{sech}\tfrac{1}{2}\,\theta. \tag{4.23}$$

We also define

$$\begin{aligned}\Psi^{\mathrm{B}}(x)\Omega \equiv (4\pi)^{-1/2}\int d\,\theta\, c^*(\theta)\begin{pmatrix} ie^{\theta/2}\\ -ie^{-\theta/2}\end{pmatrix}\\ \times e^{ix\cdot p(\theta)}\,\phi^{\mathrm{B}}(x)\Omega.\end{aligned} \tag{4.24}$$

Theorem 4B.1: For any $t>0$ the Schwinger function

$$S^{\mathrm{B}}(t)\equiv\|\phi^{\mathrm{B}}(\tfrac{1}{2}\mathrm{i}t,0)\Omega\|^2 \tag{4.25}$$

is finite-valued and satsifies

$$S^{\mathrm{B}}(t)=\{\det[1-A^{\mathrm{B}}(t)^*A^{\mathrm{B}}(t)]\}^{-1/2}, \tag{4.26}$$

where $A^{\mathrm{B}}(t)$ is the integral operator on $L^2(\mathbb{R})$ with kernel

$$A^{\mathrm{B}}(t,\theta_1,\theta_2)=(1/2\pi)\exp[-\tfrac{1}{2}t(\cosh\theta_1+\cosh\theta_2)]\times\operatorname{sech}[\tfrac{1}{2}(\theta_1-\theta_2)]. \tag{4.27}$$

Moreover,

$$S^{\mathrm{B}}(t)=1+O(e^{-2t}),\quad t\rightarrow\infty, \tag{4.28}$$

and

$$\lim_{t\to 0}\frac{\ln[S^{\mathrm{B}}(t)]}{\ln(1/t)}=\frac{1}{4}. \tag{4.29}$$

Finally, $S^{\mathrm{B}}(t)$ admits a spectral representation

$$S_{\mathrm{B}}(t)=1+\int_2^\infty d\rho^{\mathrm{B}}(m)\,K_0(mt), \tag{4.30}$$

where the measure satisfies

$$\int_2^\infty\frac{d\rho^{\mathrm{B}}(m)}{m^\gamma}\begin{cases}<\infty, & \forall\ \gamma>\frac{1}{4},\\ =\infty, & \forall\ \gamma<\frac{1}{4}.\end{cases} \tag{4.31}$$

Theorem 4B.2: For any $t>0$ the Schwinger function

$$\mathscr{S}^{\mathrm{B}}(t)\equiv\|\Psi^{\mathrm{B}}(\tfrac{1}{2}it,0)\Omega\|^2 \tag{4.32}$$

is finite-valued and satisfies

$$\mathscr{S}^{\mathrm{B}}_{\epsilon,\epsilon'}(t)=\tfrac{1}{2}G_{\epsilon\epsilon'}(t)S^{\mathrm{B}}(t),\quad \epsilon,\epsilon'=+,-, \tag{4.33}$$

where

$$G_\delta(t)\equiv\delta\sum_{n=0}^\infty\frac{1}{(2\pi)^{2n+1}}\int d\theta_1\cdots d\theta_{2n+1}e^{(\theta_1+\delta\theta_{2n+1})/2}\times\exp\Big(-t\sum_{j=1}^{2n+1}\cosh\theta_j\Big)\prod_{i=1}^{2n}\operatorname{sech}[\tfrac{1}{2}(\theta_{i+1}-\theta_i)]. \tag{4.34}$$

Moreover, $\mathscr{S}^{\mathrm{B}}$ satisfies

$$\mathscr{S}^{\mathrm{B}}(t)=S_{\mathrm{D}}(t)+O(e^{-3t}),\quad t\rightarrow\infty; \tag{4.35}$$

its diagonal elements satisfy

$$\lim_{t\to 0}\frac{\ln[\mathscr{S}^{\mathrm{B}}_{\epsilon,\epsilon}(t)]}{\ln(1/t)}=\frac{5}{4}, \tag{4.36}$$

and its off-diagonal elements satisfy

$$-\int_0^\infty dt\,t^\alpha\mathscr{S}^{\mathrm{B}}_{\epsilon,-\epsilon}(t)\begin{cases}<\infty, & \forall\ \alpha>\frac{1}{4},\\ =\infty, & \forall\ \alpha<\frac{1}{4}.\end{cases} \tag{4.37}$$

Finally, $\mathscr{S}^{\mathrm{B}}$ admits a spectral representation

$$\mathscr{S}^{\mathrm{B}}(t)=\frac{1}{2\pi}\begin{pmatrix}K_1(t) & -K_0(t)\\ -K_0(t) & K_1(t)\end{pmatrix}+\int_3^\infty m\,d\rho_1(m)\begin{pmatrix}K_1(mt) & 0\\ 0 & K_1(mt)\end{pmatrix}-d\rho_2(m)\times\begin{pmatrix}0 & K_0(mt)\\ K_0(mt) & 0\end{pmatrix}. \tag{4.38}$$

Here, the measures are positive and satisfy the inequalities (2.49) and (2.50) with $m(1)=1$; furthermore,

$$\int_3^\infty d\rho_j(m)\,m^{-\gamma+2-j}\begin{cases}<\infty, & \forall\,\gamma>\frac{5}{4}\\ =\infty, & \forall\gamma<\frac{5}{4}\end{cases}\quad j=1,2. \tag{4.39}$$

Theorem 4B.3: The Wightman functions

$$W^{\mathrm{B}}(\overline{F},F)\equiv\|\phi^{\mathrm{B}}(F)\Omega\|^2,\quad \widetilde{F}\in C_0^\infty(\mathbb{R}^2), \tag{4.40}$$

and

$$\mathscr{W}^{\mathrm{B}}(\overline{F},F)\equiv\|\Psi^{\mathrm{B}}(F)\Omega\|^2,\quad \widetilde{F}\in C_0^\infty(\mathbb{R}^2)^2, \tag{4.41}$$

extend to tempered distributions. W^{B} admits a tempered time-zero restriction, while $\mathscr{W}^{\mathrm{B}}$ does not admit a time-zero restriction.

Proof of Theorem 4B.1: Since $\|A^{\mathrm{B}}(t)\|_2<\infty$, $\|A^{\mathrm{B}}(t)\|<1$, and $A^{\mathrm{B}}(t,\theta_2,\theta_1)=A^{\mathrm{B}}(t,\theta_1,\theta_2)$ for any $t>0$, (4.26) follows from Sec. 4 in Ref. 8.

But in view of (2.62) and (2.63) this implies

$$S^{\mathrm{B}}(t)=S^{\mathrm{B}}_{1/2}(t)^{1/2}. \tag{4.42}$$

Hence, (4.29) is a consequence of (2.65), and the other assertions can then be proven as before. □

Proof of Theorem 4B.2: Elsewhere (Ref. 8, Lemma 4.3) we have shown that vectors of the form $\exp(\frac{1}{2}A\mathrm{c}^*\mathrm{c}^*)\Omega$, where $\|A\|_2<\infty$, $\|A\|<1$, and $A=A^T$, are in the domain of all powers of the number operator $N\equiv d\Gamma(1)$. Thus one has

$$\|c^*(f)\exp(\tfrac{1}{2}Ac^*c^*)\Omega\|\leqslant\|f\|\ \|(N+1)\exp(\tfrac{1}{2}Ac^*c^*)\Omega\|\leqslant C\|f\|, \tag{4.43}$$

where $C<\infty$ only depends on A. It easily follows from this that $\mathscr{S}^{\mathrm{B}}(t)$ is finite-valued and, using the boson analog of (4.21), that (4.33) holds true. The bound (4.35) follows as before, while the assertions concerning the spectral representation and the measures occurring in it can be proven as in Theorem 2B.1, using (4.36) and (4.37) to obtain (4.39). Therefore, it remains to prove (4.36) and (4.37).

To prove (4.36), we observe that one may write

$$\begin{aligned}G_+(t)={}&-\sum_{n=1}^\infty\frac{1}{2n+1}\frac{1}{(2\pi)^{2n+1}}\partial_t\int d\theta_1\cdots d\theta_{2n+1}\\&\times\exp\Big[-t\sum_{j=1}^{2n+1}\cosh\theta_j\Big]\\&\times\prod_{i=1}^{2n}\operatorname{sech}[\tfrac{1}{2}(\theta_{i+1}-\theta_i)]\operatorname{sech}[\tfrac{1}{2}(\theta_1-\theta_{2n+1})]\\&+\frac{K_1(t)}{\pi},\\={}&\frac{2}{t}\sum_{n=1}^\infty\frac{1}{2n+1}\frac{1}{(2\pi)^{2n+1}}\int dy_1\cdots dy_{2n}\,(\widetilde{M}_{2n+1}(\mathbf{y})t)\\&\times K_1(\widetilde{M}_{2n+1}(\mathbf{y})t)\\&\times\prod_{i=1}^{2n}\operatorname{sech}\tfrac{1}{2}y_i\cdot\operatorname{sech}\Big(\tfrac{1}{2}\sum_{j=1}^{2n}y_j\Big)+\frac{K_1(t)}{\pi}.\end{aligned} \tag{4.44}$$

Now since the function $aK_1(a)$ is bounded on $(0,\infty)$ and has limit 1 for $a\rightarrow 0$, we have, by dominated convergence and a calculation analogous to that leading to (2.34),

$$\begin{aligned}\lim_{t\to 0}tG_+(t)={}&\frac{2}{\pi}\sum_{n=0}^\infty\frac{1}{2n+1}\int_0^\infty dx(\operatorname{sech}\pi x)^{2n+1}\\&-\frac{2}{\pi}\int_0^\infty dx\operatorname{sech}\pi x+\frac{1}{\pi}\\={}&\frac{1}{\pi^2}\int_0^\infty dx\ln[(\cosh x+1)(\cosh x-1)^{-1}]\end{aligned}$$

$$= \tfrac{1}{2}. \tag{4.45}$$

Combining this with (4.29), we obtain (4.36).

To prove (4.37), we write

$$\begin{aligned} -G_-(t) = {} & -\sum_{n=1}^{\infty} \frac{1}{2n+1} \frac{1}{(2\pi)^{2n+1}} \partial_t \int d\theta_1 \cdots d\theta_{2n+1} \\ & \times \exp\Big[-t \sum_{j=1}^{2n+1} \cosh\theta_j \Big] \prod_{i=1}^{2n} \operatorname{sech}[\tfrac{1}{2}(\theta_{i+1} - \theta_i)] \\ & \times \operatorname{sech}[\tfrac{1}{2}(\theta_1 + \theta_{2n+1})] + \frac{K_0(t)}{\pi} \\ = {} & \sum_{n=1}^{\infty} \frac{1}{2n+1} \frac{1}{(2\pi)^{2n+1}} \int dy_0 dy_1 \cdots dy_{2n} \, \widetilde{M}_{2n+1}(\mathbf{y}) \cosh y_0 \\ & \times \exp[-t\widetilde{M}_{2n+1}(\mathbf{y}) \cosh y_0] \prod_{i=1}^{2n} \operatorname{sech}\tfrac{1}{2} y_i \\ & \times \operatorname{sech}[y_0 + f_n(\mathbf{y})] + \frac{K_0(t)}{\pi}, \end{aligned} \tag{4.46}$$

where

$$f_n(\mathbf{y}) \equiv \ln\Big\{ \widetilde{M}_{2n+1}(\mathbf{y}) \Big/ \Big[1 + \sum_{i=1}^{2n} \exp\Big(\sum_{k=1}^{i} y_k \Big) \Big] \Big\} + \tfrac{1}{2} \sum_{i=1}^{2n} y_i . \tag{4.47}$$

Hence, for any $\epsilon > 0$ we have, using (2.26),

$$\begin{aligned} \int_0^{\infty} dt\, t^{\epsilon}(-G_-(t)) < {} & \frac{\Gamma(1+\epsilon)}{2\pi} \int dy_0 (\operatorname{sech} y_0)^{\epsilon} \sum_{n=1}^{\infty} \frac{1}{2n+1} \\ & \times \int dy_1 \cdots dy_{2n} \, \widetilde{M}_{2n+1}(\mathbf{y})^{-\epsilon} \prod_{i=1}^{2n} (2\pi \cosh\tfrac{1}{2} y_i)^{-1} + C_{\epsilon} \\ & < C'_{\epsilon} \sum_{n=1}^{\infty} (2n+1)^{-1-\epsilon} + C_{\epsilon} < \infty . \end{aligned} \tag{4.48}$$

On the other hand, by virtue of (4.46),

$$\int_0^{\infty} dt\,(-G_-(t)) = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{2n+1} = \infty . \tag{4.49}$$

Combining this with (4.48) and (4.29), (4.37) follows.□

Proof of Theorem 4B.3: The statements follow from Theorems 4B.1 and 4B.2 as in preceding proofs.□

Remark 4B.4: It is of interest to point out that nonexistence of a time-zero restriction for the diagonal elements already follows from the fact that the three-particle component of $\Psi^{B}_{\epsilon}(0,f)$ [where $f \in \mathcal{S}(\mathbb{R})$] is not square integrable, which can be seen as in the proof of Theorem 2B.2. Note, however, that for the off-diagonal elements all $(2n+1)$-particle contributions are finite, since they are bounded above by a multiple of the integral

$$I_n \equiv \int d\theta_1 \cdots d\theta_{2n+1} \Big| \hat{f}\Big(\sum_{j=1}^{2n+1} \sinh\theta_j \Big) \Big|^2 . \tag{4.50}$$

[To see that I_n is finite, one can either change variables $\theta_j \rightarrow p_j \equiv \sinh\theta_j$ and use Young's and Hölder's inequality, or change to the center of mass variables (2.22), from which one infers directly that

$$\begin{aligned} I_n & = \int dy_0 d\mathbf{y} \, |\hat{f}(\widetilde{M}_{2n+1}(\mathbf{y}) \sinh y_0)|^2 \\ & \leqslant \int dp \, |\hat{f}(p)|^2 \int d\mathbf{y} \widetilde{M}_{2n+1}(\mathbf{y})^{-1} \\ & < \infty , \end{aligned} \tag{4.51}$$

where the last step follows by using (2.24) and, e.g., the arithmetic–geometric mean inequality.] Thus, in this case information on the sum of the terms is essential to conclude no time-zero restriction exists.

Remark 4B.5: If one combines the results of Sato *et al.*[9] (who introduced the bosonic Ising model considered in this section) with Ref. 4, one can get more detailed information on the short-distance behavior of S^{B} and $G_{\pm}$, but this only follows after the introduction of considerable machinery. It would be of interest to reobtain such results (and their analogs in the fermion case) in a simpler and more direct way by using only the underlying operators $A^{B}(t)$ and $A^{F}(t)$.

[1] S. N. M. Ruijsenaars, Ann. Phys. **132**, 328–82 (1981).

[2] S. N. M. Ruijsenaars, "Scattering theory for the Federbush, massless Thirring and continuum Ising models," Texas A&M University preprint, 1980.

[3] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964).

[4] B. M. McCoy, C. A. Tracy, and T. T. Wu, J. Math. Phys. **18**, 1058–92 (1977).

[5] S. N. M. Ruijsenaars, Ann. Phys. **126**, 399–449 (1980).

[6] J. L. Challifour, J. Math. Phys. **9**, 1137–45 (1968).

[7] S. N. M. Ruijsenaars, J. Math. Phys. **18**, 517–26 (1977).

[8] S. N. M. Ruijsenaars, Ann. Phys. **116**, 105–34 (1978).

[9] M. Sato, T. Miwa, and M. Jimbo, Publ. RIMS **15**, 871–972 (1979).

Particle states and scattering theory in abelian gauge model with spontaneously broken symmetry

Garner Bishop[a)]
Department of Physics, University of Rhode Island, Kingston, Rhode Island 02881
Kurt Haller
Department of Physics, University of Connecticut, Storrs, Connecticut 06268

(Received 17 June 1982; accepted for publication 11 August 1982)

This is an exposition of the electrodynamics of the ϕ^4 theory with a spontaneously broken U(1) symmetry. The Lagrangian in a manifestly covariant gauge is used to construct the Hamiltonian and canonical commutation rules. Part of the Hamiltonian is a "free-field" Hamiltonian which describes noninteracting massive spin-1 Proca bosons, Higgs bosons, and ghosts. The remainder of the Hamiltonian is a perturbation for which an interaction picture is developed and Feynman rules are systematically derived by the Dyson–Wick procedure. The choice of the "free-field" Hamiltonian is based on earlier work, in which the particle spectrum was identified and verified explicitly by implementing Lorentz boosts on particle states. It is shown that application of the subsidiary condition and inclusion of interactions in evaluating positive frequency parts of gauge fixing fields is not only essential for consistency, but in this case also of practical importance in the proper identification of the Proca bosons.

PACS numbers: 11.10.Ef, 11.15.Bt, 11.15.Ex

I. INTRODUCTION

This paper reports extensions of earlier work[1,2] on the electrodynamics of the ϕ^4 theory with a spontaneously broken U(1) symmetry—the model originally proposed and discussed by Higgs, Kibble, Guralnik, Englert, and Brout.[3] In this model the vacuum state is degenerate, and the ϕ field develops a nonvanishing vacuum expectation value, which, in turn, transforms the photon into a massive vector boson. In our earlier work we formulated this model as a canonical field theory and obtained its particle spectrum as eigenstates of a Hamiltonian in which field fluctuations were expanded about the tree approximation to the vacuum expectation value of ϕ. We found that these eigenstates consist of a three-component spin-1 massive vector boson that obeys the Proca equations of motion and a Higgs scalar particle and that the spectrum of states also includes the massless ghosts required by the fact that the theory is a form of electrodynamics in a manifestly covariant gauge. We verified our identification of the components of the massive vector boson by explicitly carrying out Lorentz boosts on the helicity components and showing that they transformed correctly into each other. In this present work we continue the development of our canonical formulation of this model by choosing an "interaction free" part of the Hamiltonian which defines the interaction picture and by then deriving the Feynman rules from the Dyson–Wick expansion of the S matrix. We demonstrate the consistency of our S-matrix rules by verifying that S-matrix elements to states that include "pure gauge" ghosts vanish. We examine the mechanism which moderates and controls the highly divergent Proca boson propagator sufficiently to preserve renormalizability. In our formulation we are able to identify the particles that are eigenstates of the Hamiltonian that governs the time dependence of the interaction picture. We can also draw correspondence between these particles and associated fields, so that in the interaction picture some fields definitely have massive and others massless excitations. For this reason we can subtract on the mass shells of participating particles in the course of renormalization. We also include explicit calculations of S-matrix elements that connect incident and final states that are physically observable in this model; for example, we treat the elastic scattering of Higgs particle + massive vector boson→Higgs particle + massive vector boson. We will organize our presentation in the following way: In Sec. II of this paper we will systematically derive the Feynman rules from the canonical formulation following the general pattern set by Dyson and Wick. We will also illustrate the application of these rules to cases in which the correct identification of the zero-helicity component of the massive vector boson is essential. In Sec. III we will illustrate the dynamic detachment of pure gauge ghosts and the relation of this effect to gauge invariance. In Sec. IV we will discuss some technical questions relating to the renormalizability of this theory. In Sec. V we will make some general comments about the abelian Higgs model.

We will continue this introductory section by summarizing some of the most important results of Refs. 1 and 2: The Lagrangian that governs the behavior of this model is

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - G(x)\partial_\mu A_\mu + \tfrac{1}{2}G^2(x) - \mathrm{D}_\mu^\dagger\phi^\dagger D_\mu\phi - m^2\phi^\dagger\phi - h(\phi^\dagger\phi)^2. \tag{1.1}$$

ϕ is expressed in the form

$$\phi = (1/\sqrt{2})(\lambda + \psi + i\chi), \tag{1.2}$$

where $\lambda/\sqrt{2}$ is the tree approximation to the vacuum expectation value of ϕ and implies a vacuum state that breaks the U(1) symmetry. The resulting form of the Lagrangian is

---

[a)] Partially based on a thesis to be submitted by G. B. to the University of Connecticut in partial fulfillment of the requirements for the Ph.D degree.

0022-2488/83/040932-17$02.50

$$\begin{aligned}\mathscr{L} = & -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - \tfrac{1}{2}\partial_\mu\psi\partial_\mu\psi - \tfrac{1}{2}\partial_\mu\chi\partial_\mu\chi \\ & - \tfrac{1}{2}(2h\lambda^2)\psi^2 - \tfrac{1}{2}M^2A_\mu A_\mu + MA_\mu\partial_\mu\chi \\ & - G\partial_\mu A_\mu + \tfrac{1}{2}G^2 + eA_\mu\psi\overleftrightarrow{\partial}_\mu\chi \\ & - \tfrac{1}{2}e^2A_\mu A_\mu(\psi^2 + \chi^2 + 2\lambda\psi) \\ & - h(\lambda\psi^3 + \tfrac{1}{4}\psi^4 + \tfrac{1}{4}\chi^4 + \tfrac{1}{2}\psi^2\chi^2 + \lambda\psi\chi^2),\end{aligned} \qquad (1.3)$$

where $M = e\lambda$ and $\lambda^2 = -m^2/h$. As is always true in electrodynamics in a manifestly covariant linear gauge, the gauge-fixing field $G(x)$ obeys the free-field equation

$$\Box G = 0 \qquad (1.4)$$

even in the presence of interactions. It can therefore be represented as the sum of invariant positive and negative frequency parts with

$$G^{(+)}(x) = \frac{i}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{\sqrt{2}}e^{i(\mathbf{k}\cdot\mathbf{x}|-\mathbf{k}|x_0)}\Omega^{(+)}(\mathbf{k}) \qquad (1.5)$$

and with $G^{(-)}(x)$ the adjoint of $G^{(+)}(x)$.

In this work it is necessary to take particular care to implement the subsidiary condition

$$\Omega^{(+)}(\mathbf{k})|\nu\rangle = 0 \qquad (1.6)$$

to select admissible states that span a physical subspace of an appropriately constructed indefinite metric space. Requiring this subsidiary condition guarantees that both $\partial_\mu A_\mu = 0$ and $\nabla\cdot\mathbf{E} - \rho = 0$ hold for matrix elements in the physical subspace; and it is the implementation of Gauss' law that provides the zero-helicity component of the vector boson with its proper mass. As in earlier work on quantum electrodynamics, in which the U(1) symmetry is not spontaneously broken, the technique for implementing the subsidiary condition is based on constructing a pseudo-unitary similarity transformation (i.e., one that is unitary in the indefinite metric space so that $U^*U = UU^* = 1$). This similarity transformation establishes one-to-one correspondence between $\Omega^{(+)}(\mathbf{k})$ and a c-number multiple of a ghost-annihilation operator. $\Omega^{(+)}(\mathbf{k})$ has the form

$$\Omega^{(+)}(\mathbf{k}) = \kappa(k)[B_Q(\mathbf{k}) + X(k)B_R{}^\star(-\mathbf{k})] + \rho(\mathbf{k})/\sqrt{2}|\mathbf{k}|, \qquad (1.7a)$$

where $\kappa(k) = (k_0 + |\mathbf{k}|)/2k_0$, $X(k) = (k_0 - |\mathbf{k}|)/(k_0 + |\mathbf{k}|)$, and $\rho(\mathbf{k})$ is the Fourier transform of the charge density. Its pseudo-unitary transform is

$$\widetilde{\Omega}^{(+)}(\mathbf{k}) = \kappa(k)B_Q(\mathbf{k}), \qquad (1.7b)$$

where ~ designates the pseudo-unitarily transformed quantity $\tilde{\xi} = U\xi U^{-1}$ with $U^{-1} = U^*$. $B_Q(\mathbf{k})$ refers to one of the ghost annihilation operators in this theory. $B_R(\mathbf{k})$ refers to the annihilation operator for the other ghost. The corresponding creation operators are $B_Q{}^\star(\mathbf{k})$ and $B_R{}^\star(\mathbf{k})$. The kinematic relations that govern these operators are described in Ref. 2. The transformed operator $\widetilde{\Omega}^{(+)}(\mathbf{k})$ allows us to restate the subsidiary condition in a simple and very useful form—namely, $\widetilde{\Omega}^{(+)}(\mathbf{k})|n\rangle = 0$, which is equivalent to $B_Q(\mathbf{k})|n\rangle = 0$. This condition simply requires that the states $|n\rangle$ contain no $B_R{}^\star(\mathbf{k})$ ghosts, since, for the state $|n'\rangle = B_R{}^\star(\mathbf{k})|n\rangle$, $B_Q(\mathbf{k})|n'\rangle \neq 0$. If $|n\rangle$ is devoid of ghosts or contains $B_Q{}^\star(\mathbf{k})$ ghosts alone or in combination with any particles other than $B_R{}^\star(\mathbf{k})$, then $B_Q(\mathbf{k})|n\rangle = 0$ and these states too satisfy the subsidiary condition in the new transformed representation. In order to transpose the entire theory into this new representation, the transformation $\xi \rightarrow \tilde{\xi}$, where $\tilde{\xi} = U\xi U^{-1}$ is extended to all other operators in this theory; the transformed Hamiltonian $\widetilde{H}$, given by $\widetilde{H} = UHU^{-1}$, will be expressed as $\widetilde{\mathrm{H}} = \widetilde{H}_A + \widetilde{H}_B$, where $\widetilde{H}_A$ has the form

$$\begin{aligned}\widetilde{H}_A = \int\frac{d\mathbf{k}}{2k_0}\Big\{ & k_0\Big[\sum_{i=1}^{2}A_i{}^\star(\mathbf{k})A_i(\mathbf{k}) + \alpha^\star(\mathbf{k})\alpha(\mathbf{k})\Big] \\ & + |\mathbf{k}|[B_R{}^\star(\mathbf{k})B_Q(\mathbf{k}) + B_Q{}^\star(\mathbf{k})B_R(\mathbf{k})] \\ & + \frac{(k_0 - 3|\mathbf{k}|)(k_0 + |\mathbf{k}|)}{8k_0}[B_Q(\mathbf{k})B_Q(-\mathbf{k}) \\ & + B_Q{}^\star(\mathbf{k})B_Q{}^\star(-\mathbf{k})] - \frac{(k_0 + |\mathbf{k}|)^2}{4k_0}B_Q{}^\star(\mathbf{k})B_Q(\mathbf{k})\Big\} \\ & + \int\frac{d\mathbf{p}}{2p_0}[p_0\beta^\star(\mathbf{p})\beta(\mathbf{p})]\end{aligned} \qquad (1.8)$$

and $\widetilde{H}_B$ has a complicated nonlocal structure given in Ref. 2. $\widetilde{H}_A$ is the part of $\widetilde{H}$ that survives in the limit $e \rightarrow 0$, but $e\lambda = M$ is kept fixed, i.e., $\widetilde{H}_A$ is the so-called "antidipole" limit of $\widetilde{H}$. $\widetilde{H}_B$ is the part of $\widetilde{H}$ that vanishes in the antidipole limit. $\widetilde{H}_B$ consists of interactions for transverse vector bosons and Higgs scalars, and of parts of the nonlocal Coulomb interaction. There are no ghost-mediated interactions in $\widetilde{H}_B$; the ghost-mediated interactions that are normally part of the manifestly covariant formulation of this model have been eliminated by the similarity transformation and replaced with the Coulomb interaction. There are parts of $\widetilde{H}_B$ that couple the $B_Q(\mathbf{k})$ $[B_Q{}^\star(\mathbf{k})]$ ghost annihilation [creation] operators to other particles, but they play no dynamical role, because $B_Q(\mathbf{k})$ and $B_Q{}^\star(\mathbf{k})$ commute. The only role that these ghost interaction terms have is to preserve the equations of motion and the kinematic relations of the manifestly covariant Lorentz gauge. In all of these respects this model is very similar to "regular" quantum electrodynamics, i.e., to the case in which the vacuum state is nondegenerate and carries the U(1) symmetry of the Lagrangian. But the fact that the scalar field has the vacuum expectation value $\lambda/\sqrt{2}$ has some important consequences that distinguish this model from quantum electrodynamics with an unbroken U(1) symmetry.

In quantum electrodynamics without a degenerate vacuum the Coulomb interaction, $\int\rho(\mathbf{x})\rho(\mathbf{y})\,d\mathbf{x}d\mathbf{y} \times(8\pi|\mathbf{x}-\mathbf{y}|)^{-1}$, is proportional to e^2 and quadratic in particle creation and/or annihilation operators. In this, the "Higgs" model, one contribution to the charge density operator, originates from the cross term in which the $\lambda/\sqrt{2}$ from the real part of ϕ combines with Π_χ, the momentum conjugate to χ. This leads to a term in the Coulomb interaction that is quadratic in particle creation and annihilation operators and proportional to $(e\lambda)^2$; this term survives the antidipole limit. Since $e\lambda = M$, the mass of the transverse component of the vector bosons in this model, we are led to suspect, correctly as it turns out, that the excitation modes of the transformed χ and Π_χ constitute the zero-helicity component of the spin-1 particle. The χ- and Π_χ-dependent parts of the transformed Hamiltonian that survive the antidipole limit, and become part of $\widetilde{H}_A$, are

$$\widetilde{H}_{A(\chi)} = \tfrac{1}{2}\int d\mathbf{x}\ [\nabla\chi(\mathbf{x})\cdot\nabla\chi(\mathbf{x}) + \Pi_\chi{}^2(\mathbf{x})] + \frac{M^2}{8\pi}\int d\mathbf{x}\,d\mathbf{y}\,\frac{\Pi_\chi(\mathbf{x})\Pi_\chi(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}. \tag{1.9}$$

The last of these terms has originated from the Coulomb interaction. In order to arrive at a form of $\widetilde{H}_A$ that is diagonal in the number representation for each particle species, the χ field may not be represented in terms of elementary scalar spin-0 excitations. If we made such an expansion, for example, by setting

$$\chi(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}}\int \frac{d\mathbf{k}}{2k_0}\,[c(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} + c^\star(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}}] \tag{1.10}$$

and

$$\Pi_\chi(\mathbf{x}) = \frac{-i}{(2\pi)^{3/2}}\int \frac{d\mathbf{k}}{2k_0}\,[c(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} - c^\star(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}}], \tag{1.11}$$

then, although $\frac{1}{2}\int d\mathbf{x}\ [\nabla\chi(\mathbf{x})\cdot\nabla\chi(\mathbf{x}) + \Pi_\chi{}^2(\mathbf{x})]$ would be diagonal in the particle number $c^\star(\mathbf{k})c(\mathbf{k})$, the nonlocal $M^2\int d\mathbf{x}d\mathbf{y}\ \Pi_\chi(\mathbf{x})\Pi_\chi(\mathbf{y})(8\pi|\mathbf{x}-\mathbf{y}|)^{-1}$ would destroy that property. In order to arrive at a form in which all of $\widetilde{H}_{A(x)}$ is diagonal in the number representation, including the nonlocal $M^2\int d\mathbf{x}d\mathbf{y}\ \Pi_\chi(\mathbf{x})\Pi_\chi(\mathbf{y})(8\pi|\mathbf{x}-\mathbf{y}|)^{-1}$ term, we must make the expansion

$$\chi(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}}\int \frac{d\mathbf{k}}{2k_0}\left[\frac{k_0}{|\mathbf{k}|}\right][\alpha(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} + \alpha^\star(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}}] \tag{1.12}$$

and

$$\Pi_\chi(\mathbf{x}) = \frac{-i}{(2\pi)^{3/2}}\int \frac{d\mathbf{k}}{2}\left[\frac{|\mathbf{k}|}{k_0}\right][\alpha(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}} - \alpha^\star(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}}], \tag{1.13}$$

and we then arrive at

$$\widetilde{H}_{A(\chi)} = \int \frac{d\mathbf{k}}{2k_0}\,[k_0\alpha^\star(\mathbf{k})\alpha(\mathbf{k})]. \tag{1.14}$$

In these features we see that the degenerate vacuum, which allows the ϕ field to have the vacuum expectation value $\lambda/\sqrt{2}$, forces a number of significant changes in the spectrum of particle states that characterize this model: The transverse excitations of A_μ have now become massive; the excitations of the χ field, which would constitute the Goldstone mode in the model with degenerate vacuum if the electromagnetic interactions were not included, are here involved in the Coulomb interaction. In this way these excitations develop a mass identical to the mass of the transverse boson modes. Moreover, because of the extra factor $k_0/|\mathbf{k}|$ in Eq. (1.12) and $|\mathbf{k}|/k_0$ in Eq. (1.13), the α, $\alpha^\star$ excitations refer to the zero-helicity mode of a spin-1 system and no longer transform like scalar (spin-0) particle states. Since the ghost modes, represented by B_Q and $B_Q{}^\star$ remain massless (as required by the free-field equation $\Box G = 0$), $\bar{A}_\mu$ has a very different significance when time translated by $\widetilde{H}_A$ from what would be the case if ϕ had a vanishing vacuum expectation value. Under a Lorentz transformation the four components of $\bar{A}_\mu$ no longer constitute the four components of a vector field which transform among themselves. The two transverse modes and the α and $\alpha^\star$ modes all refer to particles of mass $M = e\lambda$ and transform like the three helicity components of a spin-1 Proca system. They completely fail to mix with the massless electrodynamic ghosts in this model.

In Sec. III of Ref. 2 we discussed the dynamics of this model in the antidipole limit. We gave the explicit time dependence of a number of fields, and the others, not explicitly given, can be obtained by a simple straightforward calculation. This time dependence is always an explicit c-number, but not always the plane wave $\exp(ik_\mu x_\mu)$ usual in the "interaction-free" limit of theories with nondegenerate vacuum states that carry the symmetries of the Lagrangian. The reason for this complication in the Higgs model stems from the fact that, although $\widetilde{H}_A$ is diagonal in the number operator for "observable" particles, it inevitably fails to be diagonal in the number operator for electrodynamic ghosts. Nevertheless, in this "antidipole" limit it is not only possible but also very convenient to define a Proca field, $\bar{Z}_\mu(\mathbf{x})$, and use it in the formulation of this model.[4] $\bar{Z}_\mu(\mathbf{x})$ will be seen to be an ordinary and unexceptional Proca field. It obeys the equations

$$(\Box - M^2)\bar{Z}_\mu = 0 \tag{1.15}$$

and

$$\partial_\mu \bar{Z}_\mu = 0, \tag{1.16}$$

as well as the commutation rules

$$[\bar{Z}_\mu(x),\bar{Z}_\nu(y)] = i[\delta_{\mu,\nu} - \partial_\mu\partial_\nu/M^2]\Delta(x-y) \tag{1.17}$$

with

$$\Delta(x-y) = -i\int \frac{d\mathbf{k}}{2k_0}\,[e^{ik_\mu(x-y)_\mu} - e^{-ik_\mu(x-y)_\mu}]. \tag{1.18}$$

In the remaining sections of this paper we will discuss the iteration of the S matrix in this theory in an interaction picture in which the "noninteracting" time dependence is dictated by $\widetilde{H}_A$ instead of the completely "free-field" Hamiltonian $\widetilde{H}_0$, in which $e \to 0$ without any constraint that keeps e from vanishing; in $\widetilde{H}_0$, therefore, $e\lambda = 0$ as well. There are compelling reasons for choosing $\widetilde{H}_A$ as the generator of time displacements in the interaction picture. The Hilbert space generated by $\widetilde{H}_0$ consists of Higgs scalars of mass $(2h\lambda^2)$, massless noninteracting transverse and ghost photons, and χ particles which in the completely interaction free limit lack the mass term which, in the exact theory, stems from interactions with the electromagnetic field. Whereas in conventional quantum electrodynamics, with the nondegenerate vacuum, the mass spectrum of the exact theory must be consistent with the massless photons, because of the possibility that initial and final scattering states contain such massless photons in the absence of, or far from charges, this is not the case in the abelian Higgs model. In the abelian Higgs model we cannot expect to isolate collective excitations with any definite charge, in particular not with zero charge, and the concept of photons "far from" or "in the absence of" charges does not apply. Because of the nonvanishing vacuum expectation value of ϕ, there are terms in the exact Hamiltonian that are proportional to $e\lambda$, which stem from the interactions between photons and charges, but which no longer involve any field fluctuations other than photons. These terms, which contribute to $\widetilde{H}_A$ and are not represented in $\widetilde{H}_0$, originate from the substitution of $\lambda/\sqrt{2}$ for part of ϕ, and are bilinear in creation and/or annihilation operators of elementary field excitations. They are manifestly involved in changing the masses of these excitations and, if these terms were eliminated from consideration when calculating these masses, and in calculating the time dependence of the interaction picture, they would reappear in trivial ways to modify the spectrum of incident and scattered particles. They would need to be summed, and, after summing, they would force the changes in the particle spectrum that are more easily incorporated into $\widetilde{H}_A$ directly in the quantization procedure. It is not wholly arbitrary how the Hamiltonian is divided into two parts so that one (the "interaction-free" part) defines the spectrum of participating particles and the time evolution in the interaction picture, while the remaining part is responsible for the interactions whose effects the S matrix reflects. In order to make this separation consistently, the following are necessary: The particle spectrum due to the "noninteracting" part must have a continuum part that can be made identical with that of the exact Hamiltonian once mass renormalization has been carried out. As we point out in Sec. IV, $\widetilde{H}_A$ can satisfy that requirement of an "interaction-free" part of a Hamiltonian, but $\widetilde{H}_0$ cannot.

The fact that the nonvanishing vacuum expectation value of ϕ modifies the spectrum and transformation properties of the excitations in nontrivial ways forces us to make an important change in the way we treat the subsidiary condition. In ordinary quantum electrodynamics in which the vacuum state is nondegenerate and shares the U(1) symmetry of the Lagrangian, one of us (K.H.) previously showed the connection between the consistent formulation of the theory and the formulation that leads to the Dyson–Wick expansion and the Feynman rules. In the consistent formulation the "minimal coupling" Hamiltonian is accompanied by the subsidiary condition, $a_Q(\mathbf{k}) + \rho(\mathbf{k})(2|\mathbf{k}|^{3/2})^{-1}|\nu\rangle = 0$, where $\rho(\mathbf{k})$ designates the charge density and $k_\mu a_\mu(\mathbf{k})/(\sqrt{2}|\mathbf{k}|) = a_Q(\mathbf{k})$ is the operator that annihilates one one of the massless ghosts. Alternatively, we can carry out a similarity transformation to a representation in which the subsidiary condition is $a_Q(\mathbf{k})|n\rangle = 0$, and the Hamiltonian is transformed so that ghost-mediated interactions are replaced by terms that include the nonlocal Coulomb interaction. The Dyson–Wick expansion, however, is based on the "hybrid" combination of the minimal coupling Hamiltonian and the subsidiary condition $a_Q(\mathbf{k})|n\rangle = 0$, in which the charge density has been amputated more or less by fiat. The validity of the hybrid combination is demonstrated by a theorem that the S-matrix elements for the hybrid combination and the consistent formulation agree modulo renormalization constants, and that therefore the *ad hoc* amputation of the charge density terms from the subsidiary condition is legitimate in calculating S-matrix elements. In the Higgs model, however, it is no longer permissible to calculate S-matrix elements while totally ignoring the charge density in $\Omega^{(+)}(\mathbf{k})$. It is safe to amputate the part of ρ that is a higher power than linear in creation and/or annihilation operators. The part of ρ that is proportional to e and survives the antidipole limit, however, must be kept. This point will be discussed in more detail in later sections.

II. DERIVATION OF FEYNMAN RULES

A. Antidipole interaction picture

In order to derive the S-matrix rules from a canonical formalism, using the Dyson–Wick expansion, we must establish the time dependence of the fields in an appropriately constructed interaction picture; we must specify the incident and scattered particle states. In addition, it is necessary to implement Matthews' rule[5] by transforming the S matrix from a functional of a T-ordered product of an interaction Hamiltonian into a functional of a T^*-ordered product of an interaction Lagrangian. Lastly, it is important to establish a theorem that relates the S matrix in the formalism in which the subsidiary condition has been implemented with the S matrix in the interaction picture in which the subsidiary condition has been modified by fiat.

There are a number of possible formulations of the model we are discussing in this paper. A formulation of a gauge theory consists of a Hamiltonian and a constraint equation; the latter, in a manifestly covariant gauge, is a subsidiary condition. This constraint has an important effect on the dynamical behavior of the system because it determines the subspace on which the time evolution operator may act. In one formulation the subsidiary condition given in Eq. (1.7a) is combined with the Hamiltonian that stems from the Lagrangian in Eq. (1.3). That Hamiltonian is given by

$$H = H_A + H_B \tag{2.1}$$

with

$$\begin{aligned} H_A = \int d\mathbf{x}\, [& -\tfrac{1}{4} F_{jk} F_{jk} + \tfrac{1}{2}\Pi_\mu \Pi_\mu - i(A_4 \nabla\cdot\mathbf{\Pi} \\ & + \Pi_4 \nabla\cdot\mathbf{A}) + \tfrac{1}{2}M^2 \mathbf{A}\cdot\mathbf{A} + iMA_4\Pi_\chi - M\,\mathbf{A}\cdot\nabla\chi \\ & + \tfrac{1}{2}|\nabla\psi|^2 + \tfrac{1}{2}\Pi_\psi{}^2 + \tfrac{1}{2}(2h\lambda^2)\psi^2 \\ & + \tfrac{1}{2}|\nabla\chi|^2 + \tfrac{1}{2}\Pi_\chi{}^2] \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} H_B = \int d\mathbf{x}\, [& ieA_4(\psi\Pi_\chi - \chi\Pi_\psi) + e\mathbf{A}\cdot(\chi\overleftrightarrow{\nabla}\psi) \\ & + \tfrac{1}{2}e^2 \mathbf{A}\cdot\mathbf{A}(\psi^2 + \chi^2 + 2\lambda\psi) \\ & + h\,(\lambda\psi^3 + \tfrac{1}{4}\psi^4 + \tfrac{1}{4}\chi^4 + \tfrac{1}{2}\psi^2\chi^2 + \lambda\psi\chi^2)]. \end{aligned} \tag{2.3}$$

Since this formulation combines the Hamiltonian H with the subsidiary condition in which the positive frequency part of the guage fixing field has been evaluated exactly, we call this the "exact formulation." Except for trivial multiplicative c-numbers, $\Omega^{(+)}$ is the positive frequency part of G, where frequency is defined with respect to the Hamiltonian H, which includes all interactions. The exact formulation, therefore, represents a consistent formulation of the theory in a manifestly covariant gauge. We can express the exact formulation in a different representation by subjecting all operators and states to the similarity transformation. In that case $\Omega^{(+)}$ transforms into $\widetilde{\Omega}^{(+)}$, which is given by $\widetilde{\Omega}^{(+)} = \kappa(k)B_Q(\mathbf{k})$, and H transforms into $\widetilde{H}$ given in Ref. 2. In the first of these representations of the exact formulation it is a simple task to construct H from the corresponding local Lagrangian given in Eq. (1.3). The subsidiary condition that selects the physically admissible states is very complicated, however. The states that satisfy the subsidiary condition are not Fock states, but complicated coherent superpositions of single mode excitations. It is only with these states that the theory, in this representation, maintains the validity of Gauss' law for matrix elements in the physical subspace. In the second, transformed representation of the exact formulation, the subsidiary condition has a simple form and an obvious interpretation. The admissible particle states are Fock states. The complications that keep the particle states from being Fock states in the first representation reappear in this representation as nonlocalities in the Hamiltonian $\widetilde{H}$. $\widetilde{H}$ is given by $\widetilde{H} = \widetilde{H}_A + \widetilde{H}_B$, and $\widetilde{H}_B$, given in Ref. 2, contains so many complicated nonlocal terms that it is very difficult to use in perturbation calculations, and it is impossible to find a local Lagrangian, in this second representation, that could be used to determine Feynman rules.

It is necessary to circumvent the dilemma that in one representation of the exact formulation the spectrum of admissible states, which satisfy the subsidiary condition, is very complicated, and that in the other representation the Hamiltonian is very complicated. We will circumvent it by constructing a form of the theory in which we combine the Hamiltonian H given in Eq. (2.1) with a substitute subsidiary condition, in which the so-called "positive frequency part of G" is no longer the exact positive frequency part with respect to the Hamiltonian H. This involves changing $\Omega^{(+)}$ by fiat, and the only justification for this procedure lies in our ability to prove (as we will do later in this section) that this arbitrary alteration of the subsidiary condition leaves S-matrix elements unaffected. This parallels the procedure used in ordinary quantum electrodynamics, in which the U(1) symmetry of the Lagrangian is shared by the nondegenerate vacuum. It is easy to lose sight of the fact that the very familiar Feynman rules for that case are based upon ignoring, rather than implementing, the subsidiary condition; and that the simple injunction to avoid longitudinal or timelike photons in initial states stems from a form of the subsidiary condition in which all terms proportional to the electric charge have been amputated by fiat. In this, the Higgs model, however, we will not make so extreme a change in the subsidiary condition. Were we to impose so draconian a measure in this case, we would alter the transformation properties and the spectrum of the particle states, and the theorem that the S-matrix elements are unaffected by the change in the subsidiary condition could no longer be proven. We will, instead, propose the subsitute subsidiary condition

$$\Omega_A{}^{(+)}(\mathbf{k})|\nu\rangle_A = 0\,, \tag{2.4}$$

where

$$\Omega_A{}^{(+)}(\mathbf{k}) = \kappa(k)[B_Q(\mathbf{k}) + \chi(k)B_R{}^\star(-\mathbf{k})] - \frac{M}{\sqrt{2}|\mathbf{k}|}\int \frac{d\mathbf{x}}{(2\pi)^{3/2}}\,\Pi_\chi(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.5}$$

with $M = e\lambda$. $\Omega_A{}^{(+)}(\mathbf{k})$ is the antidipole limit of $\Omega^{(+)}(\mathbf{k})$, i.e., the limit as $e\rightarrow 0$ but $e\lambda$ is kept constant. $\Omega_A{}^{(+)}$ is obtained by amputating the term $e(\chi\Pi_\psi - \psi\Pi_\chi)$ from ρ in $\Omega^{(+)}$ and leaving intact the part of ρ that combines the vacuum expectation value of ϕ, with the fluctuation in the momentum conjugate to χ. When we combine H with the subsidiary condition in Eq. (2.4), we are making a more drastic change than just carrying out a similarity transformation. It is apparent that H and $\Omega_A{}^{(+)}$ cannot both be unitarily equivalent to H and $\Omega^{(+)}$, respectively. We will generalize earlier terminology and denote formulations that are related to the exact formulation by having some operators and states unitarily equivalent, but others not, as "hybrid formulations." Even though $\Omega_A{}^{(+)}$ has a simpler structure than $\Omega^{(+)}$, because the charge density operator has been truncated, the states $|\nu\rangle_A$ that obey Eq. (2.4) still are coherent superpositions of single mode excitations and do not constitute a Fock representation. Since that makes it difficult to use the states $|\nu\rangle_A$, the next step is to subject the entire hybrid formulation, which consists of the Hamiltonian H and the subsidiary condition $\Omega_A{}^{(+)}|n\rangle = 0$, to a similarity transformation. This similarity transformation is chosen to transform $\Omega_A{}^{(+)}$ into $\kappa(k)B_Q(\mathbf{k})$. Since $\Omega_A{}^{(+)}$ is the antidipole limit of $\Omega^{(+)}$ and since $U\Omega^{(+)}(\mathbf{k})U^{-1} = \kappa(k)B_Q(\mathbf{k})$, it is clear that $U_A\Omega_A{}^{(+)}(\mathbf{k})U_A{}^{-1} = \kappa(k)B_Q(\mathbf{k})$ and that the unitary operator needed to transform $\Omega^{(+)}(\mathbf{k})$ into $\kappa(k)B_Q(\mathbf{k})$ is the antidipole limit of U. In Ref. 2 we showed that U can be represented as

$$U = \nu e^{\sigma}e^{D}. \tag{2.6}$$

ν and σ do not depend on e, except when e appears in combination with λ as $M = e\lambda$. Therefore, ν and σ are both unaffected by the antidipole limit. The antidipole limit of D, represented as D_A, is obtained by substituting $-M\Pi_\chi$ for ρ. It is given by

$$D_A = \frac{-M}{2}\int \frac{d\mathbf{k}}{\sqrt{2}k_0|\mathbf{k}|} \times\Big\{[B_R(-\mathbf{k}) - B_R{}^\star(\mathbf{k}) + B_Q(-\mathbf{k}) - B_Q{}^\star(\mathbf{k})] \times\int\frac{d\mathbf{x}}{(2\pi)^{3/2}}\,\Pi_\chi(\mathbf{x})e^{-i\mathbf{k}\cdot\mathbf{x}}\Big\}\,. \tag{2.7}$$

Under the pseudo-unitary transformation $\bar{\xi} = U_A\xi U_A{}^{-1}$, H_A [given in Eq. (2.2)] is transformed into $\widetilde{H}_A$ [given in Eq. (1.8)].[4] This can be shown by direct calculation, but follows more simply from

$$UHU^{-1} = \widetilde{H}_A + \widetilde{H}_B; \tag{2.8}$$

if we take the antidipole limit of Eq. (2.8), we find

$$U_AH_AU_A{}^{-1} = \widetilde{H}_A\,. \tag{2.9}$$

The similarly transformed H_B, given by $\bar{H}_B = U_AH_BU_A{}^{-1}$, can be expressed in a number of ways. The most useful for us is to keep the functional dependence of H_B on fields and their adjoint momenta intact, but to transform each individual operator-valued field and momentum as given by $\xi\rightarrow\bar{\xi}$, where $\bar{\xi} = U_A\xi U_A{}^{-1}$. The transformed H_B is given by

$$\bar{H}_B = \int d\mathbf{x}\,\{ie\bar{A}_4(\psi\bar{\Pi}_\chi - \bar{\chi}\Pi_\psi) + e\mathbf{A}\cdot(\bar{\chi}\nabla\psi) + \tfrac{1}{2}e^2\bar{\mathbf{A}}\cdot\bar{\mathbf{A}}(\psi^2 + \bar{\chi}^2 + 2\lambda\psi) + h(\lambda^3 + \tfrac{1}{4}\psi^4 + \tfrac{1}{4}\bar{\chi}^4 + \tfrac{1}{2}\bar{\chi}^2\psi^2 + \lambda\bar{\chi}^2\psi\}. \tag{2.10}$$

Since ψ and $\bar{\psi}$ (as well as Π_ψ and $\bar{\Pi}_\psi$) are trivially identical, never appears on either ψ or Π_ψ. Superficially, $\bar{H}_B$ may appear very similar to H_B. In fact, H_B and $\bar{H}_B$ are profoundly different because of crucial differences between A_μ and $\bar{A}_\mu$, between χ and $\bar{\chi}$, and similar differences between corresponding adjoint momenta. The gauge-fixing field $G(x)$, for example, is given by

$$G(x)=\frac{i}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{\sqrt{2}}\,[\Omega^{(+)}(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-|\mathbf{k}|x_0)}\\ -\Omega^{(-)}(\mathbf{k})e^{-i(\mathbf{k}\cdot\mathbf{x}-|\mathbf{k}|x_0)}]\,, \tag{2.11}$$

where $\Omega^{(-)}(\mathbf{k})=[\Omega^{(+)}(\mathbf{k})]^\star$. $\Omega^{(+)}(\mathbf{k})$ and $\Omega^{(-)}(\mathbf{k})$ are the massless ghost excitations in this representation; but $\bar{G}(x)$ is given by

$$\bar{G}(x)=\frac{i}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{\sqrt{2}}\,\kappa(k)\,[B_Q(\mathbf{k})e^{i(\mathbf{k}\cdot\mathbf{x}-|\mathbf{k}|x_0)}\\ -B_Q{}^\star(\mathbf{k})e^{-i(\mathbf{k}\cdot\mathbf{x}-|\mathbf{k}|x_0)}] \tag{2.12}$$

and, in this representation, $B_Q(\mathbf{k})$ and $B_Q{}^\star(\mathbf{k})$ denote the massless ghost excitations. $\bar{A}_\mu$, which is time-translated by $\exp(-i\tilde{H}_A t)$, consists of two massive transverse modes and a mixture of modes in longitudinal and timelike components; parts of the longitudinal and timelike modes have the time dependence appropriate for massive excitations, and other parts have the time dependence of massless ghost excitations. The ghost modes in $\bar{A}_\mu$ never mix with massive modes under Lorentz boosts, and $\bar{A}_\mu$ does not represent any physical excitations unambiguously. The most meaningful identification of $\bar{A}_\mu$ that can be given is

$$\bar{A}_\mu=\bar{Z}_\mu+\frac{1}{M}\partial_\mu\bar{\chi}+\frac{1}{M^2}\partial_\mu\bar{G}, \tag{2.13}$$

where $\bar{Z}_\mu$ is the massive Proca field. $\bar{Z}_\mu$ has the momentum representation

$$\bar{Z}_\mu(x)=\frac{1}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}\left\{\left[\sum_{i=1}^{2}\epsilon_\mu{}^{(i)}(\mathbf{k})A_i(\mathbf{k})+\left(\frac{k_0k_j}{M|\mathbf{k}|}\delta_{\mu,j}\right.\right.\right.\\ \left.\left.+\frac{i|\mathbf{k}|}{M}\delta_{\mu,4}\right)\alpha(\mathbf{k})\right]e^{ik_\mu x_\mu}+\left[\sum_{i=1}^{2}\epsilon_\mu{}^{(i)}(\mathbf{k})A_i{}^\star(\mathbf{k})\right.\\ \left.\left.+\left(\frac{k_0k_j}{M|\mathbf{k}|}\delta_{\mu,j}+i\frac{|\mathbf{k}|}{M}\delta_{\mu,4}\right)\alpha^\star(\mathbf{k})\right]e^{-ik_\mu x_\mu}\right\}. \tag{2.14}$$

Equation (2.13) is convenient primarily because it allows us to use $\bar{A}_\mu$ as the abbreviation for the right-hand side of that equation.

The most startling differences arise between χ and $\bar{\chi}$, whereas χ, given by Eq. (1.12) is represented by α and $\alpha^\star$ operators, $\bar{\chi}$ is given by

$$\bar{\chi}(\mathrm{x})=\frac{iM}{(2\pi)^{3/2}}\int\frac{d\mathbf{k}}{2k_0}\\ \times\left[\frac{\sqrt{2}k_0}{|\mathbf{k}|}\right]\left[\left(-\frac{1}{k_0+|\mathbf{k}|}B_R(\mathbf{k})-\frac{2k_0-|\mathbf{k}|}{4k_0(k_0-|\mathbf{k}|)}B_Q(\mathbf{k})\right)\right.\\ \times e^{i\mathbf{k}\cdot\mathbf{x}}+\left(\frac{1}{k_0+|\mathbf{k}|}B_R{}^\star(\mathbf{k})\right.\\ \left.\left.+\frac{2k_0-|\mathbf{k}|}{4k_0(k_0-|\mathbf{k}|)}B_Q{}^\star(\mathbf{k})\right)e^{-i\mathbf{k}\cdot\mathbf{x}}\right]; \tag{2.15}$$

$\bar{\chi}$ is entirely represented by excitations of electrodynamic ghosts and is totally devoid of α and $\alpha^\star$ excitations! The α and $\alpha^\star$ excitations have been completely absorbed into Z_μ, where they represent the zero-helicity mode of the massive spin-1 boson. This hybrid formulation permits correct identification of incident and scattered particles, by representing them as fluctuations about the stable degenerate vacuum (to the extent to which the stable vacuum equilibrium point is given by the tree approximation). In the representation we have used for the hybrid formulation these fluctuations are Fock states. This representation leads to a conveniently managable expression for the interaction Hamiltonian, and allows us to derive simple Feynman rules.

B. Particle states and γ-vacuum

We will discuss the vacuum and the particle states in the hybrid formulation in the representation in which the subsidiary condition is $\bar{\Omega}_A{}^{(+)}(\mathbf{k})|n\rangle=0$ or, equivalently, $B_Q(\mathbf{k})|n\rangle=0$. In this interaction picture the vacuum is the ground state for the system of massive vector bosons and Higgs particles. In order to obey the subsidiary condition, the vacuum must be free of $B_R{}^\star(\mathbf{k})$ excitations since $B_Q(\mathbf{k})$ and $B_R{}^\star(\mathbf{k})$ fail to commute. We can choose a state $|0\rangle$ as the vacuum state, and define it by $B_Q(\mathbf{k})|0\rangle=0$, $\alpha(\mathbf{k})|0\rangle=0$, $A_i(\mathbf{k})|0\rangle=0$, $\beta(\mathbf{p})|0\rangle=0$, and $B_R(\mathbf{k})|0\rangle=0$. $|0\rangle$ would be a serviceable vacuum state but is not a necessary choice. $B_R(\mathbf{k})|0\rangle=0$ is neither physically necessary nor is it required for mathematical consistency. On the other hand, $|0\rangle$ is not an eigenstate of $\tilde{H}_A$. Since $\tilde{H}_A$ contains terms that are bilinear in $B_Q(\mathbf{k})$ and others that are bilinear in $B_Q{}^\star(\mathbf{k})$, $\tilde{H}_A|0\rangle$ protrudes into the ghost sector. Ghosts are not observable, and $B_Q(\mathbf{k})$ ghosts are harmless to the consistency of the formulation [provided they are not accompanied by other $B_R{}^\star(\mathbf{k})$ ghosts]. It is therefore not essential that $|0\rangle$ be an eigenstate of $\tilde{H}_A$ in the $B_Q{}^\star(\mathbf{k})$ ghost sector. Nevertheless, it is possible to construct a different vacuum state, which we will designate by $|\gamma\rangle$ that does satisfy

$$\tilde{H}_A|\gamma\rangle=0. \tag{2.16}$$

The vacuum state $|\gamma\rangle$ has the advantage that since it is an eigenstate of $\tilde{H}_A$, time displacement invariance holds in the interaction picture, even in the ghost components of propagators. $|\gamma\rangle$ can be represented as

$$|\gamma\rangle=\exp\left\{-\tfrac{1}{2}\int\frac{d\mathbf{k}}{16k_0{}^2|\mathbf{k}|}(k_0-|\mathbf{k}|)(k_0+|\mathbf{k}|)[B_Q{}^\star(\mathbf{k})\right.\\ \left.-B_Q(-\mathbf{k})][B_Q{}^\star(-\mathbf{k})-B_Q(\mathbf{k})]\right\} \tag{2.17}$$

and has the norm $\langle\gamma|\gamma\rangle=1$. Unlike $|0\rangle$, the vacuum state $|\gamma\rangle$ does not necessarily vanish when a ghost annhilation operator is applied to it. Although $A_i(\mathbf{k})|\gamma\rangle=0$, $\alpha(\mathbf{k})|\gamma\rangle=0$, $\beta(\mathbf{p})|\gamma\rangle=0$, and $B_Q(\mathbf{k})|\gamma\rangle=0$ hold, $B_R(k)|\gamma\rangle=-[(k_0-3|\mathbf{k}|)(k_0+|\mathbf{k}|)/8k_0|\mathbf{k}|]B_Q{}^\star(\mathbf{k})|\gamma\rangle$. The n-particle eigenstates of $\tilde{H}_A$ can be constructed by applying creation and annhilation operators to $|\gamma\rangle$. The scattering matrix is obtained by taking the matrix element of the operator

$$S=T\exp\left[-i\int_{-\infty}^{\infty}\bar{H}_B(t)\,dt\right] \tag{2.18}$$

between eigenstates of $\tilde{H}_A$ that describe the appropriate initial and final particle configurations. The appropriate particle states were previously identified and, in the one-particle sector, are given in Table I.

The S matrix is then expanded and the Wick theorem leads to the reduction to propagators, in the form of vacuum expectation values in the $|\gamma\rangle$ vacuum, and vertices. We find, for example, that the Proca propagator is

$$\begin{aligned}\langle\gamma|T[\bar{Z}_\mu(x)\bar{Z}_\nu(y)]|\gamma\rangle &= -i\int\frac{d^4k}{(2\pi)^4}\frac{e^{ik_\mu(x-y)_\mu}}{k^2+M^2-i\epsilon} \\ &\quad\times\left[\delta_{\mu,\nu}+\frac{k_\mu k_\nu}{M^2}\right]+\frac{i}{M^2}\delta_{\nu,4}\delta_{\mu,4}\delta^4(x-y),\end{aligned}\qquad(2.19)$$

where $M=e\lambda$; and the propagator for the Higgs scalar is

$$\langle\gamma|T[\psi(x)\psi(y)]|\gamma\rangle = -i\int\frac{d^4p}{(2\pi)^4}\frac{e^{ip_\mu(x-y)_\mu}}{p^2+(2h\lambda^2)-i\epsilon},\qquad(2.20)$$

where the Higgs mass is given by $2h\lambda^2$. The properties of the $|\gamma\rangle$ vacuum produce a complication in the case of the $\bar{\chi}$ propagator. In the case of products of $\bar{\chi}$ operators the Wick theorem is

$$\begin{aligned}T[\bar{\chi}(x)\bar{\chi}(y)] &= \langle\gamma|T[\bar{\chi}(x)\bar{\chi}(y)]|\gamma\rangle + :\bar{\chi}(x)\bar{\chi}(y): \\ &\quad-\langle\gamma|:\bar{\chi}(x)\bar{\chi}(y):|\gamma\rangle.\end{aligned}\qquad(2.21)$$

Here : : designates normal ordering and the term $\langle\gamma|:\bar{\chi}(x)\bar{\chi}(y)|\gamma\rangle$ is required because the vacuum expectation value of normally ordered products may not vanish. When the ghost annhilation operator $B_R(\mathbf{k})$ acts on $|\gamma\rangle$ the vacuum state does not vanish and that is why this generalization of Wick's theorem is necessary. The fourth-order time-ordered product of $\bar{\chi}$'s is

$$\begin{aligned}&T[\bar{\chi}(x_1)\bar{\chi}(x_2)\bar{\chi}(x_3)\bar{\chi}(x_4)] \\ &\quad= :\bar{\chi}(x_1)\bar{\chi}(x_2)\bar{\chi}(x_3)\bar{\chi}(x_4): + \tau + \tau_\gamma,\end{aligned}\qquad(2.22a)$$

where

$$\begin{aligned}\tau = {}& :\bar{\chi}(x_1)\bar{\chi}(x_2):\Delta_\chi(x_3-x_4) + :\bar{\chi}(x_1)\bar{\chi}(x_3):\Delta_\chi(x_2-x_4) \\ &+ :\bar{\chi}(x_1)\bar{\chi}(x_4):\Delta_\chi(x_2-x_3) + :\bar{\chi}(x_2)\bar{\chi}(x_3):\Delta_\chi(x_1-x_4) \\ &+ :\bar{\chi}(x_2)\bar{\chi}(x_4):\Delta_\chi(x_1-x_3) + :\bar{\chi}(x_3)\bar{\chi}(x_4):\Delta_\chi(x_1-x_2) \\ &+ \Delta_\chi(x_1-x_2)\Delta_\chi(x_3-x_4) + \Delta_\chi(x_1-x_2)\Delta_\chi(x_2-x_4) \\ &+ \Delta_\chi(x_1-x_4)\Delta_\chi(x_2-x_3)\end{aligned}\qquad(2.22b)$$

TABLE I. Particle states in the one-particle sector of the abelian Higgs model.

Particle state	Description
$A_i^\star(\mathbf{k})\|\gamma\rangle$	transverse components of massive boson
$\alpha^\star(\mathbf{k})\|\gamma\rangle$	zero-helicity component of massive vector boson
$\beta^\star(\mathbf{p})\|\gamma\rangle$	Higgs scalar
$B_Q^\star(\mathbf{k})\|\gamma\rangle$	allowed ghost, permitted by the subsidiary condition
$B_R^\star(\mathbf{k})\|\gamma\rangle$	forbidden ghost; this state violates the subsidiary condition

and

$$\begin{aligned}\tau_\gamma = {}& -:\bar{\chi}(x_1)\bar{\chi}(x_2):\langle\gamma|:\bar{\chi}(x_3)\bar{\chi}(x_4):|\gamma\rangle - :\bar{\chi}(x_1)\bar{\chi}(x_3): \\ &\times\langle\gamma|:\bar{\chi}(x_2)\bar{\chi}(x_4):|\gamma\rangle - :\bar{\chi}(x_1)\bar{\chi}(x_4):\langle\gamma|:\bar{\chi}(x_2)\bar{\chi}(x_3):|\gamma\rangle \\ &- :\bar{\chi}(x_2)\bar{\chi}(x_3):\langle\gamma|:\bar{\chi}(x_1)\bar{\chi}(x_4):|\gamma\rangle - :\bar{\chi}(x_2)\bar{\chi}(x_4): \\ &\times\langle\gamma|:\bar{\chi}(x_1)\bar{\chi}(x_3):|\gamma\rangle - :\bar{\chi}(x_3)\bar{\chi}(x_4):\langle\gamma|:\bar{\chi}(x_1)\bar{\chi}(x_2):|\gamma\rangle \\ &- \Delta_\chi(x_1-x_2)\langle\gamma|:\bar{\chi}(x_3)\bar{\chi}(x_4):|\gamma\rangle - \langle\gamma|:\bar{\chi}(x_1)\bar{\chi}(x_2):|\gamma\rangle \\ &\times[\Delta_\chi(x_3-x_4) - \langle\gamma|:\bar{\chi}(x_3)\bar{\chi}(x_4):|\gamma\rangle] - \Delta_\chi(x_1-x_3) \\ &\times\langle\gamma|:\bar{\chi}(x_2)\bar{\chi}(x_4):|\gamma\rangle - \langle|\gamma:\bar{\chi}(x_1)\bar{\chi}(x_3):|\gamma\rangle \\ &\times[\Delta_\chi(x_2-x_4) - \langle\gamma|:\bar{\chi}(x_2)\bar{\chi}(x_4):|\gamma\rangle] \\ &- \Delta_\chi(x_1-x_4)\langle\gamma|:\bar{\chi}(x_2)\bar{\chi}(x_3):|\gamma\rangle - \langle\gamma|:\bar{\chi}(x_1)\bar{\chi}(x_4):|\gamma\rangle \\ &\times[\Delta_\chi(x_2-x_3) - \langle\gamma|:\bar{\chi}(x_2)\bar{\chi}(x_3):|\gamma\rangle].\end{aligned}\qquad(2.22c)$$

The τ_γ terms are the ones that include vacuum expectation values of normally ordered products. We have observed that the nonvanishing vacuum expectation values of normally ordered products vanish identically in S-matrix expressions, and this cancellation appears to be a general feature of this model. The $\bar{\chi}$ propagator is

$$\langle\gamma|T[\bar{\chi}(x)\bar{\chi}(y)]|\gamma\rangle = -i\int\frac{d^4k}{(2\pi)^4}e^{ik_\mu(x-y)_\mu}\frac{k^2+M^2}{k^4-i\epsilon}.\qquad(2.23)$$

There is no $G(x)$ propagator in this model. Since $B_Q(\mathbf{k})$ and $B_Q^\star(\mathbf{k})$ commute, $\langle\gamma|T[\bar{G}(x)\bar{G}(y)]|\gamma\rangle = 0$. The unusual feature of this S-matrix expansion is the appearance of the mixed propagator $\langle\gamma|T[\bar{\chi}(x)\bar{G}(y)]|\gamma\rangle$. We find that

$$\begin{aligned}T[\bar{\chi}(x)\bar{G}(y)] &= :\bar{\chi}(x)\bar{G}(y): + \langle\gamma|T[\bar{\chi}(x)\bar{G}(y)]|\gamma\rangle \\ &\quad-\langle\gamma|:\bar{\chi}(x)\bar{G}(y):|\gamma\rangle,\end{aligned}\qquad(2.24)$$

where the normally ordered product $:\bar{\chi}(x)\bar{G}(y):$ has a vanishing expectation value. The mixed propagator is

$$\langle\gamma|T[\bar{\chi}(x)\bar{G}(y)]|\gamma\rangle = iM\int\frac{d^4k}{(2\pi)^4}\frac{e^{ik_\mu(x-y)_\mu}}{k^2-i\epsilon}.\qquad(2.25)$$

C. Matthews' rule, surface terms, and Feynman graphs

In the antidipole limit interaction picture, the time translation operator is $\exp(-iH_A t)$. Under this time dependence we find that $\Pi_\psi = \partial_0\psi$ and $\bar{\Pi}_\chi = \partial_0\bar{\chi} - iM\bar{A}_4$. With this substitution $\widetilde{H}_B$ can be written in the form

$$\widetilde{H}_B = -\int d\mathbf{x}\,[\bar{\mathscr{L}}_B + \tfrac{1}{2}e^2\bar{A}_4\bar{A}_4(\psi^2+\bar{\chi}^2)],\qquad(2.26)$$

where

$$\begin{aligned}\bar{\mathscr{L}}_B = {}& e\bar{A}_\mu(\psi\partial_\mu\bar{\chi} - \bar{\chi}\partial_\mu\psi) - eM\bar{A}_\mu\bar{A}_\mu\psi \\ &- \tfrac{1}{2}e^2\bar{A}_\mu\bar{A}_\mu(\psi^2+\bar{\chi}^2) \\ &- h(\lambda\psi^3 + \tfrac{1}{4}\psi^4 + \tfrac{1}{2}\bar{\chi}^2 + \lambda\bar{\chi}^2\psi).\end{aligned}\qquad(2.27)$$

When Wick's theorem is applied to the S-matrix elements discussed in Sec. II B, then the second-order combinations

$$\begin{aligned}S_\chi = {}& -\frac{e^2}{2}\int d^4x\,d^4y\,\bar{A}_4(x)\bar{A}_4(y)\psi(x)\psi(y) \\ &\times\langle\gamma|T[\partial_0\bar{\chi}(x)\partial_0\bar{\chi}(y)]|\gamma\rangle\end{aligned}\qquad(2.28a)$$

and

$$\begin{aligned}S_\psi = {}& -\frac{e^2}{2}\int d^4x\,d^4y\,\bar{A}_4(x)\bar{A}_4(y)\bar{\chi}(x)\bar{\chi}(y) \\ &\times\langle\gamma|T[\partial_0\psi(x)\partial_0\psi(y)]|\gamma\rangle\end{aligned}\qquad(2.28b)$$

arise. The time-ordered product in these expressions can be expanded as shown in

$$\begin{aligned}\theta(x_0-y_0)\partial_0\bar{\chi}(x)\partial_0\bar{\chi}(y)+\theta(y_0-x_0)\partial_0\bar{\chi}(y)\partial_0\bar{\chi}(x)\\ =\frac{\partial}{\partial x_0}[\theta(x_0-y_0)\bar{\chi}(x)\partial_0\bar{\chi}(y)\\ +\theta(y_0-x_0)\partial_0\bar{\chi}(y)\bar{\chi}(x)]\\ +\delta(x_0-y_0)[\bar{\chi}(x),\partial_0\bar{\chi}(y)].\end{aligned}\tag{2.29}$$

The commutator $[\bar{\chi}(x),\partial_0\bar{\chi}(y)]_{x_0=y_0}$ has the value $i\delta(\mathbf{x}-\mathbf{y})$ so that

$$T[\partial_0\bar{\chi}(x)\partial_0\bar{\chi}(y)]=T^*[\partial_0\bar{\chi}(x)\partial_0\bar{\chi}(y)]+i\delta^4(x-y)\tag{2.30}$$

and

$$\begin{aligned}-\frac{e^2}{2}\int d^4x d^4y\,\bar{A}_4(x)\bar{A}_4(y)\psi(x)\psi(y)\langle\gamma|T[\partial_0\bar{\chi}(x)\partial_0\bar{\chi}(y)]|\gamma\rangle\\ =-\frac{e^2}{2}\int d^4x\,d^4y\,\bar{A}_4(x)\bar{A}_4(y)\psi(x)\psi(y)\\ \times\langle\gamma|T^*[\partial_0\bar{\chi}(x)\partial_0\bar{\chi}(y)]|\gamma\rangle\\ -\frac{ie^2}{2}\int d^4x\,\bar{A}_4(x)\bar{A}_4(x)\psi^2(x).\end{aligned}\tag{2.31}$$

Similarly S_ψ turns into $-(e^2/2)\int d^4xd^4y\,\bar{A}_4(x)\bar{A}_4(y)\bar{\chi}(x)\times\bar{\chi}(y)\langle\gamma|T^*[\partial_0\psi(x)\partial_0\psi(y)]|\gamma\rangle-i(e^2/2)\int d^4x\,\bar{A}_4(x)\bar{A}_4(x)\times\bar{\chi}^2(x)$ and the substitution of the T^*-ordered product for the T-ordered product eliminates the normal dependent term $(ie^2/2)\int d^4x\,\bar{A}_4(x)\bar{A}_4(x)[\bar{\chi}^2(x)+\psi^2(x)]$ in the first-order S-matrix element. In the nth-order S-matrix element $n(n-1)$ contributions in S_n in which T^*-ordered products replace the T-ordered products cancel $n-1$ contributions in $S_{(n-1)}$, in which the normal dependent terms appear.

The effect of these substitutions is to replace S in Eq. (2.18) by the expression

$$S=T^*\exp\left[i\int d^4x\,\bar{\mathscr{L}}_B(x)\right].\tag{2.32}$$

We have found it very useful to express $\bar{\mathscr{L}}_B$ in terms of the Proca boson $\bar{Z}_\mu$ instead of the somewhat unphysical $\bar{A}_\mu$. For that purpose $\bar{\mathscr{L}}_B$ is rewritten by making the substitution

$$\begin{aligned}-e\bar{A}_\mu\bar{\chi}\partial_\mu\psi=-e\partial_\mu(\bar{A}_\mu\bar{\chi}\psi)\\ +e\bar{A}_\mu\psi\partial_\mu\bar{\chi}+e\bar{G}\bar{\chi}\psi\end{aligned}\tag{2.33}$$

and expressing $\mathscr{L}_B$ as

$$\bar{\mathscr{L}}_B=\bar{\mathscr{L}}_{B(V)}+\bar{\mathscr{L}}_{B(S)},\tag{2.34}$$

where $\bar{\mathscr{L}}_{B(S)}=-e\partial_\mu(\bar{A}_\mu\bar{\chi}\psi)$. Then the substitution shown in Eq. (2.13) is made in $\bar{\mathscr{L}}_{B(V)}$. This leads to the expression

$$\begin{aligned}\bar{\mathscr{L}}_{B(V)}=&-eM\bar{Z}_\mu\bar{Z}_\mu\psi-\frac{2e}{M}\bar{Z}_\mu\partial_\mu\bar{G}\psi+e\bar{G}\psi\bar{\chi}\\ &-\frac{e}{M^3}\partial_\mu\bar{G}\partial_\mu\bar{G}\psi+\frac{e}{M}\partial_\mu\bar{\chi}\partial_\mu\bar{\chi}\psi-\tfrac{1}{2}e^2\bar{Z}_\mu\bar{Z}_\mu\psi^2\\ &-\tfrac{1}{2}e^2\bar{Z}_\mu\bar{Z}_\mu\bar{\chi}^2-\bar{M}e^2\bar{Z}_\mu\partial_\mu\bar{\chi}\psi^2-\frac{e^2}{M^2}\bar{Z}_\mu\partial_\mu\bar{G}\psi^2\\ &-\frac{1}{2}\frac{e^2}{M^2}\partial_\mu\bar{\chi}\partial_\mu\bar{\chi}\psi^2-\frac{e^2}{M^3}\partial_\mu\bar{\chi}\partial_\mu\bar{G}\psi^2\\ &-\frac{1}{2}\frac{e^2}{M^4}\partial_\mu\bar{G}\partial_\mu\bar{G}\psi^2\\ &-\frac{e^2}{M}\bar{Z}_\mu\partial_\mu\bar{\chi}\bar{\chi}^2-\frac{e^2}{M^2}\bar{Z}_\mu\partial_\mu\bar{G}\bar{\chi}^2\\ &-\frac{1}{2}\frac{e^2}{M^2}\partial_\mu\bar{\chi}\partial_\mu\bar{\chi}\bar{\chi}^2\\ &-\frac{e^2}{M^3}\partial_\mu\bar{\chi}\partial_\mu\bar{G}\bar{\chi}^2-\frac{1}{2}\frac{e^2}{M^4}\partial_\mu\bar{G}\partial_\mu\bar{G}\bar{\chi}^2\\ &-h(\lambda\psi^3+\tfrac{1}{4}\psi^4+\tfrac{1}{4}\bar{\chi}^4+\tfrac{1}{2}\psi^2\bar{\chi}^2+\lambda\psi\bar{\chi}^2)\end{aligned}\tag{2.35a}$$

and

$$\bar{\mathscr{L}}_{B(S)}=-e\partial_\mu\left[\left(\bar{Z}_\mu+\frac{1}{M}\partial_\mu\bar{\chi}+\frac{1}{M^2}\partial_\mu G\right)\psi\bar{\chi}\right].\tag{2.35b}$$

$\bar{\mathscr{L}}_{B(S)}$, when integrated over the space–time continuum, leaves a surface term that in first order can be ignored. In higher order it gives rise to nonvanishing contributions because it appears in time-ordered products. For example, the expression $-(\frac{1}{2})e^2T\{\partial_\mu[\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)]\partial_\nu[\bar{A}_\nu(y)\bar{\chi}(y)\psi(y)]\}$, which occurs in the second-order S matrix, is given by

$$\begin{aligned}-\tfrac{1}{2}e^2T\{\partial_\mu[\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)]\partial_\nu[\bar{A}_\nu(y)\bar{\chi}(y)\psi(y)]\}\\ =-\frac{1}{2}e^2\frac{\partial}{\partial x_\mu}T\{\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)\partial_\nu[\bar{A}_\nu(y)\bar{\chi}(y)\psi(y)]\}\\ -\frac{ie^2}{2}\{\bar{A}_4{}^2(x)[\psi^2(x)+\bar{\chi}^2(x)]+\bar{\chi}^2(x)\psi^2(x)\}\delta^4(x-y).\end{aligned}\tag{2.36}$$

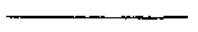
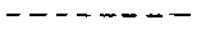
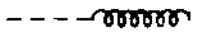
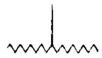
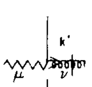
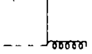
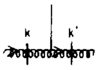
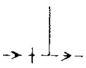
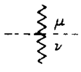
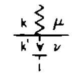
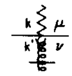
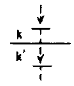
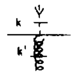
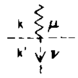
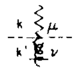
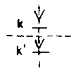
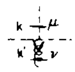
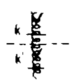
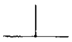
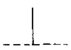
The term $(\frac{1}{2})e^2(\partial/\partial x_\nu)(\partial/\partial y_\nu)T[\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)\bar{A}_\nu(y)\times\bar{\chi}(y)\psi(y)]$ is a surface term that gives a vanishing contribution, but $\frac{1}{2}ie^2\delta(x_0-y_0)[(\partial/\partial x_\mu(x)\bar{\chi}(x)\psi(x)),\bar{A}_4(y)\bar{\chi}(y)\psi(y)]$ develops nonvanishing contributions that need to be included. Similarly the second-order term $eT\{\partial_\mu[\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)]\times\bar{\mathscr{L}}_{B(V)}(y)\}$ needs to be included. We find that

$$\begin{aligned}eT\{\partial_\mu[\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)]\bar{\mathscr{L}}_{B(V)}(y)\}\\ =e^2\frac{\partial}{\partial x_\mu}T[\bar{A}_\mu(x)\bar{\chi}(x)\psi(x)\bar{\mathscr{L}}_{B(V)}(y)]\\ +ie^2[2\bar{A}_4{}^2(x)\psi^2(x)+\psi^2(x)\bar{\chi}^2(x)]\delta^4(x-y).\end{aligned}\tag{2.37}$$

The combined contribution, in second order, of these surface terms is to add an additional seagull term $\frac{1}{2}e^2\psi^2\bar{\chi}^2$ to $\bar{\mathscr{L}}_{B(V)}$. In higher orders additional contributions from $\bar{\mathscr{L}}_{B(S)}$ arise. Analysis similar to the one completed here for the second-order case shows that quite generally the effect of $\bar{\mathscr{L}}_{B(S)}$ is to change the seagull term $-\frac{1}{2}h\psi^2\bar{\chi}^2$ to $(e^2/2M^2)\times(M^2-h\lambda^2)\psi^2\bar{\chi}^2$. In addition, the surface term $\bar{\mathscr{L}}_{B(S)}$ appears in contractions with normal-dependent terms. When these are systematically taken into account, their contributions serve to cancel the $(i/M^2)\delta_{\mu,4}\delta_{\nu,4}\delta^4(x-y)$ term in Eq. (2.19) for $\langle\gamma|T[\bar{Z}_\mu(x)\bar{Z}_\nu(y)]|\gamma\rangle$. With the changes that (1) the Z propagator exclude that normal-dependent term and (2) the seagull be modified as noted, the surface term can then be ignored. The S-matrix elements that we derived from $\bar{\mathscr{L}}_B$ can equivalently be reproduced with Feynman diagrams and the rules for propagators and vertices listed in Table II. The replacement of $\bar{A}_\mu$ by the Proca field $\bar{Z}_\mu$ and the ghost fields and $\bar{G}$ may superficially appear to complicate rather than simplify the evaluation of S-matrix elements. In fact, the elimination of $\bar{A}_\mu$ from $\mathscr{L}_B$ by the use of Eq. (2.13) has two important advantages. In the first place the field contains a mixture of massive vector boson and massless ghost particle excitations. The appearance of $\bar{A}_\mu$ in $\bar{\mathscr{L}}_B$ makes it very difficult to choose a mass parameter as a natural subtraction point in the renormalization procedure. As a result the use of $\bar{A}_\mu$ as the field variable in $\bar{\mathscr{L}}_B$ would involve us in subtractions at unphysical mass values, and require mathematically complicated and uncertain analytic continuation of renormalized quantities.

Beyond that, the Lagrangian $\bar{\mathscr{L}}_B$, when expressed in terms of $\bar{Z}_\mu$, generates many fewer graphs than it would when expressed in terms of $\bar{A}_\mu$. This is true in spite of the fact that $\bar{\mathscr{L}}_B$ may appear to contain fewer vertices when expressed in terms of $\bar{A}_\mu$ than it does after $\bar{A}_\mu$ is replaced by $\bar{Z}_\mu$ and ghost terms. This coupling of fields in propagators is such that for most processes the absence of those vertices eliminates unnecessary graphs.

TABLE II. Feynman rules for the abelian Higgs model.

Feynman rule	Graphical representation	Description
$\frac{1}{(2\pi)^{3/2}(2k_0)^{1/2}}\epsilon_\mu^{(\lambda)}(\vec{k})$		external Proca boson line[a]
$\frac{1}{(2\pi)^{3/2}(2p_0)^{1/2}}$		external Higgs Line
$\frac{1}{(2\pi)^{3/2}(2k_0)^{1/2}}\quad\frac{-iMk_0}{\lvert\vec{k}\rvert(k_0+\lvert\vec{k}\rvert)}$		external pure gauge ghost
$\frac{-i}{k^2+M^2}\left[\delta_{\mu,\nu}+\frac{k_\mu k_\nu}{M^2}\right]$		Proca propagator
$\frac{-i}{p^2+(2h\lambda^2)}$		Higgs propagator
$\frac{-i(k^2+M^2)}{k^4}$		X propagator
$\frac{iM}{k^2}$		X-G propagator
$-2ieM\delta_{\mu,\nu}$		$\bar{\mathscr{L}}_{B^1}$ vertex[b]
$2ek'_\nu\delta_{\mu,\nu}/M$		$\bar{\mathscr{L}}_{B^2}$ vertex[c]
ie		$\bar{\mathscr{L}}_{B^3}$ vertex
$-2iek\cdot k'/M^3$		$\bar{\mathscr{L}}_{B^4}$ vertex
$2iek\cdot k'/M$		$\bar{\mathscr{L}}_{B^5}$ vertex
$-2ie^2\delta_{\mu,\nu}$		$\bar{\mathscr{L}}_{B^6}$ vertex
$-2ie^2\delta_{\mu,\nu}$		$\bar{\mathscr{L}}_{B^7}$ vertex
$2e^2k'_\nu\delta_{\mu,\nu}/M$		$\bar{\mathscr{L}}_{B^8}$ vertex
$2e^2k'_\nu\delta_{\mu,\nu}/M^2$		$\bar{\mathscr{L}}_{B^9}$ vertex
$-2ie^2k\cdot k'/M^2$		$\bar{\mathscr{L}}_{B^{10}}$ vertex
$-2ie^2k\cdot k'/M^3$		$\bar{\mathscr{L}}_{B^{11}}$ vertex
$-2ie^2k\cdot k'/M^4$		$\bar{\mathscr{L}}_{B^{12}}$ vertex
$3e^2k'_\nu\delta_{\mu,\nu}/M$		$\bar{\mathscr{L}}_{B^{13}}$ vertex
$2e^2k'_\nu\delta_{\mu,\nu}/M^2$		$\bar{\mathscr{L}}_{B^{14}}$ vertex
$-2ie^4k\cdot k'/M^2$		$\bar{\mathscr{L}}_{B^{15}}$ vertex
$-3ie^2k\cdot k'/M^3$		$\bar{\mathscr{L}}_{B^{16}}$ vertex
$-2ie^2k.k'/M^4$		$\bar{\mathscr{L}}_{B^{17}}$ vertex
$-3!ih\lambda$		$\bar{\mathscr{L}}_{B^{18}}$ vertex
$-3!ih$		$\bar{\mathscr{L}}_{B^{19}}$ vertex
$-3!ih$		$\mathscr{L}_{B^{20}}$ vertex
$-ie^2(2h\lambda^2-2M^2)/M^2$		$\mathscr{L}_{B^{21}}+\mathscr{L}_{B^{23}}$ vertex
$-2!ih$		$\bar{\mathscr{L}}_{B^{22}}$ vertex

---

[a] The polarization vector $\epsilon_\mu^{(\lambda)}(\mathbf{k})$ is that for a Proca field and is given by $\epsilon_\mu^{(i)}(\mathbf{k}) = \epsilon_\mu^{(i)}(\mathbf{k}\ (i = 1, 2), k_0 k_j/M\,|\mathbf{k}|\delta_{\mu,j}, i|\mathbf{k}|/M\delta_{\mu,4}$.

[b] The vertices $\mathscr{L}_{B^i}$ (where $i = 1 \rightarrow 23$) are labeled in the order in which they appear in $\bar{\mathscr{L}}_B$ given in Eqs. (3.35a) and (3.35b).

[c] The arrows that appear in the graphical representations of the vertices indicate the direction of the momenta necessary to obtain the Feynman rules as given.

This effect is primarily due to the fact that $\bar{\mathscr{L}}_{B(V)}$ contains trilinear $\bar{Z}$–$\bar{Z}$–ψ, $\bar{Z}$–$\bar{G}$–ψ, $\bar{G}$–$\bar{\chi}$–ψ, $\bar{G}$–$\bar{G}$–ψ, and $\bar{\chi}$–$\bar{\chi}$–ψ vertices. But when $\bar{A}_\mu$ is used, the trilinear $\bar{A}$–$\bar{A}$–ψ and $\bar{A}$–$\bar{\chi}$–ψ vertices occur. Since there is no $\bar{Z}$–$\bar{\chi}$ mixed propagator, but there is an $\bar{A}$–$\bar{\chi}$ mixed propagator (because of the ghost content of the $\bar{A}$ field) there are many more ways that propagators can couple to external lines when the $\bar{A}_\mu$ is used than when the $\bar{Z}_\mu$ is used. For example, there are two second-order diagrams for ψ–$\bar{Z}$→ψ–$\bar{Z}$ scattering, but five second-order diagrams for ψ–$\bar{A}$→ψ–$\bar{A}$ scattering. In fourth-order calculations, the Lagrangian expressed in terms of $\bar{A}_\mu$ continues to produce many more graphs than $\bar{\mathscr{L}}_B$ expressed in terms of $\bar{Z}_\mu$.

D. *S*-matrix equivalence theorem

The purpose of this section is to demonstrate that the S matrix in the hybrid formulation,

$$\bar{S}_{f,i} = T^*\langle n_f|\exp\left[i\int d^4x\,\bar{\mathscr{L}}_B(x)\right]|n_i\rangle\,, \tag{2.38}$$

may safely be substituted for the S matrix, S, given by

$$S_{f,i} = \delta_{f,i} - 2\pi i\delta(E_f - E_i)\langle n_f|\widetilde{T}|n_i\rangle, \tag{2.39}$$

where

$$\widetilde{T} = \widetilde{H}_B + \widetilde{H}_B[1/(E_i - \widetilde{H}_A - \widetilde{H}_B + i\epsilon)]\widetilde{H}_B. \tag{2.40}$$

Equation (2.39) is the expression for the S matrix that follows from first principles, i.e., from the Hamiltonian $\widetilde{H}_A + \widetilde{H}_B$ and the subsidiary condition $B_Q(\mathbf{k})|n\rangle = 0$. An entirely equivalent form of S can be given by writing

$$\langle n_f|\widetilde{T}|n_i\rangle = \langle \nu_f|T|\nu_i\rangle, \tag{2.41a}$$

where

$$T = H_B + H_B[1/(E_i - H_A - H_B + i\epsilon)]H_B\,, \tag{2.41b}$$

and where $|\nu_i\rangle$ and $|\nu_f\rangle$ obey the subsidiary condition $\Omega^{(+)}|\nu\rangle = 0$. $\langle n_f|\widetilde{T}|n_i\rangle$ and $\langle \nu_f|T|\nu_i\rangle$ are identical because all operators and states in one of these expressions are unitarily equivalent to the corresponding expressions in the other. However, $\bar{S}$ differs from $\widetilde{S}$, a fact easily appreciated when $\bar{S}$ is expressed in a different form. $\bar{S}$ is also given by

$$\bar{S}_{f,i} = \delta_{f,i} - 2\pi i\delta(E_f - E_i)\langle n_f|\bar{T}|n_i\rangle, \tag{2.42a}$$

where

$$\bar{T} = \bar{H}_B + \bar{H}_B[1/(E_i - \widetilde{H}_A - \bar{H}_B + i\epsilon)]\bar{H}_B. \tag{2.42b}$$

In this formula for $\bar{S}$, the states $|n\rangle$ and the subsidiary condition $B_Q(\mathbf{k})|n\rangle = 0$ are the same as in $\widetilde{S}$, but the Hamiltonian $\bar{H}_B$ differs from $\widetilde{H}_B$. Alternatively, we can unitarily transform $\bar{S}$ so that its Hamiltonian is H, exactly as in S, but then its initial and final states would not be the set $|\nu\rangle$ that obeys $\Omega^{(+)}|\nu\rangle = 0$, but a different set of states $|\nu\rangle_A$ that obey $\Omega_A{}^{(+)}(\mathbf{k})|\nu\rangle_A = 0$. In other words, $\bar{S}$ differs from S formally because the states and operators of these two quantities are not unitarily equivalent. In $\bar{S}$ changes in the subsidiary condition have been made by fiat. It is therefore necessary to carefully consider whether the S matrix $\bar{S}$, evaluated with the rules derived from the hybrid formulation in Sec. II C, may be substituted for S.

The argument that the substitution of $\bar{S}$ for S is harmless is based upon the following assumptions. $\bar{H}$, given by $\bar{H} = \bar{H}_A + \bar{H}_B$, and $\widetilde{H}$, given by $\widetilde{H} = \widetilde{H}_A + \widetilde{H}_B$, have the same antidipole limit $\widetilde{H}_A$. $\widetilde{H}$ and $\bar{H}$ are related by a unitary transformation, namely,

$$\widetilde{H} = V\bar{H}V^{-1}\,. \tag{2.43}$$

Where $V^* = V^{-1}$. V can be given explicitly as $V = UU_A{}^{-1}$. The spectrum of states for both Hamiltonians, $\widetilde{H}$ and $\bar{H}$, is determined by $\widetilde{H}_A$. $\widetilde{H}_A$ contains all effects that contribute mass to particle states except for mass renormalization effects contributed by the perturbations that vanish in the antidipole limit. It is convenient to equate the particle masses that appear in $\widetilde{H}_A$ to the physical masses and consider appropriate counterterms to be included in the perturbing Hamiltonian, but to be suppressed because they play no important role in this argument. These assumptions and the details of the argument itself are similar to the ones that are invoked in ordinary quantum electrodynamics, in which the vacuum is nondegenerate and carries the U(1) symmetry.[6]

The argument proceeds from the observation that

$$\widetilde{H}_B = \bar{H}_B V^{-1} + (1 - V^{-1})\widetilde{H} - \widetilde{H}_A(1 - V^{-1}) \tag{2.44}$$

and similarly, that

$$\bar{H}_B = V^{-1}\widetilde{H}_B + \bar{H}(1 - V^{-1}) - (1 - V^{-1})\widetilde{H}_A. \tag{2.45}$$

From Eq. (2.44) and (2.45) we find that the "exact" scattering state $|\Psi_n\rangle$ given by

$$|\widetilde{\Psi}_n\rangle = |n\rangle + [1/(E_i - \widetilde{H} + i\epsilon)]\widetilde{H}_B|n\rangle \tag{2.46}$$

can be expressed as

$$|\widetilde{\Psi}_n\rangle = V|\bar{\Psi}_n\rangle - i\epsilon V[1/(E_n + \bar{H} + i\epsilon)](1 - V^{-1})|n\rangle, \tag{2.47a}$$

with

$$|\bar{\Psi}_n\rangle = |n\rangle + [1/(E_n - \bar{H} + i\epsilon)]\bar{H}_B|n\rangle. \tag{2.47b}$$

$\widetilde{T}_{f,i}$ can be expressed as

$$\widetilde{T}_{f,i} = \langle n_f|[\bar{H}_B V^{-1} + (1 - V^{-1})(\widetilde{H} - E_f)]|\widetilde{\Psi}_{n(i)}\rangle, \tag{2.48}$$

and, since

$$(\widetilde{H} - E_i)|\widetilde{\Psi}_{n(i)}\rangle = [i\epsilon/(E_i - \widetilde{H} + i\epsilon)]\widetilde{H}_B|n_i\rangle\,, \tag{2.49}$$

$\widetilde{T}_{f,i}$ is easily shown to be given by

$$\begin{aligned}\widetilde{T}_{f,i} = \bar{T}_{f,i} - i\epsilon\langle n_f|\\ &\times\{(1 - V^{-1})[1/(E_i - \widetilde{H} + i\epsilon)]\widetilde{H}_B\\ &-\bar{H}_B[1/(E_i - \bar{H} + i\epsilon)](1 - V^{-1})\}\\ &\times|n_i\rangle - (E_i - E_f)\langle n_f|(1 - V^{-1})|\Psi_{n(i)}\rangle.\end{aligned} \tag{2.50}$$

Equation (2.50) is the equivalence theorem that allows us to substitute $\bar{S}$ for S. On the energy shell, the only differences between $\widetilde{T}$ and $\bar{T}$ stem from the term $i\epsilon\Delta$, where $\Delta = \langle n_f|(1 - V^{-1})(E_i - \widetilde{H} + i\epsilon)^{-1}\widetilde{H}_B - \bar{H}_B(E_i - \bar{H} + i\epsilon)^{-1}(1 - V^{-1})|n_i\rangle$. The product $i\epsilon\Delta$ will vanish, in the limit $\epsilon \to 0$, unless Δ has $1/i\epsilon$ singularities. Except for diagrams that are reducible to scattering diagrams with self-energy corrections in external legs, Δ has no such $1/i\epsilon$ singularities. The diagrams with self-energy corrections in external legs affect only the wavefunction renormalization constant. In ordinary, unbroken, quantum electrodynamics, similar arguments lead to the conclusion that it is not the unrenormalized, but the renormalized S matrix for which the identity between the two representations holds.[7] Our surmise is that the same situation obtains in this case. It should be noted that when the masses in $\tilde{H}_A$ are the physical particle masses, Δ cannot produce more severe singularities than $1/i\epsilon$; singularities of the form $(i\epsilon)^{-n}$ with $n > 1$ can not occur in Δ when H has been mass renormalized.

E. Z-$\psi \rightarrow Z$-ψ scattering

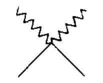
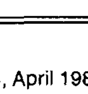
In this section we use the Feynman rules of Table II to calculate Z-$\psi \rightarrow Z$-ψ scattering which allows us to make some comparisons between the Higgs model when formulated in the ¯ representation and with other versions and representations of the theory. The $O(e^2)$ Feynman diagrams that contribute to Z-$\psi \rightarrow Z$-ψ scattering are given in Table III. When we use the Feynman rules in Table II to evaluate the Feynman diagrams in Table III, we find that

$$\begin{aligned}
&\langle Z_{\lambda'}(\mathbf{k}')\beta(\mathbf{p}')|\bar{S}|Z_{\lambda}(\mathbf{k})\beta(\mathbf{p})\rangle \\
&\quad= \frac{i\epsilon_\mu{}^{(\lambda)}(\mathbf{k})\epsilon_\nu{}^{(\lambda')}(\mathbf{k}')}{(2\pi)^2(16k_0k_0'p_0p_0')^{1/2}} \\
&\qquad\times\frac{\delta^4(k+p-k'-p')}{(2h\lambda^2+2p\cdot k')(2h\lambda^2-2p\cdot k)} \\
&\qquad\times\{2\delta_{\mu,\nu}[4M^2(-2h\lambda^2+k\cdot(p-p')) \\
&\qquad-(2h\lambda^2+2p\cdot k')(2h\lambda^2-2p\cdot k)] \\
&\qquad+p_\mu p_\nu'(-2h\lambda^2-2p\cdot k') \\
&\qquad+p_\mu p_\nu(-2h\lambda^2+2p\cdot k)\}. \qquad (2.51)
\end{aligned}$$

Again we note that when we use the form of the Lagrangian given in Eqs. (2.35a) and (2.35b) the calculation is considerably simplified since we need to consider only three diagrams whereas if we were to use the form of the Lagrangian given in Eq. (2.27), we would have to consider nine diagrams.

When the form of the Lagrangian in Eqs. (2.35a) and (2.35b) is used, the massive vector particle states that appear in the incident and final scattering states correspond unambiguously to excitations of the Proca field $\bar{Z}_\mu$. However, when we use the form of the Lagrangian in Eq. (2.27), there is no longer a simple one-to-one correspondence between massive particle states and field excitations. The nontransverse part of $\bar{A}_\mu$ consists of a linear superposition of massive α mode operators and R-type and Q-type ghost mode operators. When testing for "gauge invariance," it is necessary to project the initial and final "pure gauge" ghost, but at the same time, to avoid projecting any part of $\bar{Z}_\mu$. The polarization $k_\mu = (k, |\mathbf{k}|)$, which is the proper projection for the "pure gauge" ghost in ordinary, unbroken, electrodynamics, is quite wrong for this model. We will discuss this point in detail in the next section.

TABLE III. Feynman diagrams for Z-$\psi \rightarrow Z$-ψ scattering.

S-matrix element	Feynman diagram
$M_{12}^{(1)}$	
$M_6^{(1)}$	

We also note that since the longitudinal Proca particle is projected by $\epsilon_\mu{}^{(3)}(\mathbf{k}) = (1/M)(k_0k_j/|\mathbf{k}|\delta_{\mu,j} + i|\mathbf{k}|\delta_{\mu,4})$, in the limit in which $M \rightarrow 0$, Z-$\psi \rightarrow Z$-ψ scattering is dominated by the scattering of longitudinal particle states in agreement with Lowenstein and Schroer.[8]

III. GAUGE INVARIANCE AND THE DYNAMIC DETACHMENT OF PURE GAUGE STATES

Quantum electrodynamics is invariant under the group of U(1) local gauge transformations. The "Higgs" model is a special case of quantum electrodynamics of charged scalar fields. The features that characterize this model are the self-interactions of the scalar field, which make the vacuum degenerate. These features do not interfere with the U(1) invariance of the Lagrangian, but they inhibit the transmission of that invariance to a particular vacuum state. Gauge invariance, when it is understood to include arbitrary gauge functions, is a subtle question, which is probably not fully resolved even now. Gauge invariance in QED is, however, usually understood in a restricted sense to apply to manifestly covariant formulations, in which the gauge-fixing field is free. And it is usually restricted to invariance to transformations in which gauge functions obey D'Alembert's equation. When understood in this sense, gauge invariance is equivalent to the property we call "dynamic detachment of pure gauge ghosts." This property refers to the fact that when "pure gauge" photons are parts of state vectors, S-matrix elements vanish when they connect those state vectors to other state vectors that represent observable states. Pure gauge photon states are those that include the ghost particle that obeys the subsidiary condition. In representations in which the subsidiary condition takes the form $B_Q(\mathbf{k})|n\rangle = 0$, the "pure gauge" states have the form

$$|n\rangle_{\rm pg} = B_Q{}^\star(\mathbf{k}_1)\cdots B_Q{}^\star(\mathbf{k}_n)|n'\rangle, \qquad (3.1)$$

where $|n'\rangle$ is devoid of all ghosts. In other words, $|n\rangle_{\rm pg}$ can have any number of B_Q ghosts and observable particles, but no particles generated by $B_R{}^\star(\mathbf{k})$ acting on other states. Because the unit operator (in the one-particle ghost sector) is $B_R{}^\star(\mathbf{k})|\gamma\rangle\langle\gamma|B_Q(\mathbf{k}) + B_Q{}^\star(\mathbf{k})|\gamma\rangle\langle\gamma|B_R(\mathbf{k})$, and in the many-particle sector the same configuration of ghost creation and annihilation operators obtains, the S matrix to a pure gauge state, $\langle n|_{\rm pg}S|i\rangle$, is the probability amplitude for finding $B_R{}^\star(\mathbf{k})|n'\rangle$ states. Thus the dynamic detachment of pure gauge states implies that the final scattering will be devoid of $B_R{}^\star$ ghosts. Since the S matrix to $\langle n'|B_R(\mathbf{k})$ final states does not vanish, the final scattering state will include $B_Q{}^\star$ ghosts, but since $B_Q{}^\star$ ghosts have vanishing norm, no probability is absorbed by these states, and the dynamic detachment of pure gauge states suffices to guarantee unitarity in the physical subspace.

This account is equally valid for the "Higgs" model and for ordinary quantum electrodynamics, in which there are no self-interactions that make the vacuum state degenerate. There are, however, great differences in the relationships of the ghost particles to the excitations of the vector field that distinguish the Higgs model from ordinary electrodynamics. In ordinary quantum electrodynamics the fact that $G = \partial_\mu A_\mu$, combined with the fact that all time dependence in the interaction picture is "plane wave" [i.e., $\exp(\pm ik_\mu x_\mu)$ with $k_0 = |\mathbf{k}|$] makes the pure gauge photon annihilation and creation operators proportional to $k_\mu a_\mu(\mathbf{k})$ and $k_\mu a_\mu{}^*(\mathbf{k})$, respectively. S-matrix elements to pure gauge ghosts are evaluated by using the external particle polarization k_μ (with $k_0 = |\mathbf{k}|$) to project the pure gauge ghost particle. This accounts for the so-called "test for guage invariance" in ordinary quantum electrodynamics, in which the vacuum is nondegenerate. When a photon line in an S-matrix element is projected with a polarization k_μ, the S-matrix element vanishes.

In the Higgs model, Eq. (2.12) indicates the relation between the pure gauge ghost and vector field excitations. $\partial_\mu A_\mu$ is still equal to G, but the time dependence of the interaction picture fields is more complicated in this case. $\partial_\mu Z_\mu = 0$ is an identity, and the pure gauge ghost excitations remain in $\bar{\chi}$. The pure gauge ghost has become quite detached from the massive Proca field, and the rule for projecting it does not resemble the corresponding rule in ordinary quantum electrodynamics.

The Higgs model in the exact formulation, in the transformed representation in which the subsidiary condition is $\widetilde{\Omega}^{(+)}(\mathbf{k})|n\rangle = 0$ (or equivalently $B_Q(\mathbf{k})|n\rangle = 0$) can trivially be shown to imply vanishing S-matrix elements to pure gauge ghost states. The reason is simply that $\widetilde{H}_B$ has no $B_R(\mathbf{k})$ or $B_R{}^*(\mathbf{k})$ operators anywhere at all. This fact has as an immediate consequence that for a pure gauge state $|n\rangle_{\rm pg} = B_Q{}^*(\mathbf{k})|n'\rangle$ and an initial state $|i\rangle$, where $|i\rangle$ is an ordinary observable particle,

$$\widetilde{T}_{f,i} = \langle n'|B_Q(\mathbf{k})\{\widetilde{H}_B[1/(E_i - \widetilde{H}_A - \widetilde{H}_B + i\epsilon)]\widetilde{H}_B\}|i\rangle \quad (3.2)$$

vanishes, since $B_Q(\mathbf{k})$ commutes with $\widetilde{H}_B$. This fact implies that, in the hybrid formulation,

$$\bar{S}_{f,i} = T^*\langle n'|B_Q(\mathbf{k})\exp\left[i\int d^4x\ \bar{\mathscr{L}}_B(x)\right]|i\rangle \quad (3.3)$$

vanishes. The two formulations in this case are not related by a gauge transformation. It is the equivalence theorem, proven in Sec. II D, that requires $\bar{S}_{f,i}$ to vanish when $S_{f,i}$ does. Even though this is a conclusive argument to demonstrate that $\bar{S}_{f,i}$ to pure gauge states vanishes, we will use the Feynman rules given in Table II to evaluate the lowest-order S-matrix elements for two pure gauge processes, to demonstrate the proper projection for pure gauge ghosts, and to illustrate how these S-matrix elements vanish. First we consider the process Z–$\psi \rightarrow$pure gauge ghost–ψ to $O(e^2)$. The general form of a matrix element in the model is given by

$$\begin{aligned}\bar{S}_{f,i} &= M^{(m)}{}_{1^{n_1},2^{n_2},\ldots,23^{n_{23}}}\\ &= \frac{i^{(n_1+\cdots+n_{23})}}{(n_1+\cdots+n_{23})!}\sum_P T^*\langle\nu_f|\\ &\quad\times\prod_{j_1=1}^{n_1}\int d^4x_{j_1}\ \bar{\mathscr{L}}_{B^1}(x_{j_1})\cdots\prod_{j_{23}=1}^{n_{23}}\\ &\quad\times\int d^4x_{j_{23}}\ \bar{\mathscr{L}}_{B^{23}}(x_{j_{23}})|\nu_i\rangle.\end{aligned} \quad (3.4)$$

The index m labels the set of contractions that leads to a single Feynman graph, with graphs that are related by crossing symmetry included under a single index. The one- or two-digit sub or superscripts i or ij (or n_i or n_{ij} and $\bar{\mathscr{L}}_{B^i}$ or $\bar{\mathscr{L}}_{B^{ij}}$) refer to the interaction term listed in Table II. n_i or n_{ij} refers to the number of times $\bar{\mathscr{L}}_{B^i}$ or $\bar{\mathscr{L}}_{B^{ij}}$, respectively, occurs in the corresponding graph. Σ_P indicates a sum over all permutations of space–time variables. The matrix elements that contribute to Z–$\psi\rightarrow$pure gauge ghost–ψ scattering are obtained by replacing $|i\rangle$ by $|Z_\lambda(\mathbf{k})\beta(\mathbf{p})\rangle$ and $\langle f|$ by $\langle B_Q(k')\beta(p')|$ in Eq. (3.4). In Table IV we exhibit the Feynman diagrams that contribute to Z–$\psi\rightarrow$pure gauge ghost–ψ scattering. When we use the Feynman rules of Table II to evaluate the amplitudes represented by the Feynman diagrams of Table IV, we find

$$M_{2,5}{}^{(1)} = 4L(k,k',p,p')_\mu(s_\mu s_\nu k_\nu'/s^2 + t_\mu t_\nu k_\nu'/t^2), \quad (3.5a)$$

$$M_8{}^{(1)} = -2L(k,k',p,p')_\mu k_\mu', \quad (3.5b)$$

$$M_{2,22}{}^{(1)} = -2(2h\lambda^2)L(k,k',p,p')_\mu(s_\mu/s^2 + t_\mu/t^2), \quad (3.5c)$$

with

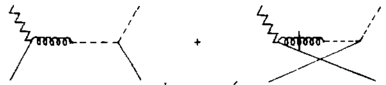
$$L(k,k',p,p')_\mu = \frac{ie^2}{(2\pi)^2(16k_0k_0'p_0p_0')^{1/2}} \times\frac{k_0\epsilon_\mu{}^{(\lambda)}(\mathbf{k})\delta^4(k+p-k'-p')}{|\mathbf{k}'|(k_0'+|\mathbf{k}'|)}, \quad (3.5d)$$

and $s_\mu = (k+p)_\mu$ and $t_\mu = (k'-p)_\mu$. When we evaluate the on-shell amplitude in Eqs. (3.5a), (3.5b), and (3.5c), we find

$$M_{2,5}{}^{(1)} + M_{2,22}{}^{(1)} + M_8{}^{(1)} = 0. \quad (3.6)$$

Equation (3.6) demonstrates the dynamic detachment of pure gauge states in the one ghost sector of the physical subspace. We note that, in order to obtain the correct result

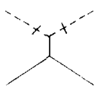
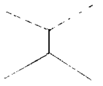
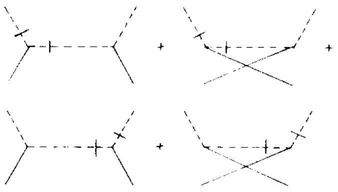
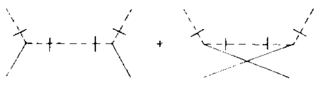
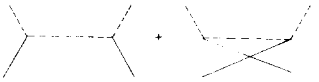
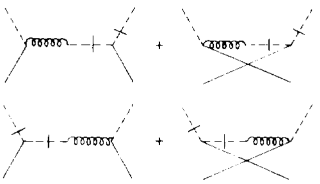
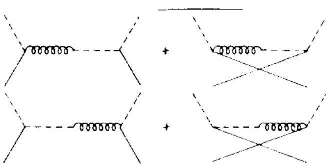
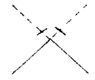
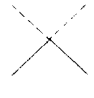
TABLE IV. Feynman diagrams for Z–$\psi\rightarrow$pure gauge ghost–ψ scattering.

S-matrix element	Feynman diagram
$M^{(1)}_{2,5}$ =	
$M^{(1)}_{2,22}$ =	
$M^{(1)}_{8}$ =	

given in Eq. (3.6), we must include the Higgs self-coupling vertex $\mathscr{L}_{B22}$. This is not surprising when we consider that the coupling coefficient h is implicitly dependent on the electric charge and can be expressed as the positive constant $h\lambda = -e(2h\lambda^2)/(2M)$. We also note that, in contrast to the established practice in ordinary quantum electrodynamics, the diagrams in Table IV cannot be obtained by replacing an external Z line with a pure gauge ghost line in Z–ψ→Z–ψ scattering.

We now proceed to demonstrate dynamic detachment of pure gauge states in the two-ghost sector of the physical state space and consider ψ–pure gauge ghost→ψ–pure gauge ghost scattering. In Table V we exhibit the Feynman diagrams that contribute to ψ–pure gauge ghost→ψ–pure gauge ghost scattering. We note that in the last seagull diagram we have included the contribution of the surface term. When we use the Feynman rules in Table II to evaluate the amplitudes represented by the Feynman diagrams in Table V, we find that

TABLE V. Feynman diagrams for ψ–pure gauge ghost→pure gauge ghost–ψ scattering.

S-matrix element	Feynman diagram
$M_{5,18}^{(1)}$ =	
$M_{18,22}^{(1)}$ =	
$M_{22,5}^{(1)}$ =	
$M_{5,5}^{(1)}$ =	
$M_{22,22}^{(1)}$ =	
$M_{5,3}^{(1)}$ =	
$M_{3,22}^{(1)}$ =	
$M_{10}^{(1)}$ =	
$M_{21}^{(1)} + M_{23}^{(1)}$ =	

$$M_{5,18}^{(1)} = N(k,k',p,p')6\,\frac{(2h\lambda^2)}{M^2}\,\frac{k\cdot k'}{(p-p')^2+2h\lambda^2}, \tag{3.7a}$$

$$M_{18,22}^{(1)} = N(k,k',p,p')3\,\frac{(2h\lambda^2)^2}{M^2}\,\frac{1}{(p-p')^2+2h\lambda^2}, \tag{3.7b}$$

$$\begin{aligned} M_{5,22}^{(1)} &= -N(k,k',p,p')2\,\frac{(2h\lambda^2)}{M^2}\left[\frac{s\cdot(k+k')(s^2+M^2)}{s^4}\right. \\ &\quad \left. + \frac{t\cdot(k+k')(t^2+M^2)}{t^4}\right], \end{aligned} \tag{3.7c}$$

$$\begin{aligned} M_{5,5}^{(1)} &= N(k,k',p,p')\,\frac{4}{M^2}\left[\frac{k\cdot sk'\cdot s(s^2+M^2)}{s^4}\right. \\ &\quad \left. + \frac{k\cdot tk'\cdot t\,(t^2+M^2)}{t^4}\right], \end{aligned} \tag{3.7d}$$

$$M_{22,22}^{(1)} = N(k,k',p,p')\,\frac{(2h\lambda^2)^2}{M^2}\left(\frac{s^2+M^2}{s^4} + \frac{t^2+M^2}{t^4}\right), \tag{3.7e}$$

$$M_{3,5}^{(1)} = N(k,k',p,p')2\left[\frac{s\cdot(k+k')}{s^2} + \frac{t\cdot(k+k')}{t^2}\right], \tag{3.7f}$$

$$M_{3,22}^{(1)} = N(k,k',p,p')2(2h\lambda^2)\left(\frac{1}{s^2}+\frac{1}{t^2}\right), \tag{3.7g}$$

$$M_{10}^{(1)} = N(k,k',p,p')\,\frac{2}{M^2}\,k\cdot k', \tag{3.7h}$$

and

$$M_{21}^{(1)} + M_{23}^{(1)} = -N(k,k',p,p')\,\frac{(2h\lambda^2 - 2M^2)}{M^2}, \tag{3.7i}$$

where

$$\begin{aligned} N(k,k',p,p') &= \frac{ie^2}{(2\pi)^2(16k_0k_0'p_0p_0')^{1/2}}\,\frac{k_0k_0'}{|\mathbf{k}|\,|\mathbf{k}'|} \\ &\quad \times \frac{\delta^4(k+p-k'-p')}{(k_0+|\mathbf{k}|)(k_0'+|\mathbf{k}'|)}. \end{aligned} \tag{3.8}$$

When we use the mass shell constraint for the external momenta and the fact that $p\cdot k' = k\cdot p'$, we find

$$\begin{aligned} &M_{5,18}^{(1)} + M_{18,22}^{(1)} + M_{5,22}^{(1)} + M_{5,5}^{(1)} + M_{22,22}^{(1)} \\ &\quad + M_{5,3}^{(1)} + M_{22,3}^{(1)} + M_{10}^{(1)} + M_{22}^{(1)} + M_{23}^{(1)} = 0. \end{aligned} \tag{3.9}$$

Equation (3.9) exhibits detachment of pure gauge states in the two-ghost sector of the physical state space. We note that the contribution of the surface term is essential in effecting important subtractions necessary for the result in Eq. (3.9). The calculations above serve to demonstrate that, in the ~ representation, detachment of pure gauge states obtains as required by the S-matrix equivalence theorem.

IV. RENORMALIZATION AND RELATED MATTERS

It is apparent from earlier parts of this paper that the S matrix for the abelian Higgs model can be given a number of alternative formulations and that these formulations are related by a much broader class of relationships than only gauge transformations. Different formulations are useful for different purposes. For example, in Sec. III we saw that the Lagrangian given in Eqs. (2.35a) and (2.35b), and represented by the Feynman rules in Table II, is very suitable for the demonstration that "gauge invariance holds," i.e., that forbidden ghost states detach dynamically from observable states. Equations (2.35a) and (2.35b) also give the form of the Lagrangian that is most directly applicable to evaluating S-matrix elements when zero-helicity vector particles are among the initial and final states. In this section we will observe that the form of the Lagrangian given in Eq. (2.27) is very useful for a discussion of ultraviolet infinities, and we will make some further comments on the renormalizability of this model.

A number of proofs can be found in the literature that the Feynman rules for the abelian Higgs model lead to renormalizable S-matrix elements.[9–12] One important observation in the published demonstrations is that the abelian Higgs model can be described by Lagrangians in which the interaction terms are products of operator fields with canonical dimensions that make it possible to extract all infinities from a diagram with a number of subtractions that depends only on the diagram's number of external lines. The form of the Lagrangian in our Eq. (2.27) has that property; as a manifestation of that fact, the $\bar{A}_\mu$ propagator is dominated by $\delta_{\mu,\nu}/k^2$ at high frequency. When such a Lagrangian is used, a finite number of counterterms suffices to remove infinities, and the model is renormalizable. We have very little to add to the technical aspects of the argument in Refs. 8–12. Instead we will deal with the dynamical processes in this model that control the highly divergent Proca propagator generated by the $\bar{Z}_\mu$ field, and discuss some other aspects of the renormalization program.

It has sometimes been suggested that the Higgs model is renormalizable because the massive vector particle is "not quite" the conventional one and that the vector propagator differs from the Proca propagator because the vector particle itself differs in some essential way from a Proca spin one particle.[13] Our findings, however, are that the vector particles in this theory (the excitations of the $\bar{Z}_\mu$ field) are quite ordinary Proca particles and give rise to Proca propagators, which by themselves would result in uncontrollable infinities. The theory becomes renormalizable in the case of the abelian Higgs model because of the relation between the Proca field and the ghosts. We remark that in a theory with vector fields and "normal" nondegenerate vacuum, the following options usually exist. Either the fields are massless (like the photon); then they have only two observable helicity modes that transform like scalars,[14] and in the Feynman rules the ghost modes are responsible for mediating the Coulomb interaction. Alternatively, the fields may be massive; then they have three helicity components which transform among themselves under the Lorentz group.[15] The theory then requires no ghost states in its formulation. The novel element in the case of the abelian Higgs model is the simultaneous presence of the Proca field $\bar{Z}_\mu$, with all helicity components included in the positive metric space, and massless ghosts, all coupled to the same sources. It is this that allows the theory to be renormalizable. The electromagnetic origin of the Lagrangian in effect imposes a coupling rule that demands that $\bar{Z}_\mu$ be coupled to other fields only in the combination $A_\mu = Z_\mu + (1/M)\partial_\mu \chi + (1/M^2)\partial_\mu G$, and this insures renormalizable propagators. Thus, for example, if empirical considerations were to indicate the need for Z–Z–ψ coupling, this rule would require us to postulate the A–A–ψ vertex; we would consequently obtain, in addition to Z–Z–ψ, the vertices $(1/M)Z_\mu \partial_\mu G\psi$, $(1/M^2)Z_\mu \partial_\mu \chi\psi$, $(1/M^2)\partial_\mu G\partial_\mu G\psi$, $(1/M^3)\partial_\mu G\partial_\mu \chi\psi$, and $(1/M^4)\partial_\mu \chi\partial_\mu \chi\psi$. However, an arbitrary selection of vertices, involving the combination of vector bosons and ghosts designated by A_μ, would not necessarily preserve unitarity in the physical subspace. The combination of vertices in $\mathscr{L}_B$ preserves unitarity because $\mathscr{L}_B$ leads to equations of motion that guarantee the free field equation $\Box G = 0$ for the gauge-fixing field G. In the transformation of Eq. (2.27) to Eqs. (2.35a) and 2.35b) the coupling rule described above is obscured because the combination of vertices in $\mathscr{L}_B$, by guaranteeing the free-field equation for G, produces cancellation among a number of vertices that result when Eq. (2.27) is expanded to show the Proca and ghost components separately. The Lagrangian given in Eqs. (2.35a) and (2.35b) preserves renormalizability through a set of relationships among different classes of diagrams. We will not discuss that feature any further in this paper, but interested readers can find a discussion elsewhere.[16]

One very important element in the theory is the "antidipole limit" Hamiltonian H_A (and its similarity transform $\tilde{H}_A$). $\tilde{H}_A$, which includes only the tree approximation to the vacuum expectation value of ϕ, gives rise to a particle spectrum with massive particles that Lorentz-transform self-consistently under the time translation imposed by $\tilde{H}_A$. This fact plays a significant role in our ability to develop and renormalize this theory. The fact that the self-mass corrections due to the perturbations imposed by $\mathscr{L}_B$ are identical for all helicity components of Z_μ and that these perturbations leave the ghost modes of G massless allows us to use $\tilde{H}_A$ to describe the spectrum of incident and scattered particle states in calculating S-matrix elements. It is apparent from the Feynman rules in Table II that $\mathscr{L}_B$ contributes self-mass corrections that are identical for all three helicity components of Z_μ. We observe that the helicity modes of the Proca particle are projected by the factor $\epsilon_\mu{}^{(i)}(\mathbf{k})$, where $\epsilon_\mu{}^{(\pm)}(\mathbf{k})$ refers to transverse (massive) helicity modes and $\epsilon_\mu{}^{(3)}(\mathbf{k}) = (1/M)(k_0 k_j/|\mathbf{k}|\delta_{\mu,j} + i|\mathbf{k}|\delta_{\mu,4})$ (with $k_0{}^2 = M^2 + |\mathbf{k}|^2$) refers to the zero-helicity mode. For all three $\epsilon_\mu{}^{(i)}(\mathbf{k})$, it is straightforward to show that $\epsilon_\mu{}^{(i)}k_\mu = 0$ and $\epsilon_\mu{}^{(i')}(\mathbf{k})\epsilon_\mu{}^{(i)}(\mathbf{k}) = 1$. The vertices in Table II are δ functions for 4-momenta, multiplied either by constants or by the momentum of a ghost line. The momentum factor can contract into $\epsilon_\mu{}^{(i)}(\mathbf{k})$ through the δ functions at vertices. $\epsilon_\mu{}^{(i)}(\mathbf{k})$ is therefore either contracted on k_μ (in which case it vanishes) or on an internal momentum. In the latter case the only surviving term in a self-energy calculation is $\Sigma_\mu \epsilon_\mu{}^{(i)}(\mathbf{k})\epsilon_\mu{}^{(i)}(\mathbf{k}) = 1$ (i not summed), multiplied by a helicity-independent integral. The fact that the self-mass corrections due to $\mathscr{L}_B$ affect the three helicity components of the Proca particles in identical ways and the fact that the continued validity of $\Box G = 0$ keeps the ghosts massless allow us to conclude that the spectrum generated by $\tilde{H}_A$ is stable to the perturbation imposed by $\tilde{\mathscr{L}}_B$. If the Hamiltonian had been divided into a "noninteracting" and a "perturbing" part in a different way, so that the spectrum of the former were not stable to the perturbing effects of the latter, we could not have used the form of the S matrix given in Eq. (2.38). We also could not have applied the equivalence theorem given in Sec. II D, which requires that the spectrum of the mass-renormalized "noninteracting" Hamiltonian is not changed by the perturbations. Consider, for example, the use of H_0 (the limit of H as $e \rightarrow 0$ and $e\lambda \rightarrow 0$, but $h\lambda^2 \nrightarrow 0$) as the "noninteracting" part of H and the remaining H_I, as the perturbation. The spectrum generated by H_0 includes massless photons (transverse as well as ghost modes), Higgs scalars, and Goldstone bosons. This spectrum, however, is not stable to the perturbation imposed by H_I (or its equivalent T^*-ordered Lagrangian form). Among the interactions that perturb H_0, two are dependent only on e in combination with the vacuum expectation value of ϕ, as the product $e\lambda = M$. These two are $-(M^2/2)A_\mu A_\mu$ and $MA_\mu \partial_\mu \chi$. They are responsible for vertices $-M^2\delta^4(k-k')$ and $iMk_\mu \delta^4(k-k')$, respectively. The self-energy function of the photon is modified by these two in all possible alternations of photon and Goldstone lines ($\delta_{\mu,\nu}/k^2$ and $1/k^2$, respectively), as shown in Fig. 1. The sum of all $-M^2\delta^4(k-k')$ insertions into the photon propagators can be calculated exactly and is $\delta_{\mu,\nu}/(M^2+k^2)$. The result of all possible Goldstone line insertions into this sum leads to $\delta_{\mu,\nu}/(M^2+k^2) + [M^2k_\mu k_\nu / k^2(M^2+k^2)^2]\{1 + M^2/(M^2+k^2) + \cdots + [M^2/(M^2+k^2)]^n + \cdots\}$, which sums to the "exact propagator $[1/(M^2+k^2)](\delta_{\mu,\nu} + M^2k_\mu k_\nu/k^4)$. For transverse photons this propagator has a pole at $k^2+M^2=0$. But for the ghosts projected with the polarization k_μ the projected propagator becomes $(k^2+M^2)/(k^2+M^2)$ and the pole at $k^2+M^2=0$ disappears. It is apparent that the perturbation due to the two terms $-(M^2/2)A_\mu A_\mu + MA_\mu \partial_\mu \chi$ does not affect the self-mass of the four components of A_μ in identical ways and therefore destabilizes the spectrum. It is also easy to establish that the Goldstone propagator is affected. The Goldstone propagator is given by the graphs in Fig. 2, which sum to $(k^2+M^2)/k^4$. These features make it clear that H_0 is not suitable as an "interaction-free" Hamiltonian to generate a spectrum of incident and scattered particles and to time-translate the interaction picture. What this analysis cannot show are the results we previously demonstrated: that $\tilde{Z}_\mu$ becomes a vector field when $\tilde{H}_A$ time-translates the interaction picture and that the ghost components no longer mix with $\tilde{Z}_\mu$ under Lorentz transformations.

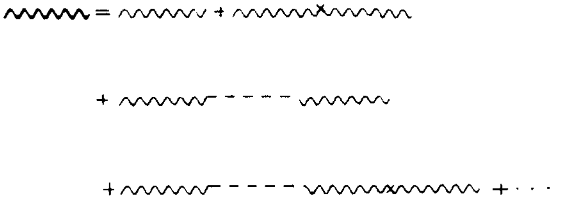


FIG. 1. Feynman diagrams in which e appears only in the combination $e\lambda = M$ and that modify the photon self-energy function.

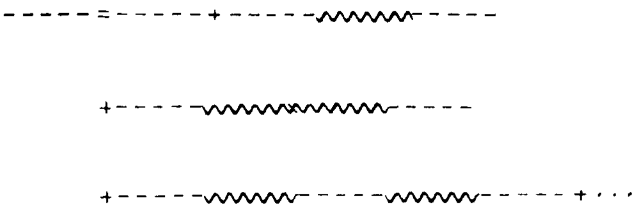


FIG. 2. Feynman diagrams in which e appears only in the combination $e\lambda = M$ and that modify the χ self-energy function.

The infrared divergences in the abelian Higgs model can arise only from ghosts, and their resolution cannot depend on cancellations involving incoherently added elastic radiative processes, such as happens in the Bloch–Nordsiek analysis of ordinary quantum electrodynamics. This is because all observable particles in this model have nonvanishing mass. It is useful to recall that of the components of the Hamiltonian H only $H_{I,T}$, $H_{C(R)}$, and H_h contribute to transition amplitudes. H_Q plays no role in evaluating the amplitude of any physical process. $H_{I,T}$, $H_{C(R)}$, and H_h describe interactions among massive particles only. The only possible source of infrared divergences are the nonlocal form factors $1/|\mathbf{k}|$ or $1/|\mathbf{k}|^2$ that arise in $H_{C(R)}$ and H_h and the transmission of infrared behavior from $H_{C(R)}$ from one factor to another through delta functions in momenta. This opens an avenue for resolving infrared infinities in this model that probably should be explored. We will not, however, address that problem any further in this paper.

V. DISCUSSION

In this section we will summarize some important differences between theories in which vector bosons have dynamical mass and those in which the mass stems from a spontaneously broken symmetry.

Quantum electrodynamics, in a manifestly covariant formulation, can only be formulated in an indefinite metric space with a subsidiary condition, whether the vacuum state is degenerate or not. When the Lagrangian for a vector boson includes a dynamical mass, then its manifestly covariant form can be, and usually is, formulated in a positive-definite metric Hilbert space without any subsidiary constraints that are independent of Lagrange's equations. Nevertheless, this difference in the formulation and in the underlying space does not mark a crucial difference between vector boson theories with dynamical and Higgs masses. It is possible, though not necessary, to formulate a vector boson theory with dynamical mass in an indefinite metric space. When such a theory represents a neutral field interacting with a charged particle, the interactions can be treated consistently by using a subsidiary condition.[17] In such a theory the Lagrangian would be given by

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - G\partial_\mu A_\mu + \tfrac{1}{2}(1-\gamma)G^2 - M^2A_\mu A_\mu - D_\mu{}^\dagger\phi^\dagger D_\mu\phi - m^2\phi^\dagger\phi \tag{5.1}$$

for a charged boson or

$$\mathscr{L} = -\tfrac{1}{4}F_{\mu\nu}F_{\mu\nu} - G\partial_\mu A_\mu + \tfrac{1}{2}(1-\gamma)G^2 - M^2A_\mu A_\mu - \bar{\psi}[m + \gamma_\mu(\partial_\mu - ieA_\mu)]\psi \tag{5.2}$$

for a charged fermion. m designates the positive mass of the charged particle and M the dynamical mass of the neutral vector field. The equations of motion demonstrate that there is a conserved current j_μ and that the gauge-fixing field obeys the equation

$$(\Box - M^2)G = 0. \tag{5.3}$$

This equation allows the imposition of a subsidiary condition

$$G^{(+)}|\nu\rangle = 0 \tag{5.4}$$

and the further development of this equation, to eliminate the ghost degrees of freedom, that has been reported in detail in previous publications. Equations (5.1) and (5.2) are invariant to gauge transformations by gauge functions that obey $(\Box - M^2)\Lambda = 0$.[18] The equations of motion that follow from Eqs. (5.1) and (5.2) include

$$\partial_\nu F_{\mu\nu} - M^2 A_\mu + \partial_\mu G = -j_\mu \tag{5.5}$$

for the appropriate conserved current j_μ, and

$$\partial_\mu A_\mu = (1-\gamma)G. \tag{5.6}$$

When the propagators are evaluated by inverting the equations for the fields, we find that the transverse components of A_μ propagate like a particle with mass M. This M is the same mass as appears in Eq. (5.3) for the ghost components; hence all components of the A_μ propagate at the same rate. The zero-helicity mode of the vector boson is part of the field A_μ and extends into the part of the space in which the metric is indefinite. It is in this respect that this theory differs most markedly and significantly from the abelian Higgs model. In the case of the latter, the transverse components of A_μ become massive, but the ghosts never do, because the equation $\Box G = 0$ never loses its validity. The zero-helicity component of the vector boson has a very different origin. It includes in its composition the particle that would have been the Goldstone boson in the absence of the electromagnetic interaction, but that, in the process of helping the transverse photon to become massive, must accept that same mass itself. To understand in some heuristic sense what mechanisms contribute to making the transverse photons massive, it is necessary to focus one's attention on the vacuum expectation value of ϕ. This quantity carries all the quantum numbers of ϕ, but has no dynamical degrees of freedom. In the evaluation of interaction terms, for example, $e^2\phi^\dagger\phi A_\mu A_\mu$, the part of ϕ that corresponds to the vacuum expectation value can contribute no particle excitations for the various momentum modes. The spatial integral of $e^2\phi^\dagger\phi A_\mu A_\mu$ is $e^2\lambda^2\int A_\mu(\mathbf{x}) \times A_\mu(\mathbf{x})\,d\mathbf{x}$. When we invert the equations of motion, $e^2\lambda^2 A_\mu A_\mu$ contributes an inertia-carrying term to the propagator, whereas $e^2\phi^\dagger\phi A_\mu A_\mu$ would have contributed to the source (vertex) terms only, had ϕ and $(\phi^\dagger)$ consisted of quantized excitation modes instead of developing a vacuum expectation value.

Finally, a comment about Goldstone's theorem. The mechanism by which the abelian Higgs model manages to either evade the Goldstone theorem (in gauges that are not manifestly covariant) or obey it (in manifestly covariant gauges) have often been discussed. In our formulation we have an opportunity to explicitly identify the Goldstone boson and to trace the steps taken to prove the Goldstone theorem, using the antidipole limit Hamiltonian as a realizations of a model in which mass is generated for a vector boson from the vacuum expectation value of a charge bearing field. We find that the charge $\bar{Q}(t) = \int \bar{j}_0(\mathbf{x},t)\,d\mathbf{x}$ is given by

$$\bar{Q}(t) = \frac{-iM}{2(2\pi)^{3/2}} \int d\mathbf{k}\, \frac{|\mathbf{k}|}{k_0}\, \delta(\mathbf{k}) \Big\{ \alpha(\mathbf{k},t) - \alpha^\star(-\mathbf{k},t) + \frac{iM}{\sqrt{2}} \frac{1}{k_0 + |\mathbf{k}|} \left[B_Q(\mathbf{k},t) + B_Q{}^\star(-\mathbf{k},t) \right] \Big\}, \tag{5.7}$$

where the time dependence is imposed by $\exp(-i\widetilde{H}_A t)$. The form of time-dependent excitation operators has been reported by us previously and are

$$\alpha(\mathbf{k},t) = \alpha(\mathbf{k})e^{-ik_0 t}, \tag{5.8a}$$

$$\alpha^\star(\mathbf{k},t) = \alpha(\mathbf{k})e^{ik_0 t}, \tag{5.8b}$$

$$B_Q(\mathbf{k},t) = B_Q(\mathbf{k})e^{-i|\mathbf{k}|t}, \tag{5.8c}$$

and

$$B_Q{}^\star(\mathbf{k},t) = B_Q{}^\star(\mathbf{k})e^{i|\mathbf{k}|t}. \tag{5.8d}$$

This explicit time dependence of Eq. (5.1) still does not permit us to unambiguously determine whether Q is time independent, because of the indeterminancies inherent in the product of the operator-valued $\alpha(\mathbf{k},t)$, $\alpha^\star(-\mathbf{k},t)$, $B_Q(\mathbf{k},t)$, $B_Q{}^\star(-\mathbf{k},t)$, and the factor $|\mathbf{k}|\delta(\mathbf{k})$. Because of these difficulties in representing the total charge operator it is not clear how to carry out the step in the proof of Goldstone's theorem that requires evaluation of $[\bar{Q},\bar{\chi}(\mathbf{x}',t')]$. However, if we evaluate the commutator $[\bar{j}_0(\mathbf{x},t),\bar{\chi}(\mathbf{x}',t')]$ first and only then perform the $\int d\mathbf{x}$, we find that

$$\int d\mathbf{x}\,[\bar{j}_0(\mathbf{x},t),\chi(\mathbf{x}',t')] = iM. \tag{5.9}$$

If we evaluate the vacuum expectation value of Eq. (5.9) in the γ vacuum, we find that the massless "Goldstone" excitations, whose existence is a necessary consequence of Eq. (5.6), are the electrodynamic ghosts.

ACKNOWLEDGMENTS

Research was supported in part by Department of Energy Grant No. DE-AC02-79ER10336.A and in part by the University of Connecticut Research Foundation.

---

[1] K. Haller and R. B. Sohn, Phys. Rev. D **14**, 479 (1976).

[2] K. Haller and G. Bishop, Phys. Rev. D **21**, 368 (1980).

[3] P. W. Higgs, Phys. Rev. **145**, 1156 (1966); G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, in *Advances in Particle Physics*, edited by R. L. Cool and R. E. Marshak (Interscience, New York, 1968), Vol. 2, p. 567; F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).

[4] In Ref. 2 we used ˜ to designate both the exact formulation, in the representation in which the subsidiary condition is given by the operator $U\Omega^{(+)}(\mathbf{k})U^{-1}$, and the antidipole limit formulation, in the representation in which the subsidiary condition is given by the operator $U_A\Omega^{(+)}U_A^{-1}$. At that time we used ˜ to designate both representations since the form of the operator that defines the subsidiary condition has the form $\kappa(k)B_Q(\mathbf{k})$ in both representations, and since in the earlier work there seemed little danger of confusion between the two. For greater clarity we now make a

notational distinction between these two different cases, and use $\bar{\ }$ for the exact theory and $\overline{\ }$ for the antidipole limit theory, when each version of the theory is formulated in the representation in which the subsidiary condition is defined by the operator $\kappa(k)B_Q(\mathbf{k})$. Note that in this notation $\widetilde{H}_A$ and $\overline{H}_A$ are identical, and we use $\widetilde{H}_A$ throughout this paper. For most other operators ξ, $\tilde{\xi}$ and $\bar{\xi}$ differ.

[5] P. T. Matthews, Phys. Rev. **76**, 684, 1489 (1949); Y. Takahashi, *An Introduction to Field Quantization* (Pergamon, London, 1968).

[6] K. Haller, Acta Phys. Austriaca **42**, 163 (1975); K. Haller and L. Landovitz, Phys. Rev. D **2**, 1498 (1970).

[7] K. Haller, Ref. 6; I. Bialynicki–Birula, Phys. Rev. **155**, 1414 (1967).

[8] J. H. Lowenstein and B. Schroer, Phys. Rev. D **10**, 2513 (1974).

[9] B. W. Lee, Phys. Rev. D **4**, 823 (1971).

[10] T. Appelquist, J. Carrazone, T. Goldman, and H. Quinn, Phys. Rev. D **8**, 1747 (1973).

[11] C. Beccei, A. Roua, and R. Stora, Commun. Math. Phys. **42**, 127 (1974).

[12] J. H. Lowenstein, M. Weinstein, and W. Zimmerman, Phys. Rev. D **10**, 1854, 2500 (1974); J. H. Lowenstein and B. Schroer, Ref. 8.

[13] J. Bernstein, Rev. Mod. Phys. **46**, 7 (1974).

[14] K. Haller and R. B. Sohn, J. Math Phys. **19**, 1589 (1977).

[15] A. Proca, J. Phys. Radium **7**, 347 (1936).

[16] G. Bishop, Ph.D. thesis, University of Connecticut, 1981.

[17] K. Haller, Phys. Rev. D **8**, 1796 (1973).

[18] G. Feldman and P. T. Matthews, Phys. Rev. **130**, 1633 (1963).

Yang–Mills theory in null-path space[a)]

S. L. Kent
Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260
E. T. Newman
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 15 April 1982; accepted for publication 11 June 1982)

We define a GL(n,C) matrix-valued function G on a six-dimensional space of null paths in Minkowski space. Such paths are defined to begin at an arbitrary spacetime point x^a and end at future null infinity. The space of these paths can thus be parametrized by giving the point x^a and null direction. We show how knowledge of G can be used to obtain the GL(n,C) Yang–Mills connection at x^a. We also derive a single equation for G, involving characteristic data given on null infinity, which is equivalent to the currentless or vacuum Yang–Mills field equations. The self-dual (anti-self-dual) nonabelian fields and the general abelian cases are described as special examples.

PACS numbers: 11.15.−q, 02.30. + g

I. INTRODUCTION

It is the purpose of this paper to present a reformulation of classical GL(n,C) Yang–Mills theory.[1,2] The reformulation is in terms of a single matrix-valued function G on a six-dimensional subspace of the space of paths in Minkowski space M. This subspace is defined as the null paths beginning at each point (x^a) of M and ending at future null infinity $\mathscr{I}^{+}$[3]. A convenient parametrization of these paths is to give the Minkowski coordinates x^a of the starting point and the (complex) stereographic coordinates ($\zeta,\bar{\zeta}$) on S^2 which label the light cone generators of x^a. A path is thus labeled by ($x^a,\zeta,\bar{\zeta}$). The function $G(x^a,\zeta,\bar{\zeta})$ is defined by the parallel propagation (with a given connection) of n linearly independent fiber vectors, from x^a to null infinity along the ($\zeta,\bar{\zeta}$) generator.

We will show how, from knowledge of $G(x^a,\zeta,\bar{\zeta})$, the connection one-form γ_a at the point x^a can be obtained. Furthermore we will show how the vacuum Yang–Mills equations can be imposed on the G. This results in a rather complicated integro-differential equation for G which involves the characteristic initial data (essentially the radiation field) acting as the driving term. Two simple special cases are immediately obtainable; in the case of self-dual (or anti-self-dual) fields we obtain a simple derivation of the Sparling equation,[4,5] namely $\eth G = -GA$, while for abelian (Maxwell) theories we obtain the equation

$$\eth\bar{\partial}\ln G-\bar{\eth}A-\eth\bar{A},$$

where A and its conjugate $\bar{A}$ are the characteristic free data given on null infinity. The latter equation is equivalent to the vacuum Maxwell equations.

In Sec. II we will review some material from vector bundle theory[6] and introduce our notation. In Sec. III we will describe a light cone calculus (involving two types of covariant differentiation) which forms the foundation of much of our later calculations. In Sec. IV we will discuss $G(x^a,\zeta,\bar{\zeta})$ and its definition in terms of the parallel propagation of the n linearly independent vectors from x^a to infinity assuming some connection γ_a, while in Sec. V we solve the inverse problem, namely how the $\gamma_a(x^a)$ can be obtained from $G(x^a,\zeta,\bar{\zeta})$. We also give the necessary conditions on G such that the γ_a exists. Sections IV and V thus establishes a correspondence between connections and the functions G. Finally, in Sec. VI we impose the vacuum Yang–Mills equations on the connection and find the resulting equation for G. The two special cases of self-dual and abelian fields are discussed there.

In the conclusion we discuss possible generalizations and specialization of this work and in addition show the relationship of the work to the holonomy group.

We include four appendices. The first three give detailed proofs of some assertions made in the text while the last gives an alternate proof of our equation for G in the self-dual and abelian cases.

II. NOTATION AND CONVENTIONS

On Minkowski space M we will consider the trivial vector bundle B (each fiber being an n-complex-dimensional vector space), i.e., $B = M\times C^n$. The global vector fields e_A ($A=1,\ldots,n$) form a basis set as does

$$e'_A = G_A{}^B(x^a)e_B, \tag{2.1}$$

with $G_A{}^B(x^a)$ being GL(n,C) matrix-valued functions on M. The connection or parallel transfer of vectors is introduced by defining ∇_a by

$$\nabla_a e_A = \gamma^B_{Aa}e_B, \tag{2.2}$$

with

$$\gamma_A{}^B = \gamma^B_{Aa}dx^a, \tag{2.3}$$

being the connection (matrix-valued) one-form. One then defines the covariant derivative of an arbitrary vector $V = V^Ae_A$ by

$$\nabla_a V = (V^A{}_{,a} + V^B\gamma^A_{Ba})e_A, \tag{2.4}$$

with a comma denoting the ordinary derivatives with respect to the Minkowski coordinates x^a.

From (2.1) it follows that under a change in basis

[a)] This work has been supported by a grant from the NSF.

$$\gamma'^{B}_{A} = G^{C}_{A,a} G^{-1B}_{C} + G^{C}_{A} \gamma^{D}_{Ca} G^{-1B}_{D}, \tag{2.5a}$$

or in matrix notation,

$$\gamma'_a = G_{,a} G^{-1} + G \gamma_a G^{-1}. \tag{2.5b}$$

The curvature tensor or gauge field of this connection is defined by

$$F_{ab} = \gamma_{b,a} - \gamma_{a,b} - [\gamma_a, \gamma_b], \tag{2.6}$$

with $[\gamma_a,\gamma_b] = \gamma_a\gamma_b - \gamma_b\gamma_a$. From (2.5) one obtains

$$F'_{ab} = G F_{ab} G^{-1}. \tag{2.7}$$

The curvature tensor satisfies the Bianchi identities

$$F_{[a,b,c]} + F_{[ab}\gamma_{c]} - \gamma_{[c}F_{ab]} = 0, \tag{2.8a}$$

or

$$\nabla_{[c}F_{ab]} = 0. \tag{2.8b}$$

The dual field is defined by

$$F^{*}_{ab} = \tfrac{1}{2}\eta_{abcd}F^{cd}, \eta_{abcd} = (-g)^{1/2}\epsilon_{abcd}, \tag{2.9}$$

with ϵ_{abcd} the alternating symbol, with $\epsilon_{0123} = -1$.

We now write the Yang–Mills field equations as

$$g^{bc}\nabla_c F_{ab} = J_a, \tag{2.10}$$

where g^{bc} is the Minkowski metric and J_a is the current. We note that if $F^{*}_{ab} = \pm iF_{ab}$ (i.e., F_{ab} is self-dual or anti-self-dual), then (2.10) is automatically satisfied with $J_a = 0$.

We next recall that if ζ, $\bar\zeta$ are stereographic coordinates labeling the points on a sphere ($\zeta = e^{i\phi}\cot\frac{1}{2}\theta$), and f is a function of these coordinates, then we can define the differential operators [7] $\eth$ and $\bar\eth$ by

$$\eth f = 2P^{1-s}\frac{\partial}{\partial\bar\zeta}(P^s f), \tag{2.11}$$

and

$$\bar\eth f = 2P^{1+s}\frac{\partial}{\partial\zeta}(P^{-s} f),$$

where $P = \frac{1}{2}(1+\zeta\bar\zeta)$ and s is the spin weight of f. For future reference we note that

$$(\eth\bar\eth - \bar\eth\eth)f = -2sf. \tag{2.12}$$

As an important application, we consider the (complex) light cone at a point x^a in M. We can parametrize directions on this cone by ζ and $\bar\zeta$ and in a Minkowski coordinate system $\{x^a = (t,x,y,z)\}$ we then define

$$l^a(\zeta,\bar\zeta) = \frac{1}{2\sqrt{2}P}(1+\zeta\bar\zeta, \zeta+\bar\zeta, i(\bar\zeta-\zeta), -1+\zeta\bar\zeta), \tag{2.13a}$$

$$m^a = \eth l^a, \tag{2.13b}$$

$$\bar m^a = \bar\eth l^a, \tag{2.13c}$$

$$n^a = l^a + \eth\bar\eth l^a, \tag{2.13d}$$

where we assign to the vector l^a in (2.13a) the spin weight 0. It is then easily verified that

$$l^a n_a = -m^a\bar m_a = 1, \tag{2.14}$$

all other scalar products being zero. It is also not difficult to show that

$$\begin{aligned} &\eth m^a = \bar\eth\bar m^a = 0,\\ &\bar\eth m^a = \eth\bar m^a = (n^a - l^a),\\ &\eth n^a = -m^a,\\ &\bar\eth n^a = -\bar m^a. \end{aligned} \tag{2.15}$$

We note that as ζ moves over the complex plane, l^a and n^a range over all real null directions. This will be of importance in the next section.

For future use we introduce the matrices[5]

$$\begin{aligned} &\|\chi_{0A}{}^{B}\| = \chi_0 = F_{ab}l^a m^b,\\ &\|\chi_{1A}{}^{B}\| = \chi_1 = \tfrac{1}{2}F_{ab}(l^a n^b + \bar m^a m^b),\\ &\|\chi_{2A}{}^{B}\| = \chi_2 = F_{ab}\bar m^a m^b, \end{aligned} \tag{2.16}$$

as the anti-self-dual components of F_{ab}, and

$$\begin{aligned} &\|\bar\chi_{0A}{}^{B}\| = \bar\chi_0 = F_{ab}l^a \bar m^b,\\ &\|\bar\chi_{1A}{}^{B}\| = \bar\chi_1 = \tfrac{1}{2}F_{ab}(l^a n^b - \bar m^a m^b),\\ &\|\bar\chi_{2A}{}^{B}\| = \bar\chi_2 = F_{ab}m^a n^b, \end{aligned} \tag{2.17}$$

as the self-dual components of F_{ab}.

Finally we recall how to define a null polar coordinate system in Minkowski space based on the null cones emanating from a timelike geodesic L.

Beginning with Minkowski coordinates x^a, we introduce null polar coordinates $(u,r,\eta,\bar\eta)$ by

$$x^a = ut^a + r\hat l^a(\eta,\bar\eta), \tag{2.18}$$

with $\hat l^a$ having the same functional form as in (2.13a), and t^a being a constant vector tangent to L with $t^a t_a = 2$. u is (proportional to) the proper time along the world line L, ($x^a = ut^a$), r is a normalized affine length measured along the null rays leaving L, and η and $\bar\eta$ are complex stereographic coordinates labeling the direction of these null rays.

In this coordinate system, $\mathscr{I}^+$ has "equation" $r = \infty$ (i.e., $\mathscr{I}^+$ is the future completion of null geodesics leaving L). We may then take $(u, \eta, \bar\eta)$ as coordinates on $\mathscr{I}^+$ where $u = u_0$ (constant) will be the intersection with $\mathscr{I}^+$ of the light cone associated with the point $x^a = u_0 t^a$ on L, while $\eta = \eta_0$, $\bar\eta = \bar\eta_0$ (constants) will label a particular generator of $\mathscr{I}^+$. We will denote the point on $\mathscr{I}^+$ corresponding to $u = \infty$ by I^+.

It is important to note here that in the interior of M, η and $\bar\eta$ are to be distinguished from ζ and $\bar\zeta$, which were previously defined to label the (arbitrary) directions of null rays leaving *any* point x^a in M (not just those on the fixed L). We have however, chosen ζ and $\bar\zeta$ so that a null geodesic leaving x^a in the direction given by ζ, $\bar\zeta$ will intersect $\mathscr{I}^+$ in the generator given by $\eta = \zeta$, $\bar\eta = \bar\zeta$.

From the above we have two different types of null tetrads: $\{\hat l^a, \hat n^a, \hat m^a, \hat{\bar m}^a\}$ which is, at any point, a fixed tetrad that is associated with the null coordinates in (2.18), and $\{l^a, n^a, m^a, \bar m^a\}$ [from (2.13)] which, at any point, can sweep out all null directions. Both tetrads will play a central role in our later work.

We also note for later reference (and prove in Appendix B) the fact that if x^a is any point of M, then the intersection with $\mathscr{I}^+$ of the light cone of x^a is given by

$$u = u(\zeta,\bar{\zeta}) = x^a l_a(\zeta,\bar{\zeta}),$$
$$\eta = \zeta, \quad \bar{\eta} = \bar{\zeta}. \tag{2.19}$$

III. A LIGHT CONE CALCULUS

Let x^a be a fixed but arbitrary point of M, and consider the future null cone, $C(x^a)$ at x^a (with equal ease we could restrict our attention to the past light cone at x^a). Then it is not difficult to see from the last section that any point y^a on $C(x^a)$ can be written as

$$y^a = x^a + R^a, \tag{3.1a}$$

with

$$R^a = s l^a(\zeta,\bar{\zeta}), \tag{3.1b}$$

where ζ, $\bar{\zeta}$ label the (unique) generator of $C(x^a)$ passing through y^a, and s is the normalized affine distance from x^a to y^a along that generator.

Suppose now that $V_0^A(x^a)$ is a section of B. We wish to define a two-point function $V^A(x^a,R^a)$ on M as follows. Fix x^a and define $V^A(x^a,R^a)$ to be the vector obtained when $V_0^A(x^a)$ is parallelly propagated (using the connection given in Sec. II), from x^a to $y^a = x^a + R^a$ along the null geodesic connecting them.

We may then define two types of covariant derivatives of $V^A(x^a,R^a)$. Both differentiations will be with respect to changes in y^a, but in one case we will change y^a by fixing x^a and varying R^a via s, ζ, or $\bar{\zeta}$, whereas in the second case we will change y^a by fixing R^a (the difference $y^a - x^a$) and varying x^a. Thus in the first case we will obtain the "covariant change" in V^A between two neighboring points on the *same* light cone, whereas in the second case we will obtain the covariant change in V^A between two "comparable" points (same R^a) on *different* light cones. Using the connection which was introduced in Sec. II we write these two covariant derivatives as

$$D_a V = \frac{\partial V}{\partial y^a}\bigg|_{x^a = \text{const}} + V\gamma_a(y^a) = \frac{\partial V}{\partial R^a} + V\gamma_a(y^a), \tag{3.2}$$

$$\nabla_a V = \frac{\partial V}{\partial y^a}\bigg|_{R^a = \text{const}} + V\gamma_a(y^a) = \frac{\partial V}{\partial x^a} + V\gamma_a(y^a). \tag{3.3}$$

We will now list some of the properties of these differential operators which will be used in later sections. Using (3.2), (3.1b), (2.13), and the chain rule, one obtains

$$l^a D_a V = \frac{\partial V}{\partial s} + V\gamma_a(y^a) l^a, \tag{3.4a}$$

$$s m^a D_a V = \eth V + s V\gamma_a(y^a) m^a, \tag{3.4b}$$

and

$$s\bar{m}^a D_a V = \bar{\eth} V + s V\gamma_a(y^a)\bar{m}^a. \tag{3.4c}$$

Furthermore (see Appendix A),

$$D_a l^b = -\frac{1}{s} h_a^{\ b}, \tag{3.5a}$$

$$D_a n^b = \frac{1}{s} h_a^{\ b}, \tag{3.5b}$$

$$D_a m^b = -\frac{1}{s}(n^b - l^b) m_a, \tag{3.5c}$$

$$D_a \bar{m}^b = -\frac{1}{s}(n^b - l^b)\bar{m}_a, \tag{3.5d}$$

$$D_a f(s) = \frac{df}{ds}(n_a - l_a), \tag{3.5e}$$

and

$$D_c h_{\ b}^{a} = -\frac{1}{s}\left[(n^a - l^a)h_{cb} + (n_b - l_b)h_{\ c}^{a}\right], \tag{3.5f}$$

where f is a scalar function of s and $h_{ab} = m_a\bar{m}_b + \bar{m}_a m_b$ is the two-dimensional metric on S^2, the sphere of null directions. We also point out that since l^b is a function of ζ and $\bar{\zeta}$ only, we must have

$$\nabla_a l^b = 0. \tag{3.6}$$

Finally, for purposes of later calculations we write down the commutation relations between D_a and ∇_a. Directly from (3.2) and (3.3) we obtain

$$[\nabla_b,\nabla_a] V = V F_{ba}, \tag{3.7a}$$

$$[D_b,\nabla_a] V = V F_{ba}, \tag{3.7b}$$

$$[D_b,D_a] V = V F_{ba}, \tag{3.7c}$$

where we have also used (2.6). Then using (3.6) and (3.7c) we obtain

$$[V_b, l^a D_a] V = V F_{ba} l^a. \tag{3.8}$$

Lastly, using (3.5a) and (3.7b) we obtain

$$[D_b, l^a D_a] V = -\frac{1}{s} h_{\ b}^{a} D_a V + V F_{ba} l^a. \tag{3.9}$$

IV. A NEW VARIABLE FOR GAUGE THEORIES

Let x^a be a fixed but arbitrary point of M, and let $g_{(\lambda)}^A(x^a)$ ($\lambda = 1,\dots,n$) be a basis for the fiber at x^a. We will define the (nonsingular) matrix-valued function g at any point y^a on $C(x^a)$ to be the value obtained when $g_{(\lambda)}^A(x^a)$ is parallelly propagated from x^a to y^a along the unique generator of $C(x^a)$ joining them. Thus g can be viewed as a function of x^a and R^a (see Sec. III).

From (3.4a) and the geometrical definition of g given above, we see that g must satisfy

$$l^a D_a g = 0, \tag{4.1a}$$

or

$$\frac{dg}{ds} + g\gamma_a l^a = 0. \tag{4.1b}$$

We now define

$$G(x^a,\zeta,\bar{\zeta}) = \lim_{s\to\infty} g(x^a,\zeta,\bar{\zeta},s), \tag{4.2}$$

as our new variable for gauge theories. It follows from (4.1) that

$$G = O\exp\left\{-\int_0^\infty \gamma_a l^a ds\right\}, \tag{4.3}$$

where the O signifies an affine ordered exponential integral. In the abelian (Maxwell) theory, if we let $f = \ln g$, we obtain

$$F = \ln G = -\int_0^\infty \gamma_a l^a ds. \tag{4.4}$$

Note that we have tacitly assumed γ_a to be in a gauge (with at least $\gamma_a \sim 1/r$) such that the integrals (4.3) or (4.4) are defined. If the fields are asymptotically flat ($F_{ab} \sim 1/r$) such a gauge will exist. If we choose not to use such a gauge, the definition of G (or F) would be altered by including another term in (4.3) [or (4.4)] involving an integral along null infinity to I^+.

V. OBTAINING THE CONNECTION FROM G

In the last section we saw how a connection gives rise to the matrix-valued function $G(x^a,\zeta,\bar{\zeta})$ via (4.3). In this section we will show how to invert this equation in the sense that if we are given $G(x^a,\zeta,\bar{\zeta})$ (satisfying a condition to be discussed), then $\gamma_a(x^a)$ can be obtained such that (4.3) is satisfied. To do so we will apply ∇_b to both sides of the parallel propagation law (4.1a), and after using the commutation relations of Sec. III and simplifying, we will obtain a conserved quantity (i.e., one whose ordinary derivative with respect to s vanishes). It is from this "conservation law" that we will be able to derive the four components of γ_a in terms of derivatives of G.

This technique will be used again in the Sec. VI [applying, however, D_b or $D_b D_c$ instead of ∇_b to both sides of (4.1a)] in order to obtain the equation imposed on G by the currentless Yang–Mills equations on M.

To proceed, we see that from (4.1a), (3.6), and (3.8) it follows immediately that

$$l^a D_a \nabla_b g = g F_{ab} l^a. \tag{5.1}$$

Contracting both sides of (5.1) with l^b and using the antisymmetry of F_{ab} yields

$$l^b l^a D_a \nabla_b g = 0. \tag{5.2}$$

Then using (3.5a) and the scalar product relations (2.14), we can rewrite (5.2) as

$$l^a D_a (l^b \nabla_b g) = 0. \tag{5.3}$$

Next we multiply both sides of (5.3) on the right by g^{-1} and noting that (4.1) implies $l^a D_a g^{-1} = 0$, we obtain

$$l^a D_a (l^b \nabla_b g \cdot g^{-1}) = 0, \tag{5.4a}$$

or

$$\frac{d}{ds}(l^b \nabla_b g \cdot g^{-1}) = 0, \tag{5.4b}$$

where (5.4b) follows from the fact that $l^b \nabla_b g \cdot g^{-1}$ has now only scalar components, i.e., $\nabla_b g \cdot g^{-1} = \nabla_b g^A_{(\lambda)} g_A^{-1(\sigma)}$ and hence has no vector bundle indices.

We will now show how the conservation law (5.4) can be used to determine γ_a at the spacetime point x^a in terms of our fundamental variables.

To this end we see using (3.3) that, (5.4b) is equivalent to

$$(l^b g_{,b} g^{-1} + g\gamma_b(x^b + sl^b) l^b g^{-1})|_{s=0}$$
$$= \lim_{s \to \infty} (l^b g_{,b} g^{-1} + g\gamma_b(x^b + sl^b) l^b g^{-1}), \tag{5.5}$$

where we have explicitly written the argument of γ_a to emphasize that it is the connection evaluated at y^a (see Sec. III). Next under our assumption on the fall-off of γ_a (see Sec. IV, and Appendix B) and (4.2), we find that the right-hand side of (5.5) simply becomes $G_{,b} G^{-1} l^b$. We further assume, without loss of generality, that $g|_{s=0} = I$, the identity matrix. This can always be accomplished by a gauge transformation. Therefore, the left-hand side of (5.5) becomes $\gamma_b(x^b) l^b$ and we have[8]

$$l^a \gamma_a(x^b) = G_{,a} G^{-1} l^a, \tag{5.6}$$

i.e., one of the four components of γ_a at the spacetime point x^a. To obtain the other three we simply apply $\eth$, $\bar{\eth}$, and $\eth\bar{\eth}$ in turn to (5.6) [see (2.13)], obtaining

$$\gamma_b(x^b) m^b = l^b \eth(G_{,b} G^{-1}) + G_{,b} G^{-1} m^b, \tag{5.7a}$$

$$\gamma_b(x^b) \bar{m}^b = l^b \bar{\eth}(G_{,b} G^{-1}) + G_{,b} G^{-1} \bar{m}^b, \tag{5.7b}$$

$$\gamma_b(x^b) n^b = l^b \eth\bar{\eth}(G_{,b} G^{-1}) + \bar{m}^b \eth(G_{,b} G^{-1})$$
$$+ m^b \bar{\eth}(G_{,b} G^{-1}) + G_{,b} G^{-1} n^b, \tag{5.7c}$$

where we have used the fact that $\gamma_a(x^a)$ does not depend on ζ or $\bar{\zeta}$. Then using (2.14) it is clear that[9]

$$\gamma_a(x^a) = G_{,a} G^{-1} - \bar{h} m_a - h \bar{m}_a + k l_a, \tag{5.8}$$

where

$$h = l^b \eth(G_{,b} G^{-1}), \tag{5.9a}$$

$$\bar{h} = l^b \bar{\eth}(G_{,b} G^{-1}), \tag{5.9b}$$

and

$$k = m^b \bar{\eth}(G_{,b} G^{-1}) + \bar{m}^b \eth(G_{,b} G^{-1}) + l^b \eth\bar{\eth}(G_{,b} G^{-1}). \tag{5.9c}$$

It is worthwhile noting that substitution of an *arbitrary* $G(x^a,\zeta,\bar{\zeta})$ into the right-hand side of (5.8) will in general lead to a γ_a which depends on ζ and $\bar{\zeta}$ and thus does not define a connection at x^a. We would like to determine necessary conditions on $G(x^a,\zeta,\bar{\zeta})$ so that (5.8) does in fact define a connection on M. To do so we require, from (5.8), that

$$\eth\gamma_a = 0, \tag{5.10a}$$

and

$$\bar{\eth}\gamma_a = 0, \tag{5.10b}$$

as the necessary conditions on G. It turns out, however, that regularity (in ζ and $\bar{\zeta}$) together with one of the above equations implies the other, so we will concentrate on solving (5.10a), which can be written in terms of G as

$$\eth(G_{,a} G^{-1}) + (k - \eth\bar{h}) m_a - \eth h \bar{m}_a$$
$$+ (\eth k - h) l_a - h n_a = 0. \tag{5.11}$$

Contracting (5.11) in turn with l^a, $\bar{m}^a$, m^a, and n^a leads to

$$l^a \eth(G_{,a} G^{-1}) - h = 0, \tag{5.12a}$$

$$\bar{m}^a \eth(G_{,a} G^{-1}) - k + \eth\bar{h} = 0, \tag{5.12b}$$

$$m^a \eth(G_{,a} G^{-1}) + \eth h = 0, \tag{5.12c}$$

$$n^a \eth(G_{,a} G^{-1}) + \eth k - h = 0, \tag{5.12d}$$

respectively. But (5.12a) and (5.12b) are identities by virtue of (5.9), and it can be checked that applying $\bar{\eth}$ to both sides of (5.12c) leads to (5.12d). Thus (5.12c) is the only restriction on G such that it satisfy (5.10a). After a short calculation one finds this restriction can be rewritten as

$$\eth^2(l^a G_{,a} G^{-1}) = 0, \tag{5.13}$$

or

$$G_{,a}G^{-1}l^a = f_a(x^a)l^a. \tag{5.14}$$

Thus (5.6) is both a necessary and sufficient condition on G in order for it to define a connection $\gamma_a(x^a)$ on M.

VI. THE FIELD EQUATIONS

In the previous section we saw that knowledge of the correctly chosen function $G(x^a,\zeta,\bar{\zeta})$ can be used to determine a connection. In this section we will derive an equation on G such that the associated connection (5.8) and field (2.6) will satisfy the vacuum Yang–Mills equations.

To study the self-dual case we will apply sm^bD_b to both sides of (4.1a), while to study the general case we must apply $s^2g^{bc}D_bD_c$ to both sides of (4.1a). In the latter case, the current J_a (see Sec. II) will appear, and then be set to zero (corresponding to the vacuum Yang–Mills equations). In the self-dual and abelian cases we will obtain a conservation law (in the same sense as in Sec. V) while in the general (non-self-dual) case we will obtain a complicated integro-differential equation for G.

We first consider the case of self-dual fields. From (3.9) and (4.1), it follows that

$$l^aD_aD_bg = \frac{1}{s}h^{\,a}_{\,b}D_ag + gF_{ab}l^a. \tag{6.1}$$

Now we multiply both sides of (6.1) by sm^b, and use (2.14) to obtain

$$sm^bl^aD_aD_bg = -m^aD_ag + sgF_{ab}l^am^b. \tag{6.2}$$

Next using (3.5c) and (3.5e) we have

$$l^aD_a(sm^bD_bg) = sF_{ab}l^am^b. \tag{6.3}$$

If we now impose the field equations in the form that F_{ab} is self-dual (see Sec. II), then the right-hand side of (6.3) vanishes, and multiplying both sides on the right by g^{-1} we obtain the conservation law

$$\frac{d}{ds}(sm^bD_bg\cdot g^{-1}) = 0. \tag{6.4}$$

Using (3.4b), we can write (6.4) as

$$\begin{aligned}(\eth g\cdot g^{-1} &+ sg\gamma_a(x^a+sl^a)g^{-1}m^a)|_{s=0}\\ &= \lim_{s\to\infty}(\eth g\cdot g^{-1}+sg\gamma_a(x^a+sl^a)g^{-1}m^a).\end{aligned} \tag{6.5}$$

Clearly the left-hand side of (6.5) is zero from $g|_{s=0} = I$. Hence using (4.2), we obtain

$$\eth G\cdot G^{-1} = -G\lim_{s\to\infty}(s\gamma_am^a)G^{-1}, \tag{6.6}$$

or (see Appendix B, for a more detailed proof)

$$\eth G = -GA, \tag{6.7a}$$

$$A = \lim_{s\to\infty}s\gamma_am^a, \tag{6.7b}$$

where A (the data) is a spin weight $+1$ matrix-valued function on $\mathscr{I}^+$ of ζ, $\bar{\zeta}$, and $u = x^al_a$ [see (2.19)]. It is easily verified that a solution to (6.7) also satisfies (5.13). Thus (6.7) is equivalent to the Yang–Mills field equations on M in the self-dual case. Eq. (6.7) is referred to as the Sparling[4] equation, and was first derived in the context of twistor theory.

We now consider the general Yang–Mills case. The abelian (or Maxwell) theory will appear as a special case. To this end we apply D_c to both sides of (6.1), and using (3.9) (with $V = D_bg$), we obtain after some simiplification

$$\begin{aligned}l^aD_aD_cD_bg = &-2\Big(D_{(b}gF_{c)a}l^a - \frac{1}{s}h^{\,a}_{(c}D_{b)}D_ag\\ &+\frac{1}{s^2}(n_{(c}-l_c)h^{\,a}_{\,b)}D_ag\Big)\\ &-\frac{1}{s^2}n^ah_{cb}D_ag + (gD_cF_{ab})l^a,\end{aligned} \tag{6.8}$$

where we have also used (3.5) and (4.1). If we now multiply both sides of (6.8) by s^2g^{cb} and use (3.5) again, we obtain

$$\begin{aligned}l^aD_a(s^2g^{cb}D_cD_bg) &- 2s(n^cl^b+l^cn^b)D_cD_bg\\ &= 2s^2g^{cb}D_bg\cdot F_{ac}l^a + 2n^aD_ag + s^2gJ_al^a.\end{aligned} \tag{6.9}$$

We have used the fact that $g^{cb} = 2l^{(c}n^{b)} - h^{cb}$, $h^{\,b}_{\,b} = -2$, and the Yang–Mills equations $g^{cb}D_cF_{ab} = J_a$.

Then (3.5) and (4.1) imply

$$\begin{aligned}\frac{d}{ds}&(s^2g^{cb}D_cD_bg\cdot g^{-1} - 2sn^bD_bg\cdot g^{-1})\\ &= -2s^2h^{cb}D_cg\cdot g^{-1}F'_{ac}l^a + s^2J'_al^a,\end{aligned} \tag{6.10}$$

where $F'_{ac} = gF_{ac}g^{-1}$, and $J'_a = gJ_ag^{-1}$. If we define

$$H = sm^bD_bg\cdot g^{-1}, \tag{6.11a}$$

and

$$\bar{H} = s\bar{m}^bD_bg\cdot g^{-1}, \tag{6.11b}$$

then it easily follows from (6.3) that

$$\frac{dH}{ds} = sF'_{ab}l^am^b, \tag{6.12a}$$

and

$$\frac{d\bar{H}}{ds} = sF'_{ab}l^a\bar{m}^b. \tag{6.12b}$$

Substituting (6.11) and (6.12) into (6.10) and imposing the source-free Yang–Mills equations (viz., $J_a = 0$), we obtain

$$\begin{aligned}\frac{d}{ds}&(s^2g^{bc}D_cD_bg\cdot g^{-1} - 2sn^bD_bg\cdot g^{-1})\\ &= -2\Big(H\frac{d\bar{H}}{ds} + \bar{H}\frac{dH}{ds}\Big).\end{aligned} \tag{6.13}$$

We can rewrite (6.13) as

$$\begin{aligned}\frac{d}{ds}&(s^2g^{bc}D_c(D_bg\cdot g^{-1}) - H\bar{H} - \bar{H}H - 2sn^bD_bg\cdot g^{-1})\\ &= -2\Big(H\frac{d\bar{H}}{ds} + \bar{H}\frac{dH}{ds}\Big),\end{aligned} \tag{6.14}$$

and then as

$$\begin{aligned}\frac{d}{ds}&(s^2F'_{bc}l^bn^c - 2sm^{(b}D_b(s\bar{m}^{c)}D_cg\cdot g^{-1}))\\ &= \Big[\frac{dH}{ds},\bar{H}\Big] + \Big[\frac{d\bar{H}}{ds},H\Big],\end{aligned} \tag{6.15}$$

where we have used (3.5) and the contraction of (4.1) with n^b. Finally, using (3.4b) and (3.4c), we have

$$\frac{d}{ds}(\eth\overline{H} + \eth\overline{H} - s^2F'_{ab}l^an^b) = \left[H, \frac{d\overline{H}}{ds}\right] + \left[\overline{H}, \frac{dH}{ds}\right]. \tag{6.16}$$

It is from (6.16) that we will obtain our equation on G equivalent to the currentless Yang–Mills field equations.

In the non-self-dual abelian case (since the right-hand side of (6.16) vanishes), Eq. (6.16) becomes a conservation law, the consequences of which we will investigate shortly. We wish to first note (with some disappointment), that if the operation between the two commutators on the right hand side of (6.16) had been subtraction instead of addition, then we would have obtained a conservation law even in the general Yang–Mills case since

$$\frac{d}{ds}[H,\overline{H}] = \left[H, \frac{d\overline{H}}{ds}\right] - \left[\overline{H}\frac{dH}{ds}\right].$$

This not being the case, we did investigate the possibility that the right-hand side of (6.16) might still be a total derivative of some other quantity with respect to s. Unfortunately we had no success.

On the other hand, since the quantity inside the total derivative on the left-hand side of (6.16) vanishes at the space-time point x^a (i.e., at $s = 0$), we will still be able to obtain a single integro-differential equation for G which will involve characteristic data given on $\mathscr{I}^+$ and will be equivalent to the Yang–Mills equations on M. To do so we first note that (6.16) can be rewritten as

$$\lim_{s\to\infty}(\eth\overline{H} + \bar{\eth}H - s^2F'_{ab}l^an^b) = \int_0^\infty\left(\left[H, \frac{d\overline{H}}{ds}\right] + \left[\overline{H}, \frac{dH}{ds}\right]\right)ds. \tag{6.17}$$

We will now evaluate the limit on the left-hand side of (6.17). We have [see (6.11), (3.4) and Appendix B] that

$$\lim_{s\to\infty}\eth\overline{H} = \eth(\bar{\eth}G\cdot G^{-1} + G\overline{A}G^{-1}), \tag{6.18}$$

$$\lim_{s\to\infty}\bar{\eth}H = \bar{\eth}(\eth G\cdot G^{-1} + GAG^{-1}), \tag{6.19}$$

where A, and the conjugate $\overline{A}$, have previously been defined. We can therefore rewrite (6.17) as

$$\eth(\bar{\eth}G\cdot G^{-1} + G\overline{A}G^{-1}) + \bar{\eth}(\eth G\cdot G^{-1} + GAG^{-1}) - \lim_{s\to\infty}s^2GF_{ab}l^an^bG^{-1} = \int_0^\infty\left(\left[H, \frac{d\overline{H}}{ds}\right] + \left[\overline{H}, \frac{dH}{ds}\right]\right)ds. \tag{6.20}$$

A sraightforward calculation then yields

$$\eth(G^{-1}\bar{\eth}G) + \bar{\eth}(G^{-1}\eth G) = -\eth\overline{A} - \bar{\eth}A + \lim_{s\to\infty}s^2F_{ab}l^an^b + [A,G^{-1}\bar{\eth}G] + [\overline{A},G^{-1}\eth G] + G^{-1}\left(\int_0^\infty\left(\left[H, \frac{d\overline{H}}{ds}\right] + \left[\overline{H}, \frac{dH}{ds}\right]\right)ds\right)G. \tag{6.21}$$

To further simplify (6.21), we have (see Appendix B) that

$$\lim_{s\to\infty}s^2F_{ab}l^an^b = -\eth\overline{A} - \bar{\eth}A - \int_{-\infty}^u([\dot{A},\overline{A}] + [\dot{\overline{A}},A])\,du'. \tag{6.22}$$

Thus we have for our final equation for G

$$\eth(G^{-1}\bar{\eth}G) + \bar{\eth}(G^{-1}\eth G) = -2(\eth\overline{A} + \bar{\eth}A) + [A,G^{-1}\bar{\eth}G] + [\overline{A},G^{-1}\eth G] + G^{-1}\left(\int_0^\infty\left(\left[H, \frac{d\overline{H}}{ds}\right] + \left[\overline{H}, \frac{dH}{ds}\right]\right)ds\right)G - \int_{-\infty}^u([\dot{A},\overline{A}] + [\dot{\overline{A}},A])\,du'. \tag{6.23}$$

We emphasize here that Eq.(6.23) recasts the full currentless Yang–Mills field equations on M in terms of a *single* equation for a new scalar variable $G(x^a,\zeta,\bar{\zeta})$ (and hence on a six-dimensional space of null paths in M; see Sec. I) involving characteristic data given on $\mathscr{I}^+$. This completely reformulates Yang–Mills theory in terms of the G; given $A(u,\zeta,\bar{\zeta})$ and its conjugate $\overline{A}(u,\zeta,\bar{\zeta})$ as arbitrary spin 1 and -1 functions on $\mathscr{I}^+$, a solution to (6.23) could be substituted into (5.8) to produce a connection γ_a which in turn would lead to a field via (2.6). Although the question of whether (6.23) implies (5.13) remains a formidable one, we nevertheless conjecture it to be the case and in fact present a proof in Appendix D that in the self-dual and abelian cases [which can be derived directly from (6.23)], we do in fact automatically satisfy (5.13) with a solution to (6.23).

In the Maxwell case (since all commutators vanish), Eq. (6.23) becomes [using (4.4)]

$$\eth\bar{\eth}F = -(\eth\overline{A} + \bar{\eth}A), \tag{6.24}$$

where $\eth\bar{\eth}$ is the two-dimensional Laplacian[10] and one can consider Eq. (6.24) as equivalent to the vacuum Maxwell equations with A and $\overline{A}$ as characteristic initial data. Equation (6.24) can be considered as a reformulation of the Kirchoff integral formulation of Maxwell fields.[11] We note that Eq. (6.24) could also be derived directly from the conservation law which one obtains by setting the commutators to zero on the right-hand side of (6.16). Again we point out that unfortunately no such conservation law could be found for the general case. This appears to be a manifestation of the non-Huygens nature of the propagation of the full Yang–Mills gauge fields.

VII. CONCLUSIONS

We have shown here how the source free Yang–Mills theory (with certain weak asymptotic conditions) may be expressed in terms of a single GL(n,C) matrix-valued function $G(x^a,\zeta,\bar{\zeta})$ on a six-dimensional space of null paths in Minkowski space. The G had been defined by parallel propagation of a basis set in the fiber over the point x^a, along the null path labelled by ζ, $\bar{\zeta}$ to future null infinity. It is easy to see that this construction of G is intimately connected with elements of the holonomy group of the point I^+ (future timelike infinity), namely

$$H(I^+,\text{path}) = G^{-1}(x^a,\zeta_1,\bar{\zeta}_1)G(x^a,\zeta_0,\bar{\zeta}_0), \tag{7.1}$$

with H an element of the holonomy group and the path given as follows: move down $\mathscr{I}^+$ from I^+ along the generator (ζ_0, $\bar{\zeta}_0$) until a null generator from x^a is met, this generator is then followed to x^a; the path is then closed by going from x^a along the $(\zeta_1, \bar\zeta_1)$ generator to $\mathscr{I}^+$, then to I^+. Since $\gamma_a n^a = 0$ on $\mathscr{I}^+$ the parallel propagation along the $\mathscr{I}^+$ generators yields the identity transformation, thus establishing (7.1). It is easily seen that knowledge of $H(I^+,\text{path})$, with the space of paths $(R^4 \times S^2 \times S^2)$ described above, yields knowledge of the $G(x^a, \zeta, \bar\zeta)$. By taking $\zeta = \zeta_0 = \zeta_1 + d\zeta$, $\bar\zeta = \bar\zeta_0 = \bar\zeta_1$, Eq. (5.1) becomes

$$G^{-1}\left(G + \eth G \frac{d\zeta}{1+\zeta\zeta}\right) = H \equiv I - \frac{h d\zeta}{1+\zeta\bar\zeta}, \qquad (7.2)$$

or

$$\eth G = -Gh.$$

Thus, the G can be determined from the infinitesmal h. Note that for the self-dual fields, $h(x^a, \zeta, \bar\zeta) = A(u = x^a l_a, \zeta, \bar\zeta)$. We have not yet investigated what equations the vacuum Yang–Mills equations would impose on $H(I^+, x^a, \zeta, \bar\zeta, \zeta_0, \bar\zeta_0)$.

Another problem we are considering is how the material of this paper could be generalized to the case where the solutions are not global, as was assumed here. There are several approaches one can use, depending on the problem. (1) If the fields and connections are given on an Alexandrov neighborhood (the intersection of the future of the point x_P^a with the past of x_F^a) we can duplicate all our previous results except that we integrate only to the past cone of x_F^a and then along the cone to x_F^a, instead of going all the way to $\mathscr{I}^+$. Data must be given on the cone of x_F^a. (2) A virtually identical treatment is to perform an inversion about the point x_P^a, thus putting the future cone of x_P^a on $\mathscr{I}^+$. We could now use the methods of the paper to produce a solution in the neighborhood of I^+ and then by conformal invariance transform it back to the original Alexandrov neighborhood. (3) For the self-dual magnetic monopole (where for the global solution one has a nontrivial topology) we are trying to obtain the G or F in the abelian case by direct integration.

Another subject being studied is the integro-differential equation for G. We wish to see if any exact results can be obtained, e.g., direct information about a classical S-matrix taking data from $\mathscr{I}^-$ directly to $\mathscr{I}^+$ without integrating through the interior. This can easily be done for the self-dual fields. Further, we would like to understand the properties of a perturbation approach in terms of multiple scatterings of the field.

As a final comment we mention that work has begun (with some success) on the generalization of the ideas of this paper to general relativity and in particular to the asymptotically flat solutions of the Einstein vacuum equations.

APPENDIX A

In this Appendix we will use the spin-coefficient formalism[12] to prove

$$D_a l^b = \frac{-1}{s} h_a^{\ b}. \qquad (3.5a)$$

The proofs of (3.5b), (3.5c), and (3.5d) are similar, with (3.5f) following easily from (3.5c) and (3.5d). We will also use (2.18) to prove

$$D_a f(s) = \frac{df}{ds}(n_a - l_a). \qquad (3.5e)$$

First we recall the definitions of the spin coefficients for any null tetrad $\{l^a, n^a, m^a, \bar m^a\}$:

$$\begin{aligned}
\alpha &= \tfrac{1}{2}(l_{a;b} n^a \bar m^b - m_{a;b} \bar m^a \bar m^b), & \pi &= -n_{a;b} \bar m^a l^b,\\
\beta &= \tfrac{1}{2}(l_{a;b} n^a m^b - m_{a;b} \bar m^a m^b), & \rho &= l_{a;b} m^a \bar m^b,\\
\gamma &= \tfrac{1}{2}(l_{a;b} n^a n^b - m_{a;b} \bar m^a n^b), & \sigma &= l_{a;b} m^a m^b,\\
\epsilon &= \tfrac{1}{2}(l_{a;b} n^a l^b - m_{a;b} \bar m^a l^b), & \tau &= l_{a;b} m^a n^b,\\
\nu &= -n_{a;b} \bar m^a n^b, & \mu &= -n_{a;b} \bar m^a m^b,\\
\lambda &= -n_{a;b} \bar m^a \bar m^b, & \kappa &= l_{a;b} m^a l^b,
\end{aligned} \qquad (A1)$$

where ; denotes covariant differentiation.

We recall that $l^a, n^a, m^a, \bar m^a$ [as defined in (2.13)] with ζ, $\bar\zeta$ fixed are constant with respect to ∇_a defined in (3.3). Thus the spin coefficients associated with ∇_a and $l^a, n^a, m^a, \bar m^a$ all vanish.

On the other hand, fixing x^a and changing ζ and $\bar\zeta$ would correspond to choosing a variable tetrad, namely the one associated with the light cone of x^a. The covariant derivative D_a is here denoted by a semicolon. The spin coefficients associated with the light cone tetrad system become[13]

$$\rho = \mu = -\frac{1}{r} \qquad (A2)$$

$$\bar\alpha + \beta = \tau = \sigma = \epsilon = \pi = \lambda = \kappa = \gamma = \nu = 0.$$

Thus we obtain (using $r = s$)

$$D_a l_b = -\frac{1}{s}(m_a \bar m_b + \bar m_a m_b) \equiv -\frac{1}{s} h_{ab}. \qquad (A3)$$

The other equations (3.5b), (3.5c), and (3.5d), follow in a similar fashion.

To prove (3.5e), we note that if L is a timelike geodesic (with tangent vector v^a such that $v^a v_a = 2$) through the arbitrary point x^a, then any point y^a can be written as

$$y^a = x^a + w v^a + s l^a(\zeta, \bar\zeta), \qquad (A4)$$

where w is proportional to the proper time along L. The geometrical intepretation of (A4) is that we have set up a null polar coordinate system $(w, s, \zeta, \bar\zeta)$ [see (2.18)] with origin at the arbitrary point x^a.

If we apply $\partial / \partial y^b$ to both sides of (A4) we obtain

$$\delta_b^a = w_{,b} v^a + s_{,b} l^a + s l^a{}_{,\zeta} \zeta_{,b} + s l^a{}_{,\bar\zeta} \bar\zeta_{,b} \qquad (A5)$$

or

$$\delta_b^a = w_{,b} v^a + s_{,b} l^a + \frac{s}{2P} m^a \zeta_{,b} + \frac{s}{2P} \bar m^a \bar\zeta_{,b}. \qquad (A6)$$

Contracting (A6) in turn with n_a and l_a and then subtracting yields

$$s_{,b} \equiv D_b s = n_b - l_b, \qquad (A7)$$

where we have used $v^a l_a = v^a n_a = 1$. Equation (3.5e) then follows from (A7) and the chain rule.

APPENDIX B

In Sec. VI we needed the evaluation of $\lim_{s \to \infty} s^2 F_{ab}(y^a) l^a n^b$, where $y^a = x^a + s l^a(\zeta, \bar\zeta)$. In other words, holding $x^a, \zeta, \bar\zeta$ fixed, we needed to know the limiting behavior of $s^2 F_{ab}(y^a) l^a n^b$ as the point y^a moved out to $\mathscr{I}^+$ in the $\zeta, \bar\zeta$ direction on the future light cone of x^a.

In this Appendix we will show how to evaluate this limit, by using already established knowledge of how the components of F_{ab} behave asymptotically in the null polar coordinate system defined in (2.18). To do so we will first fix x^a and obtain asymptotic relationships between the parameters s, ζ, $\bar{\zeta}$ and the null polar coordinates u, r, η, $\bar{\eta}$ of the point y^a. It will then be possible to find the asymptotic relatiohships between the tetrad $\{l^a, n^a, m^a, \bar{m}^a\}$ defined in (2.13) and the tetrad $\{\hat{l}^a, \hat{n}^a, \hat{m}^a, \hat{\bar{m}}^a\}$ defined by (2.18). In particular, we will be interested in how the bivector $l^{[a}n^{b]}$ can be expressed in terms of the hatted vectors. Thus we will have expressed $s^2F_{ab}l^an^b$ asymptotically in terms of the null polar coordinate system and associated tetrad, from which the evaluation of our limit will easily follow.

For ease of notation, we begin by introducing [compare with (2.10) and (2.17)]

$$\hat{\chi}_0 = F_{ab}\hat{l}^a\hat{m}^b, \qquad \hat{\bar{\chi}}_0 = F_{ab}\hat{l}^a\hat{\bar{m}}^b,$$
$$\hat{\chi}_1 = \tfrac{1}{2}F_{ab}(\hat{l}^a\hat{n}^b + \hat{\bar{m}}^a\hat{m}^b), \qquad \hat{\bar{\chi}}_1 = \tfrac{1}{2}F_{ab}(\hat{l}^a\hat{n}^b - \hat{\bar{m}}^a\hat{m}^b),$$
$$\hat{\chi}_2 = F_{ab}\hat{\bar{m}}^a\hat{n}^b, \qquad \hat{\bar{\chi}}_2 = F_{ab}\hat{m}^a\hat{n}^b,$$

$$\hat{\gamma}_{00'} = \gamma_a\hat{l}^a, \qquad \hat{\gamma}_{10'} = \gamma_a\hat{\bar{m}}^a,$$
$$\hat{\gamma}_{01'} = \gamma_a\hat{m}^a, \qquad \hat{\gamma}_{11'} = \gamma_a\hat{n}^a. \tag{B1}$$

Then it is known for asymptotically vanishing fields that[9]

$$\hat{\chi}_0 = \frac{\hat{\chi}_0^0}{r^3} + O(r^{-4}),$$
$$\hat{\chi}_1 = \frac{\hat{\chi}_1^0}{r^2} + O(r^{-3}),$$
$$\hat{\chi}_2 = \frac{\hat{\chi}_2^0}{r} + O(r^{-2}), \tag{B2}$$

(with similar relationships holding for the conjugate quantities), where $\hat{\chi}_0^0$,$\hat{\chi}_1^0$, and $\hat{\chi}_2^0$ are functions of u, η, and $\bar{\eta}$, satisfying

$$\hat{\chi}_2^0 = -\frac{\partial}{\partial u}\bar{\gamma}^0, \tag{B3a}$$

$$\frac{\partial}{\partial u}\hat{\chi}_1^0 = \eth\,\hat{\chi}_2^0 + [\hat{\chi}_2^0,\gamma^0], \tag{B3b}$$

where

$$\gamma^0 = \lim_{r\to\infty} r\hat{\gamma}_{01'} \equiv A\,(x^a\hat{l}_a,\eta,\bar{\eta}), \tag{B4a}$$

$$\bar{\gamma}_0 = \lim_{r\to\infty} r\hat{\gamma}_{10'} \equiv \bar{A}\,(x^a\hat{l}_a,\eta,\bar{\eta}). \tag{B4b}$$

Using this information we will show that

$$\lim_{s\to\infty} s^2F_{ab}l^an^b = -(\eth\bar{A} + \bar{\eth}A) - \int_{-\infty}^{u}([\dot{A},\bar{A}] + [\dot{\bar{A}},A])\,du'. \tag{B5}$$

In order to do so we will first find relationships between the coordinates $(u, r, \eta, \bar{\eta})$ and the parameters $(s, \zeta, \bar{\zeta})$ associated with x^a. Since any point in M can be written using coordinates (2.16) or (3.1a) we must have

$$x^a + sl^a(\zeta,\bar{\zeta}) = ut^a + r\hat{l}^a(\eta,\bar{\eta}). \tag{B6}$$

In Eq. (B6) we regard the s, ζ, $\bar{\zeta}$ as known (with x^a fixed), and we wish to solve asymptotically for the u, r, η, $\bar{\eta}$ in terms of them. Geometrially, we are finding the asymptotic intersection of the light cone of x^a with the system of light cones leaving the fixed timelike geodesic $\underset{\wedge}{L}$ [see Eq. (2.18)].

Contracting (B6) with $\hat{l}_a$, $\hat{m}_a$, $\hat{\bar{m}}_a$, and $\hat{n}_a$, yields

$$x^a\hat{l}_a + sl^a\hat{l}_a = u, \tag{B7a}$$
$$x^a\hat{m}_a + sl^a\hat{m}_a = 0, \tag{B7b}$$
$$x^a\hat{\bar{m}}_a + sl^a\hat{\bar{m}}_a = 0, \tag{B7c}$$
$$x^a\hat{n}_a + sl^a\hat{n}_a = u + r, \tag{B7d}$$

while contracting with l_a, m_a, $\bar{m}_a$, and n_a, yields

$$l \equiv x^al_a = u + r\hat{l}^al_a, \tag{B8a}$$
$$m \equiv x^am_a = r\hat{l}^am_a, \tag{B8b}$$
$$\bar{m} \equiv x^a\bar{m}_a = r\hat{l}^a\bar{m}_a, \tag{B8c}$$
$$n + s \equiv x^an_a + s = u + r\hat{l}^an_a. \tag{B8d}$$

It is not difficult to show simply from the definitions [Eq. (2.13)] of the various vectors involved that

$$n^a\hat{n}_a = l^a\hat{l}_a = \left(\frac{1}{4P\hat{P}}\right)(\zeta - \eta)(\bar{\zeta} - \bar{\eta}), \tag{B9a}$$

$$l^a\hat{m}_a = \left(\frac{1}{4P\hat{P}}\right)(\bar{\eta} - \bar{\zeta})(1 + \zeta\bar{\eta}),$$

$$l^a\hat{\bar{m}}_a = \left(\frac{1}{4P\hat{P}}\right)(\eta - \zeta)(1 + \bar{\zeta}\eta), \tag{B9b}$$

$$n^a\hat{m}_a = \left(\frac{1}{4P\hat{P}}\right)(\bar{\zeta} - \bar{\eta})(1 + \zeta\bar{\eta}),$$

$$n^a\hat{\bar{m}}_a = \left(\frac{1}{4P\hat{P}}\right)(\zeta - \eta)(1 + \bar{\zeta}\eta), \tag{B9c}$$

$$m^a\hat{m}_a = \left(\frac{1}{4P\hat{P}}\right)(\bar{\eta} - \bar{\zeta})^2, \quad \bar{m}^a\hat{\bar{m}}_a = \left(\frac{1}{4P\hat{P}}\right)(\eta - \zeta)^2, \tag{B9d}$$

$$l^a\hat{n}_a = \left(\frac{1}{4P\hat{P}}\right)(1 + \zeta\bar{\eta})(1 + \bar{\zeta}\eta),$$

$$m^a\hat{\bar{m}}_a = -\left(\frac{1}{4P\hat{P}}\right)(1 + \bar{\zeta}\eta)^2, \tag{B9e}$$

where $\hat{P} = \frac{1}{2}(1 + \eta\bar{\eta})$. Note that interchanging the role of hatted and unhatted vectors on the left hand sides of (B9) simply interchanges the roles of ζ and η, $\bar{\zeta}$ and $\bar{\eta}$ on the right-hand sides.

We will now regard s, ζ, $\bar{\zeta}$ as known and solve for u, r, η, $\bar{\eta}$ from (B7). Since $x^a\hat{m}_a$ and $x^a\hat{\bar{m}}_a$ both remain finite as $s\to\infty$, we must have from (B7b) and (B7c) that $l^a\hat{m}_a$ and $l^a\hat{\bar{m}}_a$ both tend to zero as $s\to\infty$. Thus from (B9b) we must have $\eta = \zeta$ and $\bar{\eta} = \bar{\zeta}$ as $s\to\infty$. (We eliminate the possibility that $\eta = -1/\bar{\zeta}$, $\bar{\eta} = -1/\zeta$ since this would ultimately contradict $r \geqslant 0$). It then follows from (B9a) that $l^a\hat{l}_a$ has a double zero as $s\to\infty$ and hence $\hat{l}_a \to l_a$ and $sl^a\hat{l}_a \to 0$ as $s\to\infty$. Thus (B7a) leads to

$$l \equiv x^al_a = u, \tag{B10}$$

in the limit $s\to\infty$. If we now let $s\to\infty$ in (B7d) and use (B10) with the first of (B9e) we obtain

$$n + \lim_{s\to\infty} s = l + \lim_{s\to\infty} r. \tag{B11}$$

Since n and l both remain finite (and fixed) as $s\to\infty$, we must have $r\to\infty$ as $s\to\infty$. This together with (B10) proves (2.19) and justifies its use in (6.7).

To proceed, we now work out the $1/r$ term in the asymptotic relationship between η, $\bar{\eta}$ and ζ, $\bar{\zeta}$ using (B8b) and (B8c). To this end we write

$$\begin{aligned} &\eta = \zeta + a/r + O(r^{-2}), \\ &\bar{\eta} + \bar{\zeta} + \bar{a}/r + O(r^{-2}), \end{aligned} \tag{B12}$$

where a and $\bar{a}$ are to be determined.

Substituting (B12) into (B8b) and using the analog of (B9b) [see the comment following (B9)], we find

$$\begin{aligned} &a = -2P\bar{m}, \\ &\bar{a} = -2Pm. \end{aligned} \tag{B13}$$

To obtain the asymptotic relationship between s and r, we subtract (B8a) from (B8d), obtaining

$$(n - l) + s = r\hat{l}^a(n_a - l_a). \tag{B14}$$

But substituting (B12) into the analog of (B9), it can easily be shown that

$$\hat{l}^a n_a = 1 + O(r^{-2}), \tag{B15a}$$

$$\hat{l}^a l_a = O(r^{-2}). \tag{B15b}$$

Therefore, (B14) yields

$$s = r + (l - n) + O(r^{-1}). \tag{B16}$$

In summary, we have for large r

$$u = x^a l_a(\zeta,\bar{\zeta}) + O(r^{-1}) \tag{B17}$$

and

$$\zeta = \eta + 2P\bar{m}/r + O(r^{-2}), \tag{B18a}$$

$$\bar{\zeta} = \bar{\eta} + 2Pm/r + O(r^{-2}), \tag{B18b}$$

$$s = r + (l - n) + O(r^{-1}). \tag{B18c}$$

Next, we express l^a and n^a in terms of the hatted vectors, writing

$$l^a = \omega\hat{l}^a + \theta\hat{m}^a + \bar{\theta}\hat{\bar{m}}^a + \xi n^a, \tag{B19a}$$

$$n^b = \delta\hat{l}^b + \psi\hat{m}^b + \bar{\psi}\hat{\bar{m}}^b + \phi\hat{n}^b, \tag{B19b}$$

where $\omega = l^c\hat{n}_c$, etc.

It then follows directly from (B9), (B18) that

$$\theta = \bar{m}/r + O(r^{-2}), \quad \bar{\theta} = m/r + O(r^{-2}), \tag{B20}$$

$$\psi = \bar{\psi} = O(r^{-1}). \tag{B21}$$

We now express $F_{ab}l^a n^b$ in terms of the hatted vectors, i.e.,

$$\begin{aligned} F_{ab}l^a n^b = F_{ab}\{&(\psi\omega - \theta\delta)\hat{l}^a\hat{m}^b + (\bar{\psi}\omega - \bar{\theta}\delta)\hat{l}^a\hat{\bar{m}}^b \\ &+ (\theta\bar{\psi} - \bar{\theta}\psi)\hat{m}^a\hat{\bar{m}}^b + (\theta\phi - \xi\psi)\hat{m}^a\hat{n}^b \\ &+ (\bar{\theta}\phi - \xi\bar{\psi})\hat{\bar{m}}^a\hat{n}^b + (\phi\omega - \xi\delta)\hat{l}^a\hat{n}^b\}. \end{aligned} \tag{B22}$$

Then using (B2) together with (B15), (B18c), (B20), and (B21), we obtain

$$\lim_{s\to\infty} s^2 F_{ab}l^a n^b = (\hat{\chi}_1^0 + \hat{\bar{\chi}}_1^0) + \overline{m\hat{\chi}}{}_2^0 + m\hat{\chi}_2^0. \tag{B23}$$

It follows from (B3) and (B4) that

$$\begin{aligned} &m\hat{\chi}_2^0 = -m\dot{\bar{A}}, \\ &\bar{m}\hat{\bar{\chi}}_2^0 = -\bar{m}\dot{A}. \end{aligned} \tag{B24}$$

Substituting (B3a) into (B3b) and integrating with respect to u yields

$$\hat{\chi}_1^0 = -\eth'\bar{A} - \int_{-\infty}^{u} [\dot{\bar{A}},A]\,du' + \lim_{u\to-\infty}(\eth\bar{A} + \hat{\chi}_1^0), \tag{B25}$$

and similarly

$$\hat{\bar{\chi}}_1^0 = \bar{\eth}'A - \int_{-\infty}^{u} [\dot{A},\bar{A}]\,du' + \lim_{u\to-\infty}(\bar{\eth}A + \hat{\bar{\chi}}_1^0), \tag{B26}$$

wheresincenow$A\,(u = x^a l_a(\zeta,\bar{\zeta}),\zeta,\bar{\zeta}),\bar{A} = \bar{A}\,(u = x^a l_a(\zeta,\bar{\zeta}),\zeta,\bar{\zeta})$, we use $\eth'$ or $\bar{\eth}'$ to denote differentiation with respect to the *explicit* ζ or $\bar{\zeta}$ behavior, respectively.

By a gauge transformation we may set $\lim_{u\to-\infty}(\eth\bar{A} + \bar{\eth}A) = 0$. We will also by assumption set $\lim_{u\to-\infty}(\hat{\chi}_1^0 + \hat{\bar{\chi}}_1^0) = 0$. Then substituting (B24), (B25), and (B26) into (B23) yields (B4), where we have also used the facts that $\eth\bar{A} = m\dot{\bar{A}} + \eth'\bar{A}$ and $\bar{\eth}A = \bar{m}\dot{A} + \bar{\eth}'A$.

In the abelian case (B5) becomes

$$\eth\bar{\eth}F = -(\eth\bar{A} + \bar{\eth}A). \tag{B27}$$

APPENDIX C

In this appendix we will present proofs that Eq. (6.7a) (the Sparling equation) and Eq. (6.24) (its non-self-dual abelian generalization) both imply Eq. (5.13) and are therefore equivalent to the self-dual nonabelian and full Maxwell equations, respectively.

To prove that (6.7a) implies (5.13) we differentiate (6.7a) with respect to x^a obtaining

$$\eth G_{,a} = -(G_{,a}A + G\dot{A}m^a), \tag{C1}$$

from which it follows that

$$(\eth G_{,a})l^a G^{-1} = -G_{,a}l^a A G^{-1}, \tag{C2a}$$

$$(\eth G_{,a})m^a G^{-1} = -G_{,a}m^a A G^{-1}. \tag{C2b}$$

If we apply $\eth$ to (C2a) and use (6.7a) we obtain (after some simplification)

$$\eth^2(G_{,a}G^{-1}l^a) + \eth(-G_{,a}m^a G^{-1}) = 0. \tag{C3}$$

On the other hand, we have from (C2b) and (6.7a) that

$$\eth(G_{,a}m^a G^{-1}) = 0, \tag{C4}$$

thus proving (5.13).

To prove that (6.24) implies (5.13), we first rewrite (6.24) as

$$\eth\bar{\eth}F = -(m\dot{\bar{A}} + \eth'\bar{A} + \bar{m}\dot{A} + \bar{\eth}'A), \tag{C5}$$

where by $\eth'$ and $\bar{\eth}'$ we mean differentiation with respect to the explicit ζ or $\bar{\zeta}$ dependence, respectively, and $m = x^a m_a$, $\bar{m} = x^a\bar{m}_a$, $(\,\cdot \equiv \partial/\partial u)$. Differentiating (C5) with respect to x^a yields

$$\begin{aligned} \eth\bar{\eth}F_{,a} = -(&m_a\dot{\bar{A}} + l_a m\ddot{\bar{A}} + l_a\eth'\dot{\bar{A}} \\ &+ \bar{m}_a\dot{A} + l_a\bar{m}\ddot{A} + l_a\bar{\eth}'\dot{A}). \end{aligned} \tag{C6}$$

If we contract (C6) with l^a, m^a, $\bar{m}^a$, and n^a in turn, we obtain

$$l^a\eth\bar{\eth}F_{,a}=0, \tag{C7a}$$

$$m^a\eth\bar{\eth}F_{,a}=\dot{A}, \tag{C7b}$$

$$\bar{m}^a\eth\bar{\eth}F_{,a}=\dot{\bar{A}}, \tag{C7c}$$

$$n^a\eth\bar{\eth}F_{,a}=-(m\ddot{\bar{A}}+\eth'\dot{\bar{A}}+\bar{m}\ddot{A}+\bar{\eth}'\dot{A}). \tag{C7d}$$

By applying $\eth$ to (C7a) and using (2.12), we obtain (after some simplification)

$$(\bar{\eth}\eth^2)(l^aF_{,a})^2-2n^a\eth F_{,a}-\bar{m}^a\eth^2F_{,a}-m^a\eth\bar{\eth}F_{,a}=0. \tag{C8}$$

We will now show that $\bar{\eth}(\eth^2(l^aF_{,a}))=0$, from which (5.13) easily follows. Applying $\bar{\eth}$ to (C7b) and $\eth$ to (C7c) yields

$$m^a\bar{\eth}\eth\bar{\eth}F_{,a}+n^a\eth\bar{\eth}F_{,a}=\ddot{A}\bar{m}+\bar{\eth}'\dot{A}, \tag{C9a}$$

$$\bar{m}^a\eth^2\bar{\eth}F_{,a}+n^a\eth\bar{\eth}F_{,a}=\ddot{\bar{A}}m+\eth'\dot{\bar{A}}. \tag{C9b}$$

Adding (C9a) to (C9b), and substituting into (C7d) we obtain [using (C7a) and (2.12)]

$$\bar{\eth}(2n^a\eth F_{,a}+\bar{m}^a\eth^2F_{,a}+m^a\eth\bar{\eth}F_{,a})=0. \tag{C10}$$

Since the expression inside the parentheses is a regular spin weight $+1$ function, it must vanish. Substitution into (C8) yields

$$\bar{\eth}(\eth^2(l^aF_{,a}))=0, \tag{C11}$$

and (5.13) follows from (4.4) and the fact that $\eth^2(l^aF_{,a})$ is a regular spin weight $+2$ function.

APPENDIX D

In this Appendix we will give alternate proofs of the Sparling equation (6.7a) and its non-self-dual abelian generalization (6.24). To prove (6.7) we will parallelly propagate a vector in the fiber at x^a around a triangular region contained entirely in the anti-self-dual two-plane spanned by l^a and m^a, and then use the fact that the self-dual Yang–Mills equations are equivalent to the fact that $F_{ab}l^am^b$ vanishes on this two-plane. The proof of (6.24) will be somewhat more involved.

Consider then the "infinitesimal triangle" T in M (see Fig. 1) with vertices at $P_0\leftrightarrow x^a$, $P_1\leftrightarrow(x^a+sl^a(\zeta,\bar{\zeta}))$, and $P_2\leftrightarrow(x^a+sl^a(\zeta+d\zeta,\bar{\zeta}))$, where s is fixed but arbitrary. (Ultimately we want $s\to\infty$ so that P_1 and P_2 will be on $\mathscr{I}^+$.)

Obviously $l^a(\zeta,\bar{\zeta})$ is tangent to P_0P_1, and $-l^a(\zeta+d\zeta,\bar{\zeta})$ is tangent to P_2P_0. Note that $m^a(\zeta,\bar{\zeta})$ is tangent to P_1P_2 since $(x^a+sl^a(\zeta+d\zeta,\bar{\zeta}))-(x^a+sl^a(\zeta,\bar{\zeta}))=s(l^a(\zeta+d\zeta,\bar{\zeta})-l^a(\zeta,\bar{\zeta}))=sl^a_{,\zeta}(\zeta,\bar{\zeta})d\zeta=(s/2P)m^a(\zeta,\bar{\zeta})d\zeta$, where we have made use of (2.11). We may therefore take

$$d\lambda=\frac{s}{2P}d\zeta, \tag{D1}$$

as a parameter for the infinitesimal segment P_1P_2.

Parallel propagation of a vector v along P_0P_1 or P_2P_0 will then be given by

$$\frac{dv}{ds}+v\gamma_al^a=0, \tag{D2}$$

and parallel propagation along P_1P_2 is given by

$$\frac{dv}{d\lambda}+v\gamma_am^a=0. \tag{D3}$$

Since the curvature F_{ab} is self-dual, we have

$$F_{ab}l^am^b=0, \tag{D4}$$

from which it follows that any vector V^A in the fiber at P_0 remains unchanged when parallelly propagated around T. Therefore we must have that

$$I=g(x^a,\zeta,\bar{\zeta},s)\left(I-\frac{s\gamma_am^a}{2P}d\zeta\right)g^{-1}(x^a,\zeta+d\zeta,\bar{\zeta},s), \tag{D5}$$

where $g(x^a,\zeta,\bar{\zeta},s)$ parallelly propagates V^A from P_0 to P_1, $I-(s\gamma_am^a/2P)d\zeta$ $(=I+\delta g)$ parallelly propagates from P_1 to P_2, and $g^{-1}(x^a,\zeta+d\zeta,\bar{\zeta},s)$ parallelly propagates from P_2 back to P_0.

Taking the limit in (D5) as $s\to\infty$ and using $\lim_{s\to\infty}g(x^a,\zeta,\bar{\zeta},s)=G(x^a,\zeta,\bar{\zeta})$, $\lim_{s\to\infty}s\gamma_am^a=A(u,\zeta,\bar{\zeta})$, we obtain

$$I=G(x^a,\zeta,\bar{\zeta})\left(I-\frac{Ad\zeta}{2P}\right)G^{-1}(x^a,\zeta+d\zeta,\bar{\zeta}). \tag{D6}$$

Finally, from the definition of $\eth$, we obtain

$$\eth G=-GA. \tag{D7}$$

We now turn to the proof of (6.24). We begin by considering the figure in M (see Fig. 2) with vertices at P_0, P_1, P_2 (as already defined), $P_3\leftrightarrow(x^a+sl^a(\zeta+d\zeta,\bar{\zeta}+d\bar{\zeta}))$, and $P_4\leftrightarrow(x^a+sl^a(\zeta,\bar{\zeta}+d\bar{\zeta}))$. If we denote the surface of the figure by f, then by the field equations $g^{bc}\nabla_cF_{ab}=0$, and $g^{bc}\nabla_cF^*_{ab}=0$ we have

$$\int_f(F_{ab}\pm iF^*_{ab})dS^{ab}=0, \tag{D8}$$

where $F^+_{ab}\equiv\frac{1}{2}(F_{ab}-iF^*_{ab})$ is the self-dual part and $F^-_{ab}\equiv\frac{1}{2}(F_{ab}+iF^*_{ab})$ is the anti-self-dual part of F_{ab}. Since $l_{[a}\bar{m}_{b]}$ is self-dual and $l_{[a}m_{b]}$ is anti-self-dual, we can rewrite (D8) as (see Fig. 2)

$$\int_{S_{12}+S_{34}}F^-_{ab}l^{[a}m^{b]}\frac{s'}{(1+\zeta'\bar{\zeta})}ds'd\zeta'$$

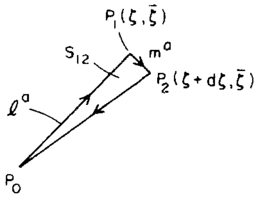


FIG. 1. Contour for derivation of the Sparling equation; P_1 is on generator $(\zeta,\bar{\zeta})$, P_2 on $(\zeta+d\zeta,\bar{\zeta})$.

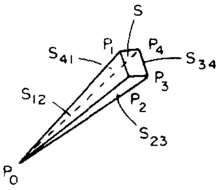


FIG. 2. Surface for derivation of generalized abelian Sparling equation; P_1 is on generator $(\zeta,\bar{\zeta})$, P_2 on $(\zeta+d\zeta,\bar{\zeta})$, P_3 on $(\zeta+d\zeta,\bar{\zeta}+d\bar{\zeta})$, P_4 on $(\zeta,\bar{\zeta}+d\bar{\zeta})$.

$$+ \int_S F^{-}_{ab} \overline{m}^{[a} m^{b]} \frac{s^2 d\zeta' d\overline{\zeta}'}{(1 + \zeta'\overline{\zeta}')^2} = 0, \tag{D9a}$$

$$\int_{S_{23} + S_{41}} F^{+}_{ab} l^{[a} \overline{m}^{b]} \frac{s'}{(1 + \zeta\overline{\zeta}')} ds' d\overline{\zeta}'$$

$$+ \int_S F^{+}_{ab} \overline{m}^{[a} m^{b]} \frac{s^2 d\zeta' d\overline{\zeta}'}{(1 + \zeta'\overline{\zeta}')^2} = 0. \tag{D9b}$$

Now since $l_{[a} n_{b]} - \overline{m}_{[a} m_{b]}$ is self-dual, and $l_{[a} n_{b]} + \overline{m}_{[a} m_{b]}$ is anti-self-dual, we can rewrite (D9) as

$$\int_{S_{12} + S_{34}} F_{ab} l^{[a} m^{b]} \frac{s'}{(1 + \zeta'\overline{\zeta})} ds' d\zeta'$$

$$+ \int_S \frac{1}{2} F_{ab} (l^{[a} n^{b]} + \overline{m}^{[a} m^{b]}) \frac{s^2 d\zeta' d\overline{\zeta}'}{(1 + \zeta'\overline{\zeta}')^2} = 0, \tag{D10a}$$

and

$$\int_{S_{23} + S_{41}} F_{ab} l^{[a} \overline{m}^{b]} \frac{s'}{(1 + \zeta\overline{\zeta}')} ds' d\overline{\zeta}'$$

$$+ \int_S \frac{1}{2} F_{ab} (l^{[a} n^{b]} - \overline{m}^{[a} m^{b]}) \frac{s^2 d\zeta' d\overline{\zeta}'}{(1 + \zeta'\overline{\zeta}')^2} = 0. \tag{D10b}$$

We now simplify (D10a). The calculation for (D10b) follows in a similar manner. Applying Stokes' theorem to the first integral on the left-hand side of (D10a), we can write

$$\int_{S_{12} + S_{34}} F_{ab} l^{[a} m^{b]} \frac{s'}{(1 + \zeta'\overline{\zeta})} ds' d\zeta'$$

$$= \int_{P_0}^{P_1} \gamma_a dx'^a + \int_{P_1}^{P_2} \gamma_a dx'^a + \int_{P_2}^{P_0} \gamma_a dx'^a + \int_{P_0}^{P_3} \gamma_a dx'^a$$

$$+ \int_{P_3}^{P_4} \gamma_a dx'^a + \int_{P_4}^{P_0} \gamma_a dx'^a. \tag{D11}$$

Taking the limit as $s \to \infty$ (so that S is now on $\mathscr{I}^+$) and substituting (D11) into (D10a), we can rewrite the latter using (4.4) as

$$- F(x^a,\zeta,\overline{\zeta}) + \frac{A(u,\zeta,\overline{\zeta}) d\zeta}{2P}$$

$$+ F(x^a,\zeta + d\zeta,\overline{\zeta}) + F(x^a,\zeta,\overline{\zeta} + d\overline{\zeta})$$

$$- \frac{A(u,\zeta,\overline{\zeta} + d\overline{\zeta}) d\zeta}{2P} - F(x^a,\zeta + d\zeta,\overline{\zeta} + d\overline{\zeta})$$

$$+ \lim_{s \to \infty} \frac{s^2}{4P^2} \chi_1 d\zeta d\overline{\zeta} = 0, \tag{D12}$$

where because of the infinitesimal nature of S, we have dropped the integral sign and have used Eq. (2.16).

This simplifies to

$$F_{,\zeta\overline{\zeta}}(x^a,\zeta,\overline{\zeta}) d\zeta d\overline{\zeta} + \frac{A_{,\overline{\zeta}}(u,\zeta,\overline{\zeta})}{2P} d\zeta d\overline{\zeta}$$

$$- \lim_{s \to \infty} \frac{s^2}{4P^2} \chi_1 d\zeta d\overline{\zeta} = 0. \tag{D13}$$

By a similar calculation, (D10b) becomes

$$F_{,\zeta\overline{\zeta}}(x^a,\zeta,\overline{\zeta}) d\zeta d\overline{\zeta} + \frac{\overline{A}_{,\zeta}(u,\zeta,\overline{\zeta})}{2P} d\zeta d\overline{\zeta}$$

$$- \lim_{s \to \infty} \frac{s^2}{4P^2} \overline{\chi}_1 d\zeta d\overline{\zeta} = 0. \tag{D14}$$

Adding (D13) to (D14) and simplifying, yields

$$8P^2 F_{,\zeta\overline{\zeta}} = -2PA_{,\overline{\zeta}} - 2P\overline{A}_{,\zeta} + \lim_{s \to \infty} s^2(\chi_1 + \overline{\chi}_1). \tag{D15}$$

Finally, using (2.10) and the abelian version of (B5), we can rewrite (D15) as

$$ð\overline{ð}F = -(ð\overline{A} + \overline{ð}A), \tag{D16}$$

which completes our alternate proof of (6.24).

It seems virtually certain that our nonabelian version of (D16), namely (6.23), can be derived in a similar fashion using the nonabelian version of Stokes' Theorem.[14]

---

[1] W. Dreschler and M. Mayer, *Fiber Bundle Techniques in Gauge Theories: Lectures in Mathematical Physics at the University of Texas at Austin* (Springer-Verlag, Berlin, 1977).

[2] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).

[3] R. Penrose, *Relativity Groups and Topology: the 1963 Les Houches Lectures.* (Gordon and Breach, New York, 1964).

[4] G. Sparling, University of Pittsburgh report (unpublished).

[5] E. T. Newman, Phys. Rev. D **22**, 3023 (1980).

[6] S. S. Chern, *Complex Manifolds without Potential Theory* (Springer-Verlag, Berlin, 1981).

[7] M. Ko, E. T. Newman, and K. P. Tod, in *Asymptotic Structure of Space-time*, edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).

[8] Though we do not fully understand the connection of our work to that of J. Isenberg and P. B. Yasskin [*Research Notes in Mathematics, Vol. 32*, edited by D. E. Lerner and P. D. Sommers (Pitman, San Francisco, 1979)], a connection clearly exists. Equation (5.6) appears to exemplify this relationship.

[9] The self-dual form of this equation was previously obtained by E. T. Newman, Phys. Rev. D **8**, 2901 (1978).

[10] $\overline{ð}ð = 4P^2(\partial^2/\partial\zeta\partial\overline{\zeta}) = (1 + \zeta\overline{\zeta})^2(\partial^2/\partial\zeta\partial\overline{\zeta})$ is the two-dimensional Laplacian on the sphere in stereographic coordinates.

[11] R. Penrose, Gen. Rel. Grav. **12**, 225 (1980).

[12] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[13] A. Janis and E. Newman, J. Math. Phys. **6**, 902 (1965).

[14] M. Iyanaga, J. Math. Phys. **22**, 2713 (1981).

On the generalization of the 't Hooft field strength in the Yang–Mills–Higgs model

P. Houston[a]
Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, 91440 France

(Received 14 December 1981; accepted for publication 25 January 1982)

The generalization to the gauge groups SU(n) and SO(n) of the 't Hooft electromagnetic field is considered for finite energy configurations of the Yang–Mills–Higgs model, and it is shown how this results in the association of sets of integer charges with any closed surface. Differential inequalities satisfied by the Higgs invariants are presented, and their implications are discussed.

PACS numbers: 11.15.−q, 11.30.−j

I. INTRODUCTION

There has been much progress recently with the Bogomolny system[1] of equations arising in the limit of vanishing potential for the classical static finite energy Yang–Mills–Higgs model in three dimensions. Initially an existence theorem was proven for a solution to the system for arbitrary topological charges and gauge groups.[2] Subsequently, for the gauge group SU(2) explicit solutions were constructed first to the axially symmetric version of the Bogomolny equations[3] and then generally.[4] These axially symmetric solutions are seen to have their charges (i.e., the zero set of the Higgs field) concentrated at a single point. This fact was not unexpected as it had been previously shown that any axially symmetric solution must have this property.[5] In the demonstration of this result [which was for the SU(2) gauge group only] a crucial part was played by the 't Hooft electromagnetic field,[6] which could be used to associate an integer with an arbitrary closed surface for any finite energy configuration of the fields.

It is our first purpose in this paper to establish what corresponds to the generalization of the 't Hooft field strength for the case of larger gauge groups of physical interest [SU(n), $n\geqslant 2$; SO(n), $n\geqslant 5$]. This is achieved by recognizing the connection between these field strengths and curvature forms induced on certain vector bundles over submanifolds of three-dimensional Euclidean space. Considering the first Chern classes of these bundles, which define a set of conserved currents, leads to the association of sets of integer charges with an arbitrary closed surface for any finite energy configuration. This we do in Sec. III.

Making the additional requirement that the field equations be satisfied, for arbitrary Higgs potential, it is of interest, with regard to these generalized field strengths, to obtain information on the invariants formed from the Higgs field (e.g., their zeros). Thus in Sec. IV we derive differential inequalities satisfied by, and involving only, these Higgs invariants. Applying these inequalities to the case of vanishing Higgs potential results in more bounds on the Higgs field than previously known.

We give a summary of our conclusions in Sec. V and Appendix A is complementary to Sec. III.

II. NOTATION AND VARIATIONAL EQUATIONS

The main purpose of this section is to establish the notation we are going to use throughout the paper.

Let M be m-dimensional Euclidean space, $\mathbb{E}^m$, with inner produce denoted by $<\cdot,\cdot>$ and let $P = M \times G$ be the trivial principal fiber bundle over M with structure group G being a compact, connected, simple Lie group and with bundle projection given by the map: $P \ni (x,g) \mapsto x \in M$. We shall, in fact, only be concerned with the case of G being SU(n), $n\geqslant 2$, or SO(n), $n\geqslant 5$. G acts on P by multiplication on the right [i.e., $(x,g)\cdot a = (x,ga)$, $\forall\ (x,g)\in P$, $a\in G$].

We consider connections on P whose connection forms are at least continuous with respect to the section of P given by the map: $M \ni x \to (x,1) \in P$, 1 being the identity element of G.

We let $\mathscr{E} = M \times \mathscr{G}$ be the trivial vector bundle associated with P with fibers isomorphic to $\mathscr{G}$, the Lie algebra of G. G acts on $\mathscr{E}$ by the adjoint bundle action [i.e., $(x, A)\cdot g = (x, \mathrm{ad}(g^{-1})A) = (x, g^{-1}Ag)$, $\forall (x, A) \in \mathscr{E}, g\in G$]. The bundle projection is $\mathscr{E} \ni (x, A) \mapsto x \in M$.

With any section of $\mathscr{E}$ or, equivalently, with any map $\Phi: M \to \mathscr{G}$, and which we require to be at least continuous on M, we can associate a tensorial 0-form on P of type (ad,$\mathscr{G}$) denoted by ϕ and given by the map $\phi: P \ni (x,g) \to \mathrm{ad}(g^{-1})\Phi(x) \in \mathscr{G}$. Such forms we will call Higgs forms.

In general, for any tensorial p-forms η_1, η_2 of type (ρ,V), where V is a vector space, $\rho: G \to \mathscr{L}(V,V)$ is a linear representation of G onto the space of linear automorphism of V, and V is endowed with a G-invariant scalar product $(\cdot,\cdot)$, we can define their inner product as the tensorial 0-form of type $(0,\mathbb{R})$, i.e., a real-valued G-invariant function, or 0-form on M, by

$$(\eta_1,\eta_2) = (p!)^{-1}\sum(\eta_1^{j_1\cdots j_p}, \eta_2^{j_1\cdots j_p}),$$

where $\eta_i = \eta_i^{j_1\cdots j_p}\,dx^{j_1}\wedge\cdots\wedge dx^{j_p}$, $i = 1,2$. Moreover, if λ and η denote a tensorial 0-form and a tensorial p-form of type (ρ,V), respectively, we can define a tensorial p-form of type $(0,\mathbb{R})$, i.e., a real-valued p-form on M, by $(\lambda,\eta) = (\lambda,\eta^{j_1\cdots j_p})$ $dx^{j_1}\wedge\cdots\wedge dx^{j_p}$.

We denote the exterior covariant derivative which

---

[a] Present address: Theoretical Physics Division, CERN, CH-1211 Geneva 23, Switzerland.

0022-2488/83/040960-09$02.50

maps pseudotensorial p-forms of type (ρ, V) onto tensorial $(p+1)$-forms of type (ρ, V) by D. The curvature Ω is then given by the tensorial 2-form of type $(\mathrm{ad}, \mathscr{G})$: $\Omega = D\omega$, where ω is the pseudotensorial connection 1-form.

For simplicity, we will represent the group SU(n), (SO(n)) by all complex (real) $n \times n$ matrices, $\mathbb{C}^{n\times n}$ ($\mathbb{R}^{n\times n}$) which are unitary (orthogonal) and unimodular, and we will represent their respective Lie algebras by all Hermitian, traceless (skew-symmetric) complex (real) $n \times n$ matrices. We will take the inner product on $V \subset \mathbb{C}^{n\times n}$ ($\mathbb{R}^{n\times n}$), invariant under the adjoint action of SU(n) (SO(n)), to be: $(A,B) = \mathrm{tr}(A^{\dagger}B)$ $[\frac{1}{2}\mathrm{tr}(A^{T}B)]$ $\forall\ A,B \in V$.

Let U: $\mathscr{G} \to \mathbb{R}^+$ be a polynomial map of $\mathscr{G}$ onto the nonnegative real numbers (usually of degree not exceeding 4) invariant by the adjoint action of G on $\mathscr{G}$, and let the zero set of U: $Z(U) = \{A \in \mathscr{G} \mid U(A) = 0\}$ be nontrivial, i.e., $\neq \{0\}$, in which case U, which is called the Higgs potential, is said to have a spontaneously broken symmetry. Specific details of the map U will not, however, be needed in the paper.

A pair (ϕ,ω) of a Higgs form and a connection form, as given above, are said to form a finite energy configuration if $\mathscr{A}(\phi,\omega) < \infty$, where $\mathscr{A}$ maps (ϕ,ω) into $\mathbb{R}^+$ by

$$\mathscr{A}(\phi,\omega) = \tfrac{1}{2}\int_{\mathrm{M}} *\{(\Omega,\Omega) + (\mathrm{D}\phi,\mathrm{D}\phi) + \mathrm{U}\circ\phi\}, \tag{2.1}$$

and $*$ denotes the Hodge duality operation taking p-forms to (m − p)-forms. Among such finite energy configurations are ones for which (Ω,Ω), $(\mathrm{D}\phi,\mathrm{D}\phi)$, and $\mathrm{U}\circ\phi$ are smooth functions on M, for which $\lim_{\mathrm{R}\to\infty}(\Omega,\Omega)(\mathrm{R}\hat{\mathrm{x}})$, $\lim_{\mathrm{R}\to\infty}(\mathrm{D}\phi,\mathrm{D}\phi)(\mathrm{R}\hat{\mathrm{x}})$, and $\lim_{\mathrm{R}\to\infty}\mathrm{U}\circ\phi\,(\mathrm{R}\hat{\mathrm{x}})$ exist and are zero, for any $\hat{\mathrm{x}} \in \mathrm{S}_1^{\mathrm{m}-1} \subset \mathrm{M}$ and for which $\lim_{\mathrm{R}\to\infty}\phi(\mathrm{R}\hat{\mathrm{x}},\mathrm{g}) = \hat{\phi}(\hat{\mathrm{x}},\mathrm{g})$ exists, for any $(\hat{\mathrm{x}},\mathrm{g}) \in \mathrm{j}^*\mathrm{P}$, where $\hat{\phi}$ is a continuous tensorial 0-form on $\mathrm{j}^*\mathrm{P}$ of type $(\mathrm{ad},\mathscr{G})$ valued in Z(U). Here $\mathrm{S}_{\mathrm{R}}^{\mathrm{m}-1} = \{\mathrm{x} \in \mathrm{M} \mid \langle \mathrm{x},\mathrm{x}\rangle = \mathrm{R}\}$, and $\mathrm{j}^*\mathrm{P}$ denotes the bundle given by the pullback of P under the inclusion map j: $\mathrm{S}_1^{\mathrm{m}-1} \hookrightarrow \mathrm{M}$. We let W denote the space of all such finite energy configurations (ϕ,ω). Associated with $\hat{\phi}$ above, since the bundle $\mathrm{j}^*\mathrm{P}$ is trivial, we have a continuous map $\mathrm{f}_{\hat{\phi}}$: $\mathrm{S}_1^{\mathrm{m}-1} \ni \hat{\mathrm{x}} \to \hat{\phi}(\hat{\mathrm{x}},1) \in \mathrm{Z(U)}$. Thus we can obtain a partitioning of W via the homotopy classification of such maps $\mathrm{f}_{\hat{\phi}}$, i.e.,

$$W = \bigcup_{\alpha \in \pi_{m-1}(Z(U))} W_\alpha, \quad W_\alpha \cap W_{\alpha'} = \text{null}, \quad \alpha \neq \alpha', \tag{2.2}$$

where W_α contains all pairs $(\phi,\omega) \in \mathrm{W}$ such that $\mathrm{f}_{\hat{\phi}}$ is in the homotopy class $\alpha \in \pi_{\mathrm{m}-1}(\mathrm{Z(U)})$.

By looking for stationary points of $\mathscr{A}$ on each W_α, we obtain the following equations:

$$*(\mathrm{D}(*\Omega)) = [\mathrm{D}\phi,\phi], \quad \Delta_\omega\phi = \mathrm{U}'\circ\phi, \tag{2.3}$$

where $[A,B] = AB - BA$, for any square matrices A,B,U': $\mathscr{G} \to \mathscr{G}$ is the map given by

$$(U'(A),B) = \left[\frac{d}{dt}U(A+tB)\right]_{t=0}, \quad \forall\ A,B \in \mathscr{G}, \tag{2.4}$$

and Δ_ω denotes the Laplacian. If η denotes a pseudotensorial or tensorial p-form of type (ρ, V), then $\Delta_\omega\eta$ is the tensorial p-form given by $\Delta_\omega\eta = *(D(*D\eta))$. On p-forms of type $(0,\mathbb{R})$, Δ_ω becomes the usual Laplacian $*d(*d)$, on $\mathbb{E}^m$, where d is the exterior derivative and we will denote the Laplacian here by Δ_0.

Henceforth, we will take $m = 3$, G will denote one of the groups SU(n), $n \geqslant 2$, or SO(n), $n \geqslant 5$, and $\mathscr{G}$ will denote its respective Lie algebra, unless stated otherwise. To avoid notational difficulties, $(-1)^{1/2}$ will denote the imaginary number.

III. CONSTRUCTION OF TOPOLOGICAL CURRENTS FOR CONFIGURATIONS AND THEIR PROPERTIES

For the Lie algebra $\mathscr{G}$ of SU(n), $n \geqslant 2$, or SO(n), $n \geqslant 5$, we can construct dense subsets $\widetilde{\mathscr{G}}_\mu$ of $\mathscr{G}$ for each μ (the label μ will be defined presently) and within each $\widetilde{\mathscr{G}}_\mu$ we can obtain an orbit of the adjoint action of G, $\widehat{\mathscr{G}}_\mu \subset \widetilde{\mathscr{G}}_\mu$, together with a deformation retraction r_μ: $\widetilde{\mathscr{G}}_\mu \to \widehat{\mathscr{G}}_\mu$ so that $\widetilde{\mathscr{G}}_\mu$ is a tubular neighborhood of $\widehat{\mathscr{G}}_\mu$. The construction is obtained as follows:

(A) For SU(n), $n \geqslant 2$, any element of $\mathscr{G}$ is conjugate to, i.e., on the same orbit under, the adjoint action of SU(n) as $\mathrm{diag}(\lambda(1),...,\lambda(n))$ for some $\lambda(1) \geqslant \lambda(2) \geqslant \cdots \geqslant \lambda(n)$, with $\Sigma_{i=1}^{n}\lambda(i) = 0$. Let $\mu = (\alpha_1,...,\alpha_l)$, where $\alpha_i \in \mathrm{N}$, $\Sigma_{i=1}^{l}\alpha_i = n$ and $l \geqslant 2$. We let $\widetilde{\mathscr{G}}_\mu$ denote the set of all elements of $\mathscr{G}$ conjugate to $\mathrm{diag}(\lambda(1),...,\lambda(n))$, with $\Sigma_{i=1}^{n}\lambda(i) = 0$, where

$$\begin{aligned} \lambda(1) &\geqslant \cdots \geqslant \lambda(\alpha_1) > \lambda(\alpha_1+1) \\ &\geqslant \cdots \geqslant \lambda(\alpha_1+\cdots+\alpha_{l-1}) \\ &\quad > \lambda(\alpha_1+\cdots+\alpha_{l-1}+1) \\ &\geqslant \cdots \geqslant \lambda(n). \end{aligned} \tag{3.1}$$

Letting $A_\mu^0 = \mathrm{diag}(\lambda^0(1),...,\lambda^0(n))$ with $\lambda^0(i)$ fixed and satisfying $\Sigma_{i=1}^{n}\lambda^0(i) = 0$ and

$$\begin{aligned} \lambda^0(1) &= \cdots = \lambda^0(\alpha_1) > \lambda^0(\alpha_1+1) \\ &= \cdots = \lambda^0(\alpha_1+\cdots+\alpha_{l-1}) \\ &\quad > \lambda^0(\alpha_1+\cdots+\alpha_{l-1}+1) = \cdots = \lambda^0(n), \end{aligned} \tag{3.2}$$

we put $\widehat{\mathscr{G}}_\mu = G(A_\mu^0)$, i.e., the orbit through A_μ^0. The map r_μ is then defined by

$$r_\mu : \widetilde{\mathscr{G}}_\mu \ni g\begin{bmatrix} \lambda(1)\cdots & 0 \\ \vdots\ddots & \\ 0 & \lambda(n) \end{bmatrix}g^{-1} \to gA_\mu^0 g^{-1} \in \widehat{\mathscr{G}}_\mu, \tag{3.3}$$

where g is any element of SU(n) and $\lambda(1),...,\lambda(n)$ satisfy Eq. (3.1).

(B) For SO(n), where $n = 2j+1$ and $j \geqslant 1$, any element of its Lie algebra is conjugate by some element of SO(n) to $\mathrm{diag}(\lambda(1)J,...,\lambda(j)J,0)$ for some $\lambda(1) \geqslant \cdots \geqslant \lambda(j) \geqslant 0$, where J denotes the skew-symmetric 2×2 real matrix with $J_{12} = 1$. Here let $\mu = (2\alpha_1,...,2\alpha_l,\beta)$, where $\alpha_i,\beta \in \mathrm{N}$, $2\Sigma_{i=1}^{l}\alpha_i + \beta = n$ and $l \geqslant 1$. We let $\widetilde{\mathscr{G}}_\mu$ contain all elements of $\mathscr{G}$ conjugate to $\mathrm{diag}(\lambda(1)J,...,\lambda(j)J,0)$, where

$$\begin{aligned} &\lambda(1) \geqslant \cdots \geqslant \lambda(\alpha_1) > \lambda(\alpha_1+1) \geqslant \cdots \geqslant \lambda(\alpha_1+\cdots+\alpha_l) \\ &\quad > \lambda(\alpha_1+\cdots+\alpha_l+1) \geqslant \cdots \geqslant \lambda(j) \geqslant 0. \end{aligned} \tag{3.4}$$

Letting $A_\mu^0 = \mathrm{diag}(\lambda^0(1)J,...,\lambda^0(j)J,0)$ with $\lambda^0(i)$ fixed and satisfying

$$\begin{aligned} \lambda^0(1) &= \cdots = \lambda^0(\alpha_1) > \lambda^0(\alpha_1+1) \\ &= \cdots = \lambda^0(\alpha_1+\cdots+\alpha_l) > \lambda^0(\alpha_1+\cdots+\alpha_l+1) \\ &= \cdots = \lambda(j) = 0, \end{aligned} \tag{3.5}$$

then $\widehat{\mathscr{G}}_\mu = G(A_\mu^0)$ and r_μ is defined by

$$r_\mu : \tilde{\mathscr{G}}_\mu \ni g \begin{bmatrix} \lambda(1)J \cdots & & 0 \\ \vdots \ddots & & \\ & \lambda(j)J & \\ 0 & & 0 \end{bmatrix} g^{-1} \rightarrow g A^0_\mu g^{-1} \in \hat{\mathscr{G}}_\mu, \tag{3.6}$$

where g is any element of SO(n) and $\lambda(1),\ldots,\lambda(j)$ satisfy Eq. (3.4).

(C) For SO(n), where $n = 2j$ and $j \geqslant 3$, any element of $\mathscr{G}$ is conjugate, by some element of SO(n), either to $A_{(+)} = \operatorname{diag}(\lambda(1)J,\ldots,\lambda(j)J)$ or to $A_{(-)} = \operatorname{diag}(\lambda(1)J,\ldots,\lambda(j-1)J, -\lambda(j)J)$, for some $\lambda(1) \geqslant \cdots \geqslant \lambda(j) \geqslant 0$. Let μ be given by one of the following three expressions:

(i$_\text{a}$) $\mu = (2\alpha_1,\ldots,2\alpha_l, +)$,

where $\alpha_i \in \mathrm{N}$, $\alpha_l > 1$, and $2\sum_{i=1}^{l} \alpha_i = n$, or

(i$_\text{b}$) $\mu = (2\alpha_1,\ldots,2\alpha_l, +)$,

where $\alpha_i \in \mathrm{N}$, $\alpha_l = 1$, and $2\sum_{i=1}^{l} \alpha_i = n$, or

(ii) $\mu = (2\alpha_1,\ldots,2\alpha_l, -)$,

where $\alpha_i \in \mathrm{N}$, $\alpha_l > 1$, and $2\sum_{i=1}^{l} \alpha_i = n$, or

(iii) $\mu = (2\alpha_1,\ldots,2\alpha_l,\beta)$,

where $\alpha_i \in \mathrm{N}$, $\beta = 4,6,\ldots$, and $2\sum_{i=1}^{l} \alpha_i + \beta = n$,

where in each case $l \geqslant 1$. [We may sometimes refer to (i$_\text{a}$) and (i$_\text{b}$) together as (i)]. $\tilde{\mathscr{G}}_\mu$ is given by all elements of $\mathscr{G}$ conjugate to $A_{(+)}$ for μ of type (i$_\text{a}$) and $A_{(-)}$ for μ of type (ii) for some $\lambda(1),\ldots,\lambda(j)$, which in each case satisfies

$$\begin{aligned} \lambda(1) &\geqslant \cdots \geqslant \lambda(\alpha_1) > \lambda(\alpha_1 + 1) \\ &\geqslant \cdots \geqslant \lambda(\alpha_1 + \cdots + \alpha_{l-1}) > \lambda(\alpha_1 + \cdots + \alpha_{l-1} + 1) \\ &\geqslant \cdots \geqslant \lambda(j) > 0. \end{aligned} \tag{3.7}$$

For μ of type (iii) $\tilde{\mathscr{G}}_\mu$ is given by all elements of $\mathscr{G}$ conjugate to $A_{(+)}$ or $A_{(-)}$ for some $\lambda(1),\ldots,\lambda(j)$ satisfying

$$\lambda(1) \geqslant \cdots \geqslant \lambda(\alpha_1) > \lambda(\alpha_1 + 1) \geqslant \cdots \geqslant \lambda(\alpha_1 + \cdots + \alpha_l) > 0$$

and

$$\begin{aligned} &\lambda(\alpha_1 + \cdots + \alpha_l) > \lambda(i) \geqslant \lambda(i') \geqslant 0, \quad \forall\ i,i' \text{ with} \\ &\quad \alpha_1 + \cdots + \alpha_l < i < i' \leqslant j. \end{aligned} \tag{3.8}$$

For μ of type (i$_\text{b}$), $\tilde{\mathscr{G}}_\mu$ is given by all elements of $\mathscr{G}$ conjugate to $A_{(+)}$ or $A_{(-)}$ for some $\lambda(1),\ldots,\lambda(j)$ satisfying

$$\begin{aligned} \lambda(1) &\geqslant \cdots \geqslant \lambda(\alpha_1) > \lambda(\alpha_1 + 1) \\ &\geqslant \cdots \geqslant \lambda(\alpha_1 + \cdots + \alpha_{l-1}) > \lambda(j) \geqslant 0. \end{aligned} \tag{3.9}$$

For μ of type (i), or (ii), we let $A^0_\mu = \operatorname{diag}(\lambda^0(1)J,\ldots,\lambda^0(j)J)$, or $\operatorname{diag}(\lambda^0(1)J,\ldots,\lambda^0(j-1)J, -\lambda^0(j)J)$, respectively, where the $\lambda^0(i)$ are fixed and satisfy

$$\begin{aligned} \lambda^0(1) &= \cdots = \lambda^0(\alpha_1) > \lambda^0(\alpha_1 + 1) \\ &= \cdots = \lambda^0(\alpha_1 + \cdots + \alpha_{l-1}) \\ &\quad > \lambda^0(\alpha_1 + \cdots + \alpha_{l-1} + 1) = \cdots = \lambda^0(j) > 0. \end{aligned} \tag{3.10}$$

For μ of type (iii) we let $A^0_\mu = \operatorname{diag}(\lambda^0(1)J,\ldots,\lambda^0(j)J)$, where the $\lambda^0(i)$ are fixed and satisfy

$$\begin{aligned} \lambda^0(1) &= \cdots = \lambda^0(\alpha_1) > \lambda^0(\alpha_1 + 1) \\ &= \cdots = \lambda^0(\alpha_1 + \cdots + \alpha_l) > 0 \end{aligned}$$

and

$$\lambda^0(i) = 0, \quad \forall i > \alpha_1 + \cdots + \alpha_l. \tag{3.11}$$

We put $\tilde{\mathscr{G}}_\mu = G(A^0_\mu)$ and r_μ is defined by

$$r_\mu : \tilde{\mathscr{G}}_\mu \ni g A_{(\delta)} g^{-1} \rightarrow g A^0_\mu g^{-1} \in \hat{\mathscr{G}}_\mu, \tag{3.12}$$

where g is any element of SO(n). Here, for μ of type (i$_\text{a}$), $A_{(\delta)} = A_{(+)}$ and, for μ of type (ii), $A_{(\delta)} = A_{(-)}$, where $\{\lambda(i)\}$ in each case satisfy Eq. (3.7), for μ of type (iii), $A_{(\delta)} = A_{(+)}$ or $A_{(-)}$ and $\{\lambda(i)\}$ satisfy Eq. (3.8), and, for μ of type (i$_\text{b}$), $A_{(\delta)} = A_{(+)}$ or $A_{(-)}$, where $\{\lambda(i)\}$ satisfy Eq. (3.9).

[*Note*: The Cartan subalgebra of $\mathscr{G}$, in our case, is given by all its diagonal or block-diagonal elements, as appropriate. The Weyl reflection groups act on this Cartan subalgebra as follows: $\lambda(i) \rightarrow \lambda(\pi(i))$, for SU($n$); $\lambda(i) \rightarrow \pm\lambda(\pi(i))$, for SO($n$), n odd; and $\lambda(i) \rightarrow \pm\lambda(\pi(i))$, with an even number of ($-$)'s, for SO(n), n even. Here π denotes a permutation of the diagonal or block-diagonal elements. The fundamental Weyl chamber may then be given as follows: For SU(n), $\lambda(1) \geqslant \cdots \geqslant \lambda(n)$; for SO($n$), n odd, $\lambda(1) \geqslant \cdots \geqslant \lambda([n/2]) \geqslant 0$; and for SO($n$), n even, $\lambda(1) \geqslant \cdots \geqslant \lambda([n/2] - 1) \geqslant |\lambda([n/2])|$. Thus our construction of neighborhoods $\tilde{\mathscr{G}}_\mu$ may be induced by the action of G on the various subsets of this fundamental Weyl chamber obtained by the omission of combinations of its faces, edges, etc.]

It is evident that in each case above, r_μ is a deformation retraction and that the bundle $\tilde{\mathscr{G}}_\mu$, with projection r_μ, is a tubular neighborhood of $\hat{\mathscr{G}}_\mu$.

For $(\phi,\omega) \in W$ we define $P_\mu = P - \phi^{-1}(\mathscr{G} - \tilde{\mathscr{G}}_\mu)$ for each μ. Since $P_\mu \subset P$, the bundle projection on P is also defined on P_μ and we denote by M_μ the image of P_μ by this map. We impose a technical restriction here, in that we only consider such configurations $(\phi,\omega) \in W$ which are nondegenerate in the sense that either M_μ is a manifold of the same dimension as M or void. [*Only* for μ of type C(i) or (ii) above do we allow the possibility that M_μ is void, i.e., it may happen that either one of $\tilde{\mathscr{G}}_{\mu'}$ or $\tilde{\mathscr{G}}_{\mu''}$ does not intersect the image of M under ϕ, where $\mu' = (2\alpha_1,\ldots,2\alpha_l, +)$ and $\mu'' = (2\alpha_1,\ldots,2\alpha_l, -)$, $2\Sigma_{i=1}^{l} \alpha_i = n$; however, $\tilde{\mathscr{G}}_{\mu'} \cup \tilde{\mathscr{G}}_{\mu''}$ *must* intersect the image of M under ϕ.] The discussion of the paper will now proceed for those M_μ which are nonvoid. For any μ we have that $P_\mu = M_\mu \times G$ is a trivial principal fiber bundle over M_μ. Letting $i_\mu : P_\mu \hookrightarrow P$ denote the inclusion map of P_μ in P, it is clear that the map $\phi \circ i_\mu : P_\mu \rightarrow \tilde{\mathscr{G}}_\mu \subset \mathscr{G}$ is a tensorial 0-form of type (ad, $\mathscr{G}$) on P_μ, valued in $\tilde{\mathscr{G}}_\mu$. Moreover, we can define the map $\phi_\mu = r_\mu \circ \phi \circ i_\mu : P_\mu \rightarrow \hat{\mathscr{G}}_\mu \subset \mathscr{G}$ which is a $\hat{\mathscr{G}}_\mu$-valued tensorial 0-form of type (ad, $\mathscr{G}$) on P_μ (we will regard such maps as ϕ_μ to be valued in $\hat{\mathscr{G}}_\mu$, $\tilde{\mathscr{G}}_\mu$, $\mathscr{G}$, $\mathbb{C}^{n \times n}$, $\cdots$ as appropriate, throughout the paper). The map $i_\mu : P_\mu \hookrightarrow P$ together with the identity map on G defines a bundle homomorphism $P_\mu \rightarrow P$. It follows (Proposition 6.2, p. 81, Ref. 7) that the connection on P with connection form ω determines uniquely, via this bundle homomorphism, a connection on P_μ with connection form $i^*_\mu \omega$.

For each μ, the stability subgroup of any point of $\hat{\mathscr{G}}_\mu$ is conjugate to $H_\mu = G_{A^0_\mu}$, the stability subgroup of A^0_μ. Moreover, it is clear that, via the map $G/H_\mu \ni gH_\mu \to gA_\mu^0 g^{-1} \in \hat{\mathscr{G}}_\mu$, the coset space G/H_μ and the orbit $\hat{\mathscr{G}}_\mu$ may be identified. Specifically, the stability subgroups H_μ are: for case A, with $\mu = (\alpha_1,...,\alpha_l)$, $\Sigma_{i=1}^l \alpha_i = n$, $H_\mu \cong S(U(\alpha_1)\times\cdots\times U(\alpha_l))$ (here S means that the overall determinant is 1); for the cases B and C (iii) with $\mu = (2\alpha_1,...,2\alpha_l,\beta)$, $\beta + 2\Sigma_{i=1}^l \alpha_i = n$, $H_\mu \cong U(\alpha_1)\times\cdots\times U(\alpha_l)\times \mathrm{SO}(\beta)$; for the cases C(i) and (ii) with $\mu = (2\alpha_1,...,2\alpha_l, +$ or $-)$, $2\Sigma_{i=1}^l \alpha_i = n$, $H_\mu \cong U(\alpha_1)\times\cdots\times U(\alpha_l)$. In each case H_μ has the block diagonal form compatible with A_μ^0. Now the trivial bundle $E_\mu = M_\mu \times \hat{\mathscr{G}}_\mu \cong M_\mu \times G/H_\mu$ with bundle projection onto M_μ gives an associated bundle with P_μ with fiber G/H_μ. Denoting by σ_0 the section of P_μ, σ_0: $M_\mu \ni x \to (x,1) \in P_\mu$, then the map $\phi_\mu^0 = \phi_\mu \circ \sigma_0$: $M_\mu \to \hat{\mathscr{G}}_\mu$ gives the section of E_μ: $M_\mu \ni x \to (x,\phi_\mu^0(x)) \in E_\mu$. It then follows, by Proposition 5.6, p. 57, Ref. 7, that P_μ is reducible to a principal fiber bundle Q_μ with structure group H_μ. Q_μ is, in fact, uniquely determined by the map ϕ_μ^0 to be the pullback bundle $Q_\mu = \phi_\mu^{0*}G$, where G here is viewed as the principal fiber bundle over $\hat{\mathscr{G}}_\mu \cong G/H_\mu$ with structure group H_μ and bundle projection: $G \ni g \to gA_\mu^0 g^{-1} \in \hat{\mathscr{G}}_\mu$. Letting $\mathscr{H}_\mu$ denote the subalgebra of $\mathscr{G}$ which is the Lie algebra of $H_\mu < G$, we need the following result:

Proposition 3.1: H_μ is a reductive subgroup of G, i.e., $\mathscr{G} = \mathscr{H}_\mu \oplus \mathscr{M}$ and $\mathrm{ad}(H_\mu)\mathscr{M} = \mathscr{M}$.

Proof: The scalar product on $\mathscr{G}$, $(\cdot,\cdot)$, is the G-invariant Cartan–Killing form. Let $\mathscr{M} = \mathscr{H}^\perp$ with respect to $(\cdot,\cdot)$. For any $h \in H_\mu$, $A \in \mathscr{M}$, $B \in \mathscr{H}_\mu$ we have that $(\mathrm{ad}(h)A,B) = (A,\mathrm{ad}(h^{-1})B)$. However, since $\mathrm{ad}(h^{-1})B$ is in $\mathscr{H}_\mu$, it follows that $(\mathrm{ad}(h)A,B)$ is zero, $\forall\, B \in \mathscr{H}_\mu$. Hence $\mathrm{ad}(h)A \in \mathscr{M}$, $\forall\, h \in H_\mu$, $A \in \mathscr{M}$.

From this and Proposition 6.4, p. 83, Ref. 7, a connection is induced on the bundle Q_μ from the connection on P_μ and its connection form is given by $(\epsilon_\mu^*\omega)_{\mathscr{H}_\mu}$, where $(A)_{\mathscr{H}_\mu}$ denotes the $\mathscr{H}_\mu$ component of any element A in $\mathscr{G}$ under the decomposition $\mathscr{G} = \mathscr{H}_\mu \oplus \mathscr{M}$, and $\epsilon_\mu = i_\mu \circ \delta_\mu$, where δ_μ: $Q_\mu \hookrightarrow P_\mu$ denotes the inclusion map of Q_μ in P_μ.

We now define maps $p_{i\mu}$: $\hat{\mathscr{G}}_\mu \to G_{\alpha_i n}$ onto the complex Grassman manifolds $G_{\alpha_i n}$, $i = 1,...,l$, for each μ [where each μ is given by $(\alpha_1,...,\alpha_l)$, $(2\alpha_1,...,2\alpha_l,\beta)$, or $(2\alpha_1,...,2\alpha_l, +$ or $-)$ as appropriate]. Here, for convenience, we take $G_{\alpha n} = \{A \in \mathbb{C}^{n\times n} | A^\dagger = A, A^2 = A, \mathrm{tr}A = \alpha\}$, $\alpha \leqslant n$. $G_{\alpha n}$ may be identified with $G_\alpha(\mathbb{C}^n)$, the space of α-planes in $\mathbb{C}^n$. The maps $p_{i\mu}$, $i = 1,...,l$, are associated with projections onto the various eigenspaces of $\hat{\mathscr{G}}_\mu$. Specifically these maps are given as follows:

(A) For SU(n) with $\mu = (\alpha_1,...,\alpha_l)$, $\Sigma_{i=1}^l \alpha_i = n$, $p_{i\mu}$ is defined by the map

$$p_{i\mu}: \hat{\mathscr{G}}_\mu \ni gA_\mu^0 g^{-1} \to g\Theta_{i\mu} g^{-1} \in G_{\alpha_i n} \subset \mathbb{C}^{n\times n} \tag{3.13}$$

(we will regard the range of $p_{i\mu}$ as belonging to $G_{\alpha_i n}$ or $\mathbb{C}^{n\times n}$ as is appropriate), where $\Theta_{i\mu} = \mathrm{diag}(0,...,0, 1,...,1, 0,...,0)$, with 1 in the $\alpha_1 + \cdots + \alpha_{i-1} + 1$ to $\alpha_1 + \cdots + \alpha_i$ diagonal positions and 0 elsewhere. Since $\Theta_{i\mu}^\dagger = \Theta_{i\mu}$, $\Theta_{i\mu}^2 = \Theta_{i\mu}$, and $\mathrm{tr}\Theta_{i\mu} = \alpha_i$, it follows that $g\Theta_{i\mu}g^{-1} \in G_{\alpha_i n}$. The maps given by Eq. (3.13), on $G_{\alpha_i n}$, are onto and $G_{\alpha_i n}$ may be identified with the orbit homeomorphic to the coset space SU(n)/S$(U(\alpha_i) \times U(n-\alpha_i))$. Not all of the maps $p_{i\mu}$ are independent in the sense that if we regard the range of $p_{i\mu}$ as belonging to $\mathbb{C}^{n\times n} \supset G_{\alpha_i n}$, then, since $\mathbb{C}^{n\times n}$ is a vector space, linear summation of $\{p_{i\mu}\}$ is defined, and we have the relation

$$\sum_{i=1}^{l} p_{i\mu} = \mathbb{1}_n, \tag{3.14}$$

where $\mathbb{1}_n$ denotes the identity $n\times n$ matrix.

(B) For SO(n), n odd, or C(iii) for SO(n), n even, where $\mu = (2\alpha_1,...,2\alpha_l,\beta)$ and $n = 2\Sigma_{i=1}^l \alpha_i + \beta$, the maps $p_{i\mu}$ are again given by Eq. (3.13) but where now

$$\Theta_{i\mu} = -\tfrac{1}{2}[\Theta_{i\mu}'^{\,2} + (-1)^{1/2}\Theta_{i\mu}'], \tag{3.15}$$

and $\Theta_{i\mu}' = \mathrm{diag}(0,...,0, J,...,J, 0,...,0)$ with J in the $\alpha_1 + \cdots + \alpha_{i-1} + 1$ to $\alpha_1 + \cdots + \alpha_l$ (2×2)-block-diagonal positions and 0 everywhere else. Since

$$\Theta_{i\mu}'^{\,3} = -\Theta_{i\mu}',$$
$$\Theta_{i\mu}'^{\,T} = -\Theta_{i\mu}', \quad \text{and} \quad \mathrm{tr}(-\Theta_{i\mu}'^{\,2}) = 2\alpha_i, \tag{3.16}$$

it follows that, indeed, $g\Theta_{i\mu}g^{-1}$ is in $G_{\alpha_i n}$. The image of $\hat{\mathscr{G}}_\mu$ by $p_{i\mu}$, given by Eqs. (3.13) and (3.15), defines the algebraic submanifold of $G_{\alpha_i n}$ obtained from all elements A of $G_{\alpha_i n}$ which satisfy $A^TA = 0$. This submanifold is homeomorphic to the coset space SO(n)/U$(\alpha_i)\times$SO$(n - 2\alpha_i)$.

(C(i) or (ii)) For SO(n), n even, and $\mu = (2\alpha_1,...,2\alpha_l, +$ or $-)$, $2\Sigma_{i=1}^l \alpha_i = n$, the maps $p_{i\mu}$ are again given by Eqs. (3.13) and (3.15), where $\Theta_{i\mu}$, $i = 1,...,l-1$ are defined in the same way as for Eq. (3.15), but where $\Theta_{l\mu}$ is now given by $\mathrm{diag}(0,...,0, J,...,J, \pm J)$, with J in the $\alpha_1 + \cdots + \alpha_{l-1} + 1$ to $\alpha_1 + \cdots + \alpha_l - 1$ (2×2)-block-diagonal positions, $+j$ or $-J$ in the last (2×2)-block diagonal position depending on whether $\mu = (2\alpha_1,...,2\alpha_l, +$ or $-)$ and 0 elsewhere. Since Eq. (3.16) is again satisfied, $p_{i\mu}$ maps $\hat{\mathscr{G}}_\mu$ into $G_{\alpha_i n}$. Regarding the range of $p_{i\mu}$ as belonging to $\mathbb{C}^{n\times n} \supset G_{\alpha_i n}$ [for μ of type C(i) or (ii)] so that linear summation of $\{p_{i\mu}\}$ is defined, we have that the following condition is satisfied:

$$\sum_{i=1}^{l} \mathrm{Re}(p_{i\mu}) = \tfrac{1}{2}\mathbb{1}_n. \tag{3.17}$$

For $l \geqslant 2$ the image of $\hat{\mathscr{G}}_\mu$ under $p_{i\mu}$ defines the same algebraic submanifold of $G_{\alpha_i n}$ as for the previous cases of B and C(iii). However, for $l = 1$ a special situation arises in that the image of $\hat{\mathscr{G}}_\mu$ under $p_{i\mu}$ is the algebraic submanifold of $G_{\alpha_i n}$ given by all elements A of $G_{\alpha_i n}$ which satisfy $A^TA = 0$ and $\mathrm{pf}(\mathrm{Im}\,A) = +1$ or -1 for $\mu = (n,+)$ or $(n,-)$, respectively, where pf denotes the pfaffian (also in this submanifold $\mathrm{Re}A = \frac{1}{2}\mathbb{1}_n$). Either of these algebraic submanifolds is homeomorphic to the coset space SO(n)/U(α_1).

Over $G_{\alpha n}$ we have defined the Stiefel bundle $R_{\alpha n} \subset G_{\alpha n} \times \mathbb{C}^{n\times\alpha}$ with $(A,B) \in R_{\alpha n}$ if and only if $AB = B$ and $B^\dagger B = \mathbb{1}_\alpha$; the bundle projection is given by the map: $R_{\alpha n} \ni (A,B) \mapsto A \in G_{\alpha n}$ and the group U(α) acts on $(A,B) \in R_{\alpha n}$ by $(A,B)\cdot g = (A,Bg)$, for any $g \in$ U(α) [U(α) is represented by all unitary elements of $\mathbb{C}^{\alpha\times\alpha}$]. Associated with $R_{\alpha n}$ we have the vector bundle $E_{\alpha n} \subset G_{\alpha n} \times \mathbb{C}^{n\times\alpha}$ over $G_{\alpha n}$ and $(A,B) \in E_{\alpha n}$ if and only if $AB = B$ and $\det(B^\dagger B) \neq 0$. The bundle projection and group action of $GL(n,\mathbb{C})$ on $E_{\alpha n}$ is the same as for $R_{\alpha n}$. Through the maps $p_{i\mu}: \hat{\mathscr{G}}_\mu \to G_{\alpha_i n}$, $i = 1,\dots,l$ [$\mu = (\alpha_1,\dots,\alpha_l)$, $(2\alpha_1,\dots,2\alpha_l,\beta, +$, or $-)$ as appropriate] defined above, we can pull back the bundles $E_{\alpha_i n}$ and $R_{\alpha_i n}$ to define the bundles $p^*_{i\mu}E_{\alpha_i n}$ and $p^*_{i\mu}R_{\alpha_i n}$, respectively, over $\hat{\mathscr{G}}_\mu$. We can further pull back using the maps $\phi^0_{i\mu} = p_{i\mu}\circ\phi^0_\mu$: $M_\mu \to G_{\alpha_i n}$ to define the bundles $\phi^{0*}_{i\mu}E_{\alpha_i n}$ and $\phi^{0*}_{i\mu}R_{\alpha_i n}$, respectively, over M_μ. Thus we have

$$\begin{tikzcd}
\phi^{0*}_{i\mu}E_{\alpha_i n}(R_{\alpha_i n}) \arrow[d] & p^*_{i\mu}E_{\alpha_i n}(R_{\alpha_i n}) \arrow[d] & E_{\alpha_i n}(R_{\alpha_i n}) \arrow[d] \\
M_\mu \arrow[r, "\phi^0_\mu"] & \hat{\mathscr{G}}_\mu \arrow[r, "p_{i\mu}"] & G_{\alpha_i n}
\end{tikzcd} \tag{3.18}$$

for $i = 1,\dots,l$.

For each μ [of the form $(\alpha_1,\dots,\alpha_l)$ or $(2\alpha_1,\dots,2\alpha_l,\beta, +$, or $-)$], we now wish to see that the following three maps give a bundle homeomorphism: $Q_\mu \to \phi^{0*}_{i\mu}R_{\alpha_i n}$, $i = 1,\dots,l$:

(i) $f_{i\mu}: Q_\mu \ni (x,g) \to (x,\phi^0_\mu(x),\phi^0_{i\mu}(x),g\Gamma_{i\mu}) \in \phi^{0*}_{i\mu}R_{\alpha_i n}$, where $g\in G$ is such that $gA^0_\mu g^{-1} = \phi^0_\mu(x)$.

(ii) $f'_{i\mu}: H_\mu \ni h \to f'_{i\mu}(h) = \Gamma^\dagger_{i\mu}h\Gamma_{i\mu} \in U(\alpha_i)$.

(iii) $f''_{i\mu}: M_\mu \ni x \to x \in M_\mu$, i.e., the identity map on M_μ.

Here $\Gamma_{i\mu}$ denotes the $n\times\alpha_i$ complex-valued matrix defined as follows:

For case A, with $\mu = (\alpha_1,\dots,\alpha_l)$, $\Sigma^l_{i=1}\,\alpha_i = n$, the $(\alpha_1 + \cdots + \alpha_{i-1} + r,r)$ component of $\Gamma_{i\mu}$ is 1 for $r = 1,\dots,\alpha_i$ and all other components are zero, where $i = 1,\dots,l$.

For case B or C(i), (ii), or (iii), with $\mu = (2\alpha_1,\dots,2\alpha_l,\beta,$ $+$, or $-)$, $2\Sigma^l_{i=1}\,\alpha_i + \beta = n$ and with $i = 1,\dots,l$ in the case of B or C(iii) and with $i = 1,\dots,l-1$ in the case of C(i), (ii), the $(2\alpha_1 + \cdots + \alpha_{i-1} + 2r - 1,r)$ component of $\Gamma_{i\mu}$ is $(\frac{1}{2})^{1/2}$ and the $(2\alpha_1 + \cdots + 2\alpha_{i-1} + 2r,r)$ component of $\Gamma_{i\mu}$ is $(-1)^{1/2}\cdot(\frac{1}{2})^{1/2}$ for $r = 1,\dots,\alpha_i$ and all other components are zero.

For cases C(i) or (ii) with $\mu = (2\alpha_1,\dots,2\alpha_l,$ $+$ or $-)$ the $(2\alpha_1 + \cdots + 2\alpha_{l-1} + 2r - 1,r)$ component of $\Gamma_{l\mu}$ is $(\frac{1}{2})^{1/2}$, for $r = 1,\dots,\alpha_l$, the $(2\alpha_1 + \cdots + 2\alpha_{l-1} + 2r,r)$ component is $(-1)^{1/2}\cdot(\frac{1}{2})^{1/2}$, for $r = 1,\dots,\alpha_l - 1$, the (n,α_l) component is $(-1)^{1/2}\cdot(\frac{1}{2})^{1/2}$ for $\mu = (2\alpha_1,\dots,2\alpha_l,$ $+$ or $-)$, respectively, and all remaining components are zero.

The three forms for $\Gamma_{i\mu}$ are respectively:

$$\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & & 0 \\ 1 & \cdots & 0 \\ \vdots & \ddots & \\ 0 & & 1 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 0 \\ 1 & 0 & & 0 \\ (-1)^{1/2} & 0 & \cdots & 0 \\ 0 & 1 & & \\ 0 & (-1)^{1/2} & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & 1 \\ 0 & 0 & & (-1)^{1/2} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & & 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & & 0 & 0 \\ 1 & \cdots & 0 & 0 \\ (-1)^{1/2} & & 0 & 0 \\ \vdots & \ddots & & \\ & & & \\ 0 & \cdots & 1 & 0 \\ 0 & & (-1)^{1/2} & 0 \\ 0 & & 0 & 1 \\ 0 & \cdots & 0 & \pm(-1)^{1/2} \end{bmatrix} \tag{3.19}$$

From the following properties of $\Gamma_{i\mu}$,

$$\Gamma_{i\mu}\Gamma^\dagger_{i\mu} = \Theta_{i\mu}, \quad \Theta_{i\mu}\Gamma_{i\mu} = \Gamma_{i\mu}, \quad \Gamma^\dagger_{i\mu}\Gamma_{i\mu} = \mathbb{1}_{\alpha_i}, \tag{3.20}$$

and

$$\Gamma^\dagger_{i\mu}h\Gamma_{i\mu} \in U(\alpha_i), \quad h\Gamma_{i\mu} = \Gamma_{i\mu}f'_{i\mu}(h), \quad \forall\ h\in H_\mu, \tag{3.21}$$

it follows that the maps $f'_{i\mu}, f'_{i\mu}$, and $f''_{i\mu}$ (in fact $f''_{i\mu}$ is induced by the bundle projections on Q_μ and $\phi^{0*}_{i\mu}R_{\alpha_i n}$ from $f_{i\mu}$ and $f'_{i\mu}$) do indeed define a bundle homomorphism of Q_μ to $\phi^{0*}_{i\mu}R_{\alpha_i n}$ [i.e., that $f_{i\mu}(x,gh) = f_{i\mu}(x,g)f'_{i\mu}(h)$, $\forall\ (x,g) \in Q_\mu$, $h\in H_\mu$ and that $f'_{i\mu}$: $H_\mu \to U(\alpha_i)$ is a group homomorphism]. The group homomorphism $f'_{i\mu}$ induces a map, which we also denote $f'_{i\mu}: \mathscr{H}_\mu \to \mathscr{U}(\alpha_i)$, between the Lie algebras of H_μ and $U(\alpha_i)$, and is given by

$$f'_{i\mu}(A) = \Gamma^\dagger_{i\mu}A\Gamma_{i\mu} \in \mathscr{U}(\alpha_i), \quad \forall\ A\in\mathscr{H}_\mu. \tag{3.22}$$

Here $\mathscr{U}(\alpha)$ is represented by all complex $\alpha\times\alpha$ Hermitian matrices since the map $f''_{i\mu}$, being the identity on M_μ, is a diffeomorphism, it follows by Proposition 6.1, p. 79, Ref. 7, that the bundle homomorphism above and the connection on Q_μ uniquely induce a connection on $\phi^{0*}_{i\mu}R_{\alpha_i n}$, for each i. Thus, recalling that the connection form on Q_μ is $(\epsilon^*_\mu\omega)_{\mathscr{H}_\mu}$, using the expression for the induced connection form given in Proposition 6.1, p. 79, Ref. 7 and using Eq. (3.22), the following result is established.

Proposition 3.2: For each i the bundle $\phi^{0*}_{i\mu}R_{\alpha_i n}$ over M_μ has a connection induced on it from the connection on the bundle P over M. Letting the respective connection forms be $\omega^{(i)}$ and ω, then $\omega^{(i)}$ is determined from ω by

$$f^*_{i\mu}\omega^{(i)} = f'_{i\mu}((\epsilon^*_\mu\omega)_{\mathscr{H}_\mu}) = \Gamma^\dagger_{i\mu}(\epsilon^*_\mu\omega)\Gamma_{i\mu}. \tag{3.23}$$

A tensorial 0-form on P_μ of type $(\mathrm{ad},\mathbb{C}^{n\times n})$ is defined by the map: $\phi_{i\mu} = p_{i\mu}\circ\phi_\mu: P_\mu \to G_{\alpha_i n} \subset \mathbb{C}^{n\times n}$, for each i. Since $\phi_\mu\circ\delta_\mu: Q_\mu \to \hat{\mathscr{G}}_\mu \subset \mathbb{C}^{n\times n}$ is a tensorial 0-form of type $(\mathrm{ad},\mathbb{C}^{n\times n})$ on Q_μ which is constant, its value being A^0_μ, we have the following proposition for $\phi_{i\mu}$.

Proposition 3.3: For each i, $\delta^*_\mu\phi_{i\mu}: Q_\mu \to G_{\alpha_i n} \subset \mathbb{C}^{n\times n}$ is a constant tensorial 0-form on Q_μ of type $(\mathrm{ad},\mathbb{C}^{n\times n})$, its constant value being $\Theta_{i\mu}$, i.e., $\delta^*_\mu\phi_{i\mu} = \Theta_{i\mu}$.

Here, let us make the general remark that for any tensorial 0-form f of type (ρ,V) on any principal fiber bundle with structure group G and on which a connection is defined with connection form ω, letting ρ also denote the induced representation of the Lie algebra $\rho: \mathscr{G} \to \mathscr{L}(V,V)$, then the exterior covariant derivative of f, Df, may be expressed in the form: $Df(X) = df(X) + \rho(\omega(X))f$, where X is any tangent vector field on the bundle. In particular, when ρ is defined by the adjoint action of G on $V = \mathscr{L}(V',V')$ for some vector space V', then $Df(X) = df(X) + [\omega(X), f]$ (here d is the exterior derivative on the bundle manifold).

Proposition 3.4: For each i, $\delta^*_\mu D\phi_{i\mu}(X) = [\omega(\epsilon_{\mu *}X), \Theta_{i\mu}]$, where X is any tangent vector field on Q_μ.

Proof: $D\phi_{i\mu}(Y) = d\phi_{i\mu}(Y) + [i^*_\mu \omega(Y), \phi_{i\mu}]$, for any tangent vector field Y on P_μ. Now consider $\delta^*_\mu D\phi_{i\mu}$ and use Proposition 3.3.

For convenience we denote $\Gamma^\dagger_{i\mu}(\cdot)\Gamma_{i\mu}$ by $(\cdot)_{i\mu}$.

Lemma 3.5: For any $A,B \in \mathbb{C}^{n\times n}$

$$([A,B])_{i\mu} - [(A)_{i\mu}(B)_{i\mu}] = -([[A,\Theta_{i\mu}],[B,\Theta_{i\mu}]])_{i\mu}. \tag{3.24}$$

Proof: Suppressing the subscript on $\Theta_{i\mu}$ and letting RHS and LHS denote right and left hand sides, then for Eq. (3.24) we have: RHS $= -(A\Theta B\Theta + \Theta A\Theta B - A\Theta^2 B - \Theta AB\Theta - B\Theta A\Theta - \Theta B\Theta A + B\Theta^2 A + \Theta BA\Theta)_{i\mu}$. But by Eq. (3.20): $\Theta^2 = \Theta$, $\Theta\Gamma_{i\mu} = \Gamma_{i\mu}$, $\Theta = \Gamma_{i\mu}\Gamma^\dagger_{i\mu}$; thus, RHS $= (AB - BA - A\Gamma_{i\mu}\Gamma^\dagger_{i\mu}B + B\Gamma_{i\mu}\Gamma^\dagger_{i\mu}A)_{i\mu} =$ LHS.

We are now in a position to state and prove the main result of this section which concerns the generalization of the formula employed by 't Hooft.

Theorem 3.6: The curvature form on the bundle $\phi^{0*}_{i\mu}R_{\alpha,n}$ associated with the connection form $\omega^{(i)}$, $\Omega^{(i)}$, is related to the curvature form Ω on the bundle P by

$$f^*_{i\mu}\Omega^{(i)}(X,Y) = (\epsilon^*_\mu\{\Omega(X,Y) + \tfrac{1}{2}[D\phi_{i\mu}(X), D\phi_{i\mu}(Y)]\})_{i\mu}, \tag{3.25}$$

where X, Y are any tangent vector fields on Q_μ.

Proof: Let $X' = f_{i\mu *}X, Y' = f_{i\mu *}Y$, the curvature on $\phi^{0*}_{i\mu}R_{\alpha,n}$ is given by (see Ref. 7): $\Omega^{(i)}(X',Y') = d\omega^{(i)}(X',Y') + \tfrac{1}{2}[\omega^{(i)}(X'), \omega^{(i)}(Y')]$. Now using expression (3.23) for $\omega^{(i)}$ implies $\Omega^{(i)}(X',Y') = (\epsilon^*_\mu(d\omega(X,Y)))_{i\mu} + \tfrac{1}{2}[(\epsilon^*_\mu\omega(X))_{i\mu}, (\epsilon^*_\mu\omega(Y))_{i\mu}] = (\epsilon^*_\mu\Omega(X,Y))_{i\mu} - \tfrac{1}{2}\{([\epsilon^*_\mu\omega(X), \epsilon^*_\mu\omega(Y)])_{i\mu} - [(\epsilon^*_\mu\omega(X))_{i\mu}, (\epsilon^*_\mu\omega(Y))_{i\mu}]\}$. By Lemma 3.5, then, $\Omega^{(i)}(X',Y') = (\epsilon^*_\mu\Omega(X,Y))_{i\mu} + \tfrac{1}{2}([[\epsilon^*_\mu\omega(X),\Theta_{i\mu}],[\epsilon^*_\mu\omega(Y),\Theta_{i\mu}]])_{i\mu}$. Applying Proposition 3.4 to this last expression proves the theorem.

Having thus obtained for each i the expression for the curvature $\Omega^{(i)}$ on the bundle $\phi^{0*}_{i\mu}R_{\alpha,n}$, which we recognize as corresponding to the generalization of the 't Hooft electromagnetic field, we shall conclude this section by discussing the role it plays. Let q_i denote the bundle projection on $\phi^{0*}_{i\mu}R_{\alpha,n}$, for each i, let $H_*(\cdot, A)$, and $H^*(\cdot, A)$ denote homology and cohomology spaces over a ring A, respectively, and let $H^*_{\mathrm{DR}}(\cdot)$ denote the de Rham cohomology space $[H^*_{\mathrm{DR}}(\cdot) \cong H^*(\cdot,\mathbb{R})]$. The first Chern class on the associated bundle $\phi^{0*}_{i\mu}E_{\alpha,n}$, for each i, $c_1(\phi^{0*}_{i\mu}E_{\alpha,n}) \in H^2_{\mathrm{DR}}(M_\mu)$ can be represented by the closed 2-form $\gamma^{(i)}$ on M_μ, given by (see Theorem 3.1, p. 307, Ref. 8)

$$q^*_i\gamma^{(i)} = [(-1)^{1/2}/2\pi]\operatorname{tr}\Omega^{(i)}. \tag{3.26}$$

From Eq. (3.25), $\operatorname{tr}\Omega^{(i)}$ is given by the expression

$$f^*_{i\mu}\operatorname{tr}\Omega^{(i)}(X,Y) = \epsilon^*_\mu(\phi_{i\mu}, \{\Omega(X,Y) + \tfrac{1}{2}[D\phi_{i\mu}(X), D\phi_{i\mu}(Y)]\}), \tag{3.27}$$

where X, Y are any vector fields on Q_μ. Since $\gamma^{(i)}$ is a closed 2-form, it defines a conserved current (in the physical sense). We may obtain another conserved current on M_μ by using the following property (see Ref. 8, p. 306)

$$c_1(\phi^{0*}_{i\mu}E_{\alpha,n}) = \phi^{0*}_{i\mu}c_1(E_{\alpha,n}), \tag{3.28}$$

where, here, $\phi^{0*}_{i\mu}$ is also used to denote the map on cohomology, i.e., $\phi^{0*}_{i\mu}: H^*(G_{\alpha,n},\cdot) \to H^*(M_\mu,\cdot)$. If we now represent $c_1(E_{\alpha,n})$, the generator of $H^2(G_{\alpha,n},\mathbb{Z}) \cong \mathbb{Z}$ by $k_{\alpha,n}$, a closed 2-form on $G_{\alpha,n}$, then, for each $i, \phi^{0*}_{i\mu}k_{\alpha,n}$ gives a closed 2-form on M_μ different from $\gamma^{(i)}$, in general. To be specific, since $G_{\alpha n}$ is a Kähler manifold, we take $k_{\alpha n}$ to be proportional, by a numerical constant, to the Kähler form on $G_{\alpha n}$. To make $k_{\alpha n}$ explicit, we must obtain coordinates on $G_{\alpha n}$. For the open neighborhood U given by all elements A in $G_{\alpha n}$ such that

$$A = \begin{bmatrix} B \\ C \end{bmatrix}(B^\dagger C^\dagger), \tag{3.29}$$

where $B \in \mathbb{C}^{\alpha\times\alpha}$, $C \in \mathbb{C}^{(n-\alpha)\times\alpha}$, and $\det B \neq 0$, coordinates are defined by the $\alpha(n-\alpha)$ components of the matrix $T = CB^{-1} \in \mathbb{C}^{(n-\alpha)\times\alpha}$ (similarly for other open neighborhoods). On U then

$$k_{\alpha n} = \xi\partial\bar{\partial}\log\ \det(\mathbb{1}_\alpha + T^\dagger T), \tag{3.30}$$

where ∂, $\bar{\partial}$, respectively, denote the exterior derivative with respect to the components of T and their complex conjugate (see Ref. 8, p. 160). The constant ξ is fixed by putting

$$\int_{C^0_{\alpha n}} k_{\alpha n} = 1, \tag{3.31}$$

where $C^0_{\alpha n}$ denotes the closed 2-cycle in $G_{\alpha n}$, given by all elements $AA^\dagger$ in $G_{\alpha n}$, where $A \in \mathbb{C}^{n\times\alpha}$ is such that

$$A = \begin{bmatrix} z_1 \\ z_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ for } \alpha = 1,$$

$$A = \begin{bmatrix} z_1 & -z_2^* & 0 & \cdots & 0 \\ z_2 & z_1^* & & & \\ 0 & & 1 & & \\ \vdots & & & \ddots & \\ 0 & & & & 1 \\ 0 & & & & 0 \\ \vdots & & & & \\ 0 & & & & 0 \end{bmatrix} \quad \text{for } \alpha > 1, \tag{3.32}$$

for any $z_1,z_2 \in \mathbb{C}$ with $z_1z_1^* + z_2z_2^* = 1$ ($\mathrm{C}^0_{\alpha n}$ defines a copy of the one-dimensional complex projective line in $G_{\alpha n}$). Both expressions $\gamma^{(i)}$ and $\phi^{0*}_{i\mu}k_{\alpha,n}$ have appeared in the physics literature for the SU(2) case, and the identity relating them there shown in Ref. 9 we now generalize by writing

$$C_1(\phi^{0*}_{i\mu}E_{\alpha,n}) = [\gamma^{(i)}] = \phi^{0*}_{i\mu}[k_{\alpha,n}], \tag{3.33}$$

(we let [·] denote cohomology or homology class). That the latter expression in Eq. (3.33) does not contain the connection form ω follows from the general property that Chern classes are independent of the connection. Since conserved currents were obtained on M_μ in two steps by using $p^*_{i\mu}$ and ϕ^{0*}_μ to map $H^2_{\rm DR}(G_{\alpha_i n})$ onto $H^2_{\rm DR}(M_\mu)$, i.e.,

$$H^2_{\rm DR}(M_\mu)\xleftarrow{\phi^{0*}_\mu}H^2_{\rm DR}(\hat{\mathscr{G}}_\mu)\xleftarrow{p^*_{i\mu}}H^2_{\rm DR}(G_{\alpha_i n}) \tag{3.34}$$

it is clear that all possible conserved currents are obtained if $\mathscr{P}=\{p^*_{i\mu}c_1(E_{\alpha_i n})\}$ generates $H^2_{\rm DR}(\hat{\mathscr{G}}_\mu)$. We show in an appendix that this is indeed true. However, $\mathscr{P}$ in general will not be a linearly independent set; in particular for SU(n) (see Appendix)

$$\sum_i p^*_{i\mu}c_1(E_{\alpha_i n})=0. \tag{3.35}$$

Using the 2-forms $\gamma^{(i)}$ or $\phi^{0*}_{i\mu}k_{\alpha_i n}$, an integer can be associated, for each i, with any closed, compact, connected two-dimensional submanifold without boundary, i.e., any elementary 2-cycle N in M_μ, for some μ. Since $[N]$ generates $H_2(N,\mathbb{Z})\cong\mathbb{Z}$ and $[C^0_{\alpha_i n}]$ generates $H_2(G_{\alpha_i n},\mathbb{Z})\cong\mathbb{Z}$ and since the map $\phi^0_{i\mu}\circ i_N: N\to G_{\alpha_i n}$ induces the map $(\phi^0_{i\mu}\circ i_n)_*$: $H_2(N,\mathbb{Z})\to H_2(G_{\alpha_i n},\mathbb{Z})$ on homology where i_N denotes the inclusion $i_N: N\hookrightarrow M_\mu$ we can write

$$(\phi^0_{i\mu}\circ i_N)_*[N]=r_i(\phi^0_\mu,N)[C^0_{\alpha_i n}], \tag{3.36}$$

for each i, where $r_i(\phi^0_\mu,N)$ is an integer given by any one of the following expressions:

$$\begin{aligned}r_1(\phi^0_\mu,N)&=\int_{(\phi^0_{i\mu}\circ i_N)_*N}k_{\alpha_i n}\\&=\int_N(\phi^0_{i\mu}\circ i_N)^*k_{\alpha_i n}=\int_N i^*_N\gamma^{(i)}.\end{aligned} \tag{3.37}$$

Since the integers $r_i(\phi^0_\mu,N)$, $i=1,\dots,l$, depend only on the homology class $i_{N*}[N]\in H_2(M_\mu,\mathbb{R})$, they will be unchanged if N is replaced by any other elementary 2-cycle which, with N, defines the boundary of a 3-cycle in M_μ. Moreover, $r_i(\phi^0_\mu,N)$, $i=1,\dots,l$, will remain unaltered under a deformation of the map $\phi^0_\mu\circ i_N: N\to\hat{\mathscr{G}}_\mu$, which leaves $(\phi^0_{i\mu}\circ i_N)_*[N]\in H_2(G_{\alpha_i n},\mathbb{R})$ unchanged for all i. For SU(n), since (see Appendix) there are only $l-1$ independent generators of $H^2(\hat{\mathscr{G}}_\mu,\mathbb{R})$ [$\mu=(\alpha_1,\dots,\alpha_l)$], we may expect a constraint to exist on the l integers $r_i(\phi^0_\mu,N)$, $i=1,\dots,l$ in this case. Because of Eq. (3.35) this constraint is given by

$$\sum_i r_i(\phi^0_\mu,N)=0. \tag{3.38}$$

Finally let us remark that, for some μ and for sufficiently large R, $V_R=\{x\in M\,|\,\langle x,x\rangle>R\}$ must be contained in M_μ, and that specifying the integers $r_i(\phi^0_\mu,S^2_{R'})$, $i=1,\dots,l$, and some R' such that $R<R'<\infty$ is equivalent to specifying the homotopy class $\alpha\in\pi_2(Z(U))$ defined by the configuration $(\phi,\omega)\in W$.[10]

IV. FURTHER PROPERTIES ARISING FROM THE VARIATIONAL EQUATIONS

In this section we consider some analytical properties which can be inferred for the invariants (or tensorial 0-forms) constructed from the Higgs form, as a consequence of the variational equations (2.3). These properties take the form of differential inequalities satisfied by, and involving only, the above invariants, first on all of P and then on P_μ for a certain μ (specified below).

We define the tensorial 0-forms of type $(0,\mathbb{R})$ on P by $\theta_r=(\phi^r,\phi^r)$, $r\in\mathbb{N}$. Only a finite number of the θ_r's are algebraically independent, this number being given by the rank of G.

Theorem 4.1: At every point of M

$$\Delta_0\theta_r-2r(\phi^r,\phi^{r-1}\Delta_\omega\phi)\geqslant0,\quad\forall\ r\in\mathbb{N}. \tag{4.1}$$

Proof: We have the following identity: $\Delta_0\theta_r=r\Sigma^{2r-2}_{s=0}(A_s,A_s)+(\phi^{r-1}D\phi,\phi^{r-1}D\phi)+2r(\phi^r,\phi^{r-1}\Delta_\omega\phi)$, where $A_s=\phi^{r-s-2}D\phi\phi^{s+1}+\phi^{r-s-1}D\phi\phi^s$. Since $(A_s,A_s)\geqslant0$ and $(\phi^{r-1}D\phi,\phi^{r-1}D\phi)\geqslant0$, the result follows.

When the variational equations (2.3) are satisfied, Eq. (4.1) gives differential inequalities only involving the θ_r's.

Corollary 4.2: At every point of M [$\alpha_r=1$ for SU(n); $a_r=(-1)^{r-1}$ for SO(n)]

$$\Delta_0\theta_r-2ra_r\left[\frac{d}{dt}U(\phi+t\phi^{2r-1})\right]_{t=0}\geqslant0,\quad\forall\ r\in\mathbb{N}. \tag{4.2}$$

We consider now the submanifolds M_μ and P_μ of M and P, respectively, where μ is given by: $(1,\dots,1)$ for SU(n); $(2,\dots,2,1)$ for SO(n), n odd; and $(2,\dots,2,+)$ for SO(n), n even. In each case $\hat{\mathscr{G}}_\mu$ defines the generic stratum of the group action and the label μ defines the interior of the fundamental Weyl chamber [i.e., without the walls; see [*Note*] after Eq. (3.12)]. For the remainder of this section μ will be as just defined. On P_μ we have tensorial 0-forms ψ, ψ_i, ζ, ζ_i, and ζ_0 of type $(\mathrm{ad},\mathbb{F}^{n\times n})$ [$\mathbb{F}=\mathbb{C}$ for SU(n); $\mathbb{F}=\mathbb{R}$ for SO(n)], which are given as follows: for SU(n), $\zeta=\psi=\phi\circ i_\mu$ and $\zeta_i=\psi_i=\phi_{i\mu}$, $i=1,\dots,n$; for SO(n), $\psi=\phi\circ i_\mu$, $\zeta=\psi^T\psi$, $\psi_i=-2\,{\rm Im}\phi_{i\mu}$, $\zeta_i=\psi_i^T\psi_i=-2\,{\rm Re}\phi_{i\mu}$, $i=1,\dots,[n/2]$, and when n is odd $\zeta_0=\mathbb{1}_n-\Sigma^{[n/2]}_{i=1}\zeta_i$ [i.e., ζ_0 is nontrivial only for SO(n), n odd]. We then have the invariants or tensorial 0-forms on P_μ of type $(0,\mathbb{R})$, $\lambda_i=(\psi_i,\psi)$, $a_i=(\zeta_i,\zeta)$, $i\neq0$, which satisfy: for SU(n), $a_i=\lambda_i$, $\Sigma_i\lambda_i=0$, $\lambda_1>\cdots>\lambda_n$; for SO($n$), $n=2j$, $a_i=\lambda^2_i$, $\lambda_1>\cdots>\lambda_{j-1}>|\lambda_j|$; and for SO($n$), $n=2j+1$, $\alpha_i=\lambda^2_i$, $\lambda_1>\cdots>\lambda_j>0$. Noting that the space of tensorial 0-forms of type $(\mathrm{ad},\mathbb{F}^{n\times n})$ on P_μ is closed under linear summation and product, we state the following further properties: $\psi=\Sigma_i\lambda_i\psi_i$; $\zeta=\Sigma_{i\neq0}a_i\zeta_i$; $\Sigma_i\zeta_i=\mathbb{1}_n$; $\zeta_i\zeta_j=\delta_{ij}\zeta_i$, $\forall\ i,j$; $(\zeta_i,\zeta_j)=\delta_{ij}$, $(i,j)\neq(0,0)$, $(\zeta_0,\zeta_0)=\frac{1}{2}$; $\{\zeta_i\}$ are Hermitian and are self-adjoint with respect to $(\cdot,\cdot)$ [in fact, $(\zeta_iA,B)=(A,\zeta_iB)$, $(A\zeta_i,B)=(A,B\zeta_i)$, $\forall\ A,B\in\mathbb{F}^{n\times n}$]. For convenience we let D denote the covariant derivative on P_μ and $\Delta=\Delta_{i^*_\mu\omega}$.

Proposition 4.3: (i) $(\zeta_i,D\zeta_i)=0$.

(ii) $(\zeta_i,\Delta\zeta_j)=-(D\zeta_i,D\zeta_j)$.

(iii) $(D\zeta_i,D\zeta_j)=2\{(D\zeta_i,D\zeta_i)\delta_{ij}-(D\zeta_i,\zeta_jD\zeta_i)\}$.

(iv) $(D\zeta_i,\zeta_jD\zeta_i)=(D\zeta_j,\zeta_iD\zeta_j)\geqslant0$.

Proof: (i) and (ii) are obvious. (iii) Taking the derivative of $\zeta^2_i=\zeta_i$ implies $D\zeta_i=\zeta_iD\zeta_i+D\zeta_i\zeta_i$. The inner product of this equation with $D\zeta_j$ then gives the result when $i=j$. For $i\neq j$, taking the inner product of this equation with $D\zeta_i$ and using the fact that $\zeta_i\zeta_j=0$ implies $\zeta_iD\zeta_j=-D\zeta_i\zeta_j$ gives

$$(D\zeta_i,D\zeta_j) = (\zeta_i D\zeta_i,D\zeta_j) + (D\zeta_i\zeta_i,D\zeta_j)$$
$$= (D\zeta_i,\zeta_i D\zeta_j) + (D\zeta_i,D\zeta_j\zeta_i) = -2(D\zeta_i,\zeta_j D\zeta_i),$$

which is the result for $i\neq j$. (iv) follows from (iii) and the fact that $(D\zeta_i,\zeta_j D\zeta_i) = (\zeta_j D\zeta_i,\zeta_j D\zeta_i)\geqslant 0$.

Lemma 4.4: (i) For SU(n) and SO(n'), n' even, $\Delta_0 a_i = 2\Sigma_{j\neq i}\,(a_i - a_j)\,(D\zeta_i,\,\zeta_j D\zeta_i) + (\zeta_i,\Delta\zeta)$.

(ii) For SO(n), n odd, $\Delta_0 a_i = 2\Sigma_{j\neq i>0}\,(a_i - a_j)\,(D\zeta_i,\,\zeta_j D\zeta_i) + 2a_i\,(D\zeta_i,\,\zeta_0 D\zeta_i)\; + (\zeta_i,\Delta\zeta),\, i>0$.

Proof: Differentiate $a_i = (\zeta_i,\zeta)$, using Proposition 4.3 and the fact that $\Sigma_i\,\zeta_i = \mathbb{1}_n$. By Proposition 4.3(iv) and the previous lemma we get:

Theorem 4.5:

(i) For SU(n), $\Delta_0\,(\Sigma^r_{i=1}\,\lambda_i) - (\Sigma^r_{i=1}\,\zeta_i,\Delta\zeta)\geqslant 0$, $r = 1,\ldots,n-1$.

(ii) For SO(n), $\Delta_0\,(\Sigma^r_{i=1}\,\lambda^2_i) - 2\Sigma^r_{i=1}\,\lambda_i(\psi_i,\Delta\psi)\geqslant 0$, $r = 1,\ldots,[n/2]$.

On application of the variational equations (2.3), we obtian the following differential inequalities on M_μ for $\{\lambda_i\}$.

Corollary 4.6: (i) For SU(n), $\Delta_0\,(\Sigma^r_{i=1}\,\lambda_i) - [(d/dt)U(\psi + t\,\Sigma^r_{i=1}\,\psi_i)]_{t=0}\geqslant 0$, $i = 1,\ldots,n-1$.

(ii) For SO(n), $\Delta_0\,(\Sigma^r_{i=1}\,\lambda^2) - 2[(d/dt)U(\psi + t\Sigma^r_{i=1}\,\lambda_i\psi_i)]_{t=0}\geqslant 0$, $i = 1,\ldots,[n/2]$.

As an application, let us consider now the limit of vanishing Higgs potential $U\rightarrow 0$. Corollary 4.2 in this case becomes: $\Delta_0\theta_r\geqslant 0$, $\forall\ r\in\mathbb{N}$, i.e., θ_r for each r, is a subharmonic function. The asymptotic condition on the invariants θ_r (replacing the finite energy constraint for $U\not\equiv 0$) is that, for each $r = 1,\ldots,\mathrm{rank}(\mathscr{G})$, $\theta_r(x)$ tends to a finite limit $(\theta_r)_\infty$, as $\langle x,x\rangle\rightarrow\infty$. Thus at each point x of M

$$\{\Delta_0[(\theta_r)_\infty - \theta_r]\}(x)\leqslant 0,\quad\forall\ r\in\mathbb{N}. \tag{4.3}$$

By multiplying Eq. (4.3) by the Green's function $[16\pi^2\langle(x-y),(x-y)\rangle]^{-1/2}\geqslant 0$ and integrating over all points x in M, the following pointwise bounds on the Higgs invariants may be obtained:

$$0\leqslant\theta_r(y)\leqslant(\theta_r)_\infty,\forall\ r\in\mathbb{N}. \tag{4.4}$$

Previously, only the $r = 1$ case of Eq. (4.4) had been shown.[2]

V. SUMMARY AND CONCLUSIONS

The subject matter of this paper may be divided into two main parts, corresponding to Secs. III and IV. In Sec. III we establish the generalization of the 't Hooft field, for gauge groups of physical interest larger than SU(2), as the curvature forms induced on vector bundles $E_{\alpha,n}$ over submanifolds M_μ of Euclidean 3-space M. The statement of this result is given in Theorem 3.6. In the remainder of Sec. III we show how, via the first Chern classes of the bundles $E_{\alpha,n}$, to associate sets of integer charges with (almost) any closed surface. In Sec. IV we obtain differential inequalities for, and involving only, the Higgs invariants which follow from the variational equations (2.3) for arbitrary Higgs potential. These inequalities are of two kinds, the first kind hold on all of M and the second on a specific submanifold M_μ. The statements of these results correspond to Corollaries 4.2 and 4.6, respectively. In an application to the limit of vanishing Higgs potential, new bounds are obtained on the Higgs invariants [Eq. (4.4)].

ACKNOWLEDGMENTS

I would like to thank L. Michel for many illuminating discussions throughout the course of this work. Also I would like to thank J. Hayden for his help especially with the proof presented in the Appendix. Financial support is gratefully acknowledged from the following institutions: Scuola Internazionale Superiori di Studi Avanzati, Trieste, Italy; Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France; and CERN, Geneva, Switzerland. Finally, it is a pleasure to thank L. Michel and N. Kuiper for hospitality shown while visiting the IHES.

APPENDIX

Here it is shown that $\{c_1(p^*_{i\mu}E_{\alpha,n})\}$ generate $H^2(\hat{\mathscr{G}}_\mu)$ [throughout the Appendix $H^*(\cdot)$ will mean $H^*(\cdot,\mathbb{R})\cong H^*_{\mathrm{DR}}(\cdot)$]. First we introduce a more convenient notation for this purpose than the one used in the text. Let $X(\beta_1,\ldots,\beta_l;\sigma)$ consist of all elements $(\gamma_1,\ldots,\gamma_l) = (p_{l\mu}(A),\ldots,p_{1\mu}(A))$, where $A\in\hat{\mathscr{G}}_\mu$ and where $\mu = (\beta_l,\ldots,\beta_1;\sigma)$ [i.e., $\beta_i = \alpha_{l-i+1}$ for SU(n), $\beta_i = 2\alpha_{l-i+1}$ and $\sigma = +,\,-$, or β_0 for SO(n) and $n = \Sigma_i\beta_i$]. Clearly $\hat{\mathscr{G}}_\mu$ and $X(\beta_1,\ldots,\beta_l;\sigma)$ may be identified. For $l>1$ we have the fiber map $\pi_l: X(\beta_1,\ldots,\beta_l;\sigma)\ni(\gamma_1,\ldots,\gamma_l)\rightarrow\gamma_l\in B_{l,l}(\beta_l)$ [$B_{l,l}(\beta_l)\subset G_{\alpha_1 n}$ in the notation used in the text] with typical fiber $X(\beta_1,\ldots,\beta_{l-1};\sigma)$. We let i_l denote the inclusion map

$$i_l: X(\beta_1,\ldots,\beta_{l-1};\sigma)\ni(\gamma_1,\ldots\gamma_{l-1})$$
$$\hookrightarrow\left(\begin{bmatrix}0 & 0\\ 0 & \gamma_1\end{bmatrix},\ldots,\begin{bmatrix}0 & 0\\ 0 & \gamma_{l-1}\end{bmatrix},\,\Theta_{i\mu}\right)\in X(\beta_1,\ldots,\beta_l:\sigma). \tag{A1}$$

We also define the maps $p^{(l)}_i: X(\beta_1,\ldots,\beta_l;\sigma)\ni(\gamma_1,\ldots,\gamma_l)\rightarrow\gamma_i\in B_{i,l}(\beta_i)\;(p^{(l)}_l = \pi_l)$.

Each of the spaces $X(\beta_1,\ldots,\beta_l;\sigma)$ and $B_{i,l}(\beta_i)$ may be considered as an orbit of the adjoint action of the group G. For any such orbit $G(A)$, $A\in\mathscr{G}$, it has been shown[11] that $\dim[H^2(G(A))] = \dim(\text{center of } G_A)$. Hence $H^2(B_{i,l}(\beta_i))\cong\mathbb{R}$ and $H^2(X(\beta_1,\ldots,\beta_l;\sigma))\cong\mathbb{R}^d$, where $d = l-1$ for SU(n) and $d = l$ for SO(n). Moreover, both $B_{i,l}(\beta_i)$ and $X(\beta_1,\ldots,\beta_l;\sigma)$ are 0- and 1-connected and since all orbits $G(A)$, $A\in\mathscr{G}$, are Kähler manifolds,[12] it follows that $H^r(B_{i,l}(\beta_i)) = H^r(X(\beta_1,\ldots,\beta_l;\sigma)) = 0$ for r an odd integer. By use of Theorem 10.6.2C, p. 431, of Ref. 12, we may write the exact sequence.

Lemma A.1:

$$0\leftarrow H^2(X(\beta_1,\ldots,\beta_{l-1};\sigma))\xleftarrow{i^*_l}H^2(X(\beta_1,\ldots,\beta_l;\sigma))$$
$$\xleftarrow{\pi^*_l}H^2(B_{l,l}(\beta_l))\leftarrow 0. \tag{A2}$$

Letting $H^2(B_{i,l}(\beta_i))$ be generated by $k^{(l)}_i$, $i = 1,\ldots,l$ [$k^{(l)}_i$ represents the first Chern class of the vector bundle over $B_{i,l}(\beta_i)$, denoted $E_{\alpha,n}$ in the text]. Our task now reduces to proving the following theorem.

Theorem A.2: $\{p^{(l)*}_i k^{(l)}_i\}_{i\leqslant l}$ generates $H^2(X(\beta_1,\ldots,\beta_l;\sigma))$.

Proof: We used induction on l. The initial case [$l = 2$ for SU(n), $l = 1$ for SU(n)] is obviously true. Now let us define the inclusion maps

$$j^{(l)}_i:\; B_{i,l-1}(\beta_i)\ni\gamma\hookrightarrow\begin{bmatrix}0 & 0\\ 0 & \gamma\end{bmatrix}\in B_{i,l}(\beta_i),\quad i = 1,\ldots,l-1. \tag{A3}$$

Hence the following diagram commutes, for each i, $i = 1,\ldots,l-1$:

$$\begin{tikzcd} X(\beta_1,\ldots,\beta_{l-1};\sigma) \arrow[r, "i_l"] \arrow[d, "p_i^{(l-1)}"] & X(\beta_1,\ldots,\beta_l;\sigma) \arrow[d, "p_i^{(l)}"] \\ B_{i,l-1}(\beta_i) \arrow[r, "j_i^{(l)}"] & B_{i,l}(\beta_i) \end{tikzcd} \tag{A4}$$

i.e., $j_i^{(l)} \circ p_i^{(l-1)} = p_i^{(l)} \circ i_l$. On 2-cohomology, $j_i^{(l)}$ is an isomorphism, i.e., $j_i^{(l)*}: H^2(B_{i,l}(\beta_i)) \cong \mathbb{R} \cong H^2(B_{i,l-1}(\beta_i)) \cong \mathbb{R}$, for $i = 1,\ldots,l-1$, so let us choose to have $j_i^{(l)*}k_i^{(l)} = k_i^{(l-1)}$. Hence, from Eq. (A4) $p_i^{(l-1)*}k_i^{(l-1)} = p_i^{(l-1)*}j_i^{(l)*}k_i^{(l)}$ $= i_l^* p_i^{(l)*} k_i^{(l)}$, for $i = 1,\ldots,l-1$. Making the inductive hypothesis for the $l-1$ case, it follows that $\{i_l^* p_i^{(l)*} k_i^{(l)}\}$ generates $H^2(X(\beta_1,\ldots,\beta_{l-1};\sigma))$. By the exactness of the sequence (A2), $\pi_l^* k_l^{(l)}$ must define an independent element in $H^2(X(\beta_1,\ldots,\beta_l;\sigma))$ to $\{p_i^{(l)*}k_i^{(l)}\}_{i<l-1}$. Since $\dim[H^2(X(\beta_1,\ldots,\beta_l;\sigma))] = \dim[H^2(X(\beta_1,\ldots,\beta_{l-1};\sigma))] + 1$ and since $\pi_l = p_l^{(l)}$, the proof by induction is complete.

Finally, let us remark that for SU(n) only $l-1$ elements of $\{c_1(p_{i\mu}^* E_{\alpha,n})\}_{i<l}$ (in the notation of the text) are linearly independent and that the following relation is satisfied:

$$\sum_i c_1(p_{i\mu}^* E_{\alpha,n}) = 0. \tag{A5}$$

This equation is a consequence of the Whitney sum formula and the fact that $\oplus_i E_{a,n}$ is a trivial vector bundle over $\hat{\mathscr{G}}_\mu$.

[1] E. Bogomolny, Sov. J. Nucl. Phys. **24**, 449 (1976).

[2] A. Jaffe and C. Taubes, *Monopoles and Vortices* (Birkhäuser, Boston, 1980).

[3] R. Ward, Comm. Math. Phys. **79**, 317 (1981); M. Prasad, Comm. Math. Phys. **80**, 137 (1981).

[4] R. Ward, Comm. Math. Phys. **80**, 563 (1981); E. Corrigan and P. Goddard, Comm. Math. Phys. **80**, 575 (1981).

[5] P. Houston and L. O'Raifeartaigh, Phys. Lett. B **93**, 151 (1980); P. Houston and L. O'Raifeartaigh, Phys. Lett. B **94**, 153 (1980); P. Houston and L. O'Raifeartaigh, Z. Phys. C **8**, 175 (1981).

[6] G. t' Hooft, Nucl. Phys. B **79**, 276 (1974).

[7] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969), Vol. 1.

[8] S. Kobayashi and K. Nomizu, *Foundations of Differential Geomtery* (Wiley, New York, 1969), Vol. 2.

[9] J. Arafune, P. Freund, and C. Goebel, J. Math. Phys. **16**, 433 (1975).

[10] A. Schwarz, Nucl. Phys. B **112**, 358 (1976); C. Taubes, Comm. Math. Phys. **81**, 299 (1981).

[11] R. Bott, in *Representation Theory of Lie Groups*, London Math. Soc. Lecture Notes, Series 34 (Cambridge U. P., Cambridge, 1979).

[12] R. Bott, Proc. Natl. Acad. Sci. USA **40**, 1147 (1954).

[13] P. Hilton and S. Wylie, *Homology Theory* (Cambridge U. P., Cambridge, 1960).

A particle representation for Korteweg–de Vries solitons

Graham Bowtell and Allen E. G. Stuart
Department of Mathematics, The City University, Northampton Square, London EC1V OHB, England

(Received 15 November 1979; accepted for publication 5 November 1982)

In an earlier paper we established an equivalence between the dynamics of interacting sine–Gordon solitons and the motions of poles of the corresponding Hamiltonian density. In particular, we found analytic expressions for the forces acting between the solitons and used these to represent the N-soliton solution as an N-body interaction between classical particles. In this paper, we apply the methods of our previous analysis to obtain a dynamically equivalent particle representation for interacting Korteweg–de Vries solitons. The representation is faithful and a detailed analysis is present for the one- and two-soliton solutions. In these cases the particle motions accurately reflect the behavior of the solitons, giving, respectively, a uniform motion and a repulsive interaction. Furthermore, in the case of the two-soliton solutions, the phase shifts calculated from the particle trajectories are the same as those obtained from an asymptotic analysis of the waveforms. Because of the nature of the Korteweg–de Vries equation, there are important differences between the present analysis and that employed for the sine–Gordon equation and these are discussed in some detail. A comparison with related work on *other* solutions of the Korteweg–de Vries is also presented.

PACS numbers: 11.10.Qr, 11.10.Lm, 11.80.Jy

I. INTRODUCTION

One of the many interesting aspects of the theory of nonlinear evolution equations is the field–particle duality that exists between the regular and signular parts of special solutions of these equations.[1,2] To be more precise, as the regular component of the solution evolves in time according to the field equation, the singular features, such as poles, etc., which lie on certain submanifolds of a (generally) complex domain, execute well-defined motions which can be identified with those of a classical many-body problem with two-body forces. This establishes a direct link between an infinite-dimensional system represented by the field equation and a finite-dimensional system of classical particles.

For example, if we consider the Korteweg–de Vries equation (KdV)

$$u_t + uu_x + u_{xxx} = 0, \quad u:(x,t)\in\mathbb{R}^2 \rightarrow u(x,t)\in\mathbb{R}, \tag{1.1}$$

then, as shown by Airault *et al.*[3] and the Chodnovsky brothers,[4] the motions of the poles of the rational and elliptic solutions of (1.1) can be related to certain solutions of many-body systems with the Hamiltonians

$$H = \frac{1}{2}\sum_{k=1}^{n}\dot{z}_k^2 + \frac{1}{2}\sum_{\substack{k,l\\ k\neq l}} V(z_{kl}), \quad z_k\in\mathbb{C}, \equiv \frac{d}{dt}, \tag{1.2}$$

where V are repulsive pair potentials which only depend on the (complex) inter particle separation $z_{kl} = z_k - z_l$. In the case of the rational solutions these potentials are proportional to z_{kl}^{-2} or z_{kl}^{-4}, while for the elliptic solutions $V \propto \mathscr{P}(z_{kl})$ or $\mathscr{P}^2(z_{kl})$, where $\mathscr{P}$ is a Weierstrassian elliptic function.[5]

In a similar manner, though from quite a different point of view to that adopted in Refs. 3 and 4, we have established[6] a strong correlation between the dynamics of interacting solitons of the sine–Gordon equation (SGE),

$$\phi_{xx} - \phi_{tt} = \sin\phi, \quad \phi:(x,t)\in\mathbb{R}^2 \rightarrow \phi(x,t)\in\mathbb{R} \bmod 2\pi, \tag{1.3}$$

and the motions of the poles of the corresponding Hamiltonian density,

$$\mathscr{H} = \frac{1}{2}\left[(\phi_{ns})_t^2 + (\phi_{ns})_x^2 + 2(1 - \cos\phi_{ns})\right],$$
$$\phi_{ns} \equiv n\text{-soliton solution}. \tag{1.4}$$

In this case the imaginary parts of the pole positions turned out to be constants and, hence, the projection onto the real axis gave us a one-dimensional, many-body problem (with a *real* phase space), which was also a one-to-one map, i.e., one particle per soliton. The two-body problem was worked out in detail and led to the pair potentials.

$$V_{ss}(x) = 8\gamma(1 - \tanh\gamma|x|), \tag{1.5a}$$

for the repulsive, soliton–soliton scattering state and

$$V_{sa}(x) = 8\gamma(1 - \coth\gamma|x|) \tag{1.5b}$$

for the attractive, soliton–antisoliton scattering state, where γ is the Lorentz factor and $|2x|$ the (real) distance between the particles in the center-of-mass frame.

Our success with the SGE prompted us to extend the analysis to some of the other well-known soliton equations, and in this paper we report on an attempt to obtain a particle representation for the KdV solitons. The study effectively brings together two earlier works on the problem. The first is a paper by Thickstun,[7] in which the analysis is motivated by a formal analogy between the soliton solutions and the velocity field of a set of dipoles in a perfect two-dimensional fluid, an analogy originally suggested by Kruskal,[8] while the second is a paper by ourselves,[9] which, in retrospect, turned out to be a very preliminary version of the present work.

Thickstun's analysis, although interesting in its own right, does not appear to us to be a satisfactory solution to the original problem. There are three reasons for this, two physical and one mathematical. First, the association of dipoles, which are vector particles, with the solutions of a scalar field

0022-2488/83/040969-13$02.50

equation is physically inconsistent, or at best, incomplete.[4] Secondly, the interpretation of the details of the soliton interactions in terms of "slipping dipoles" seems to be more of a conjecture than an assertion, since it is not supported by any quantitative arguments. Thirdly, and what is perhaps the most relevant criticism, the method used by Thickstun for finding the poles of the two-soliton solution has three major disadvantages: (1) It only leads to solutions for rational values of the square-root of the speed ratio; (2) it does not make use of the extra degrees of freedom which are available in the complex plane; (3) it does not lead to a faithful representation, i.e., a particle problem where the number of particles is the same as the number of solitons. In fact, the solutions obtained in Ref. 7 exhibit an unstable property in that the number of poles depends explicitly on the speed ratio of the solitons and small changes in the latter can produce extremely large changes in the former, whereas if a faithful representation exists, we expect it to the stable, i.e., in the case of the two-soliton solutions it should contain two and only two poles for *all* values of the speed ratio.

Now in our preliminary analysis,[9] we showed that it is possible to obtain such a faithful representation for the two-soliton solutions. However, the method we used was somewhat ad hoc, and, as a consequence, our results were mathematically incomplete (we missed the random poles) as well as being physically obscure (we worked with complex time). In the present study we remedy both these deficiencies. Our method is more general and, working with *real* time, we obtain both the faithful and random representations. The former is then used to discuss the dynamics of interacting solitons.

Our results indicate that KdV solitons repel each other and that the forces acting between them have a short range. Moreover, these forces are local in nature, rather than of the action-at-a-distance type, and are generated by the absorption and emission of energy and momentum by the underlying field. As a consequence, the solitons *exchange*, rather than preserve, their identities during the interaction. Note that this explanation of the manner in which the interaction occurs is based on the assumption that the faithful poles can be identified with the center of mass of the individual solitons in the two-soliton waveforms. It differs considerably from the "conventional" view and from an earlier interpretation given by Lax,[10] in which the solitons are identified by the "peaks" on the waveform. We shall elaborate on this in Sec. V.

From a general point of view the results obtained for the KdV are not as satisfactory as those we obtained for the sine–Gordon equation. This stems from the fact that, apart from the presence of random poles, the KdV does not have the right symmetry for a particle interpretation. Thus, in calculating the mass spectrum of the pole particles from the waveforms, we have had to abandon the Hamiltonian density for a phenomenologically defined mass density. Nevertheless, the results are interesting in their own right and provide, we think, a satisfactory picture of the translational mode dynamics of interactng KdV solitons.

The plan of the rest of the paper is as follows. In Sec. II, we compare and contrast the objectives and methods of our approach with those of Refs. 3 and 4. We feel that this comparison is important and, for pedagogical reasons, is better at the beginning rather than the end of the paper. In Sec. III we present our physical interpretation and compute the mass spectrum of the particles. Section IV deals with the kinematics of the polemotions, and Sec. V with the dynamics of the solitons. Our concluding remarks are contained in Sec. VI.

II. TWO APPROACHES TO DUALITY

Apart from establishing a duality between fields and particles, the examples cited above also illustrate two different ways of exploiting this duality. On the one hand, there are the mathematically motivated studies of Refs. 3 and 4, and on the other, the more physically oriented analyses of Refs. 6 and 7. The contrast between them is best explained via specific examples and can be sketched in the following manner.

First, from a formal point of view, the evolution equations can be used to generate and study solvable many-body problems. To see how this can be done, consider, for example, the following fairly simple nonlinear equation:

$$u_t + u_x + u^2 = 0, \quad u:(x,t)\in\mathbb{R}^2\rightarrow u(x,t)\in\mathbb{R}, \tag{2.1}$$

which has the general solution

$$u(x,t) = f(x-t)/[1+tf(x-t)], \quad u(x,0) = f(x), \tag{2.2}$$

where $f:\mathbb{R}\rightarrow\mathbb{R}$ is an arbitrary function. To study the motions of the singularities of (2.2), for general f, we continue u to the map $u(z,t):\mathbb{C}\times\mathbb{R}\rightarrow\mathbb{C}$ via the substitution $x\rightarrow z = x+iy$. Then, from (2.2), the orbits of these singularities are given by

$$z_i(t) = f_i^{-1}(-1/t) + t\in\mathbb{C}, \quad t\in\mathbb{R}, \tag{2.3}$$

where f_i^{-1} is the ith branch of f^{-1}. Thus, both the number and nature of the orbits (2.3) can be varied by adopting different choices for the initial profiles of (2.1).

Now the most interesting examples occur when the initial profile is chosen so that $u(z,t)$ is a meromorphic function. In this case the singularities manifest themselves as a finite, or at worst an infinitely countable, set of poles. The simplest nontrivial choice for our example is the four-parameter rational function

$$f(x) = \frac{1-\gamma}{x-\alpha} + \frac{1-\delta}{x-\beta}, \quad \alpha,\beta,\gamma,\delta\in\mathbb{R}, \tag{2.4}$$

which leads to the complex solution

$$u(z,t) = \sum_{i=1}^{2} [\mathbb{1} - \dot{z}_i(t)]/[z - z_i(t)], \quad \dot{z}_i \equiv \frac{dz_i}{dt}, \tag{2.5}$$

with an obvious identification of the parameters with the initial positions and velocities of the two first-order poles. The trajectories $z_i(t)$ can be computed from (2.3) and turn out to be

$$z_{1,2}(t) = a + bt \pm (At^2 + Bt + C)^{1/2}, \quad A = (1-b)^2, \tag{2.6}$$

where, a, b, B, and C are constants which are fixed by the initial conditions.

We turn now to the dynamics of the pole motions. If we consider the poles as representing classical particles of unit mass, then we have a two-body problem with a configuration space $\mathbb{C}^2$ and a phase space $\mathbb{C}^4$, i.e., the state of the system is given at $t\in\mathbb{R}$ by the vector $v(t) = (v_1, v_2, v_3, v_4) = (z_1(t), z_2(t), \dot{z}_1(t), \dot{z}_2(t))\in\mathbb{C}^4$. The dynamical properties of the system can now be computed from (2.6) and are as follows: (1) The equations of motion take the form $(dv/dt) = X_p(v)$, where the vector field $X_p:\mathbb{C}^4\rightarrow\mathbb{C}^4$ is given by $X_p(v) = (v_3, v_4, w, -w)$, $w = 2(1-v_3)(1-v_4)/(v_1-v_2)$. Note that X_p is autonomous and globally well defined except on the collision set $z_1 = z_2$. (2) X_p defines a smooth flow on the manifold $\mathbb{C}^{4-} = \mathbb{C}^4\setminus$collision set, i.e., $v(t) = \exp(tX_p)\,[v(0)]$, $v\in\mathbb{C}^{4-}$, $t\in\mathbb{R}$. (3) The system has two independent constants of motion, $v_3 + v_4 = \alpha_1\in\mathbb{C}$ and $(1-v_3)(1-v_4)(v_1-v_2)^2 = \alpha_2\in\mathbb{C}$, and thus the flow takes place on the submanifold, $N\subset\mathbb{C}^{4-}$, defined by the values of these constants.

The situation vis-a-vis the evolution equation (2.1) is now as follows. Let X_F be the vector field of (2.1), i.e., $u_t = X_F(u)$, and M the manifold of its solutions with initial profiles given by (2.4). It is straightforward to check that M is invariant under the flow generated by X_F and, hence, that every solution in M has the form (2.5). We now consider a map ϕ: $M\rightarrow\mathbb{C}^4$, which relates the orbits of $f(x)$ in M to those of its poles in $\mathbb{C}^4$, i.e., $\exp(tX_F)f(x)\subset M\xrightarrow{\phi}\exp(tX_p)v(0)\subset\mathbb{C}^4$. This induces a map between X_F and X_p, and thus we have at least two representations of the dynamics of (2.1): (i) as a flow on an infinite-dimensional function space M (the natural flow) and (ii) as a flow on a finite-dimensional complex space $N\subset\mathbb{C}^4$ (the induced flow). The next step is to identify X_p (via transformations if necessary) with the vector field of a two-body Hamiltonian system.

Thus, consider the Hamiltonian, H:$v\in\mathbb{C}^4\mapsto H(v)\in\mathbb{C}$, defined by

$$H = (v_3^2 + v_4^2)/2 + \beta/(v_1-v_2)^2, \quad \beta\in\mathbb{C}, \tag{2.7}$$

which is a straightforward extension, from $\mathbb{R}^4$ to $\mathbb{C}^4$, of the real Hamiltonian representing two particles moving in $\mathbb{R}$ under the influence of an inverse-cube force law. The vector field X_H of H is defined on $\mathbb{C}^{4-}$, and if we choose the constant β equal to α_2, then it is easy to see that $X_{H|N} = X_p$, i.e., the flows on N generated by X_H and X_p are identical. Thus the evolution equation (2.1) can be used to study the phase portrait of (2.7) on $N\subset\mathbb{C}^{4-}$ and possibly in its immediate neighborhood.

Remarks: (1) The analysis presented above can be generalized to solutions of (2.1) which contain n simple poles, i.e., those of the form $u(z,t) = \Sigma_{k=1}^n[1-\dot{z}_k(t)]/[1-z_k(t)]$, for *any* n. In this case we make a connection with the n-body version of (2.7):

$$H(\mathbb{C}^{2n}) = \sum_{j=1}^{n}\dot{z}_j^2/2 + \beta\sum_{j<k}(z_j - z_k)^{-2}. \tag{2.8}$$

However, such an unrestricted extension is not always possible. For example, if we consider the pole flows of the rational solutions of the KdV [which are also related to those of the system (2.8)], then nontrivial connections are limited to values of n for which $(1+8n)$ is a perfect square.[3] Thus, $n=2$ is trivial and we have a peculiar (from a physical point of veiw) situation in which there are two-body forces, but no two-body problem.

(2) In our particular example, the pole flow is embedded directly into the flow of a physically meaningful Hamiltonian, that is, one which is identified with the total energy function of the system and hence is the bonafide generator of translations in time. When this is not possible, a rather less direct connection may be made as follows. Consider an n-dimensional Hamiltonian system which is known to be completely integrable. Then the system has n independent constants of motion $C_j(q,p)$, including the Hamiltonian, which are in involution, i.e., the Poisson brackets $\{C_j, C_k\} = 0$ for all j, k. Thus, there are n basic vector fields, $X_j = (\nabla_p, \nabla_q)C_j$, which are linearly independent and commute with each other. It may now be possible to identify the vector field of the pole flow with some linear combination of the X_j. For example, the real version of (2.8) is known to be completely integrable[11] and, in Ref. 3, the poleflows of the rational solutions of the KdV are related to (2.8) in this indirect manner. Of course, the physical meaning of this relationship now becomes obscure, since we no longer have an embedding into *the* Hamiltonian flow. Nevertheless, from a mathematical point of view, the connection is interesting and may prove useful.

We now turn our attention to the more physical approaches of Refs. 6 and 7, which concentrate on soliton solutions. The example discussed above shows, quite clearly, that the main objectives of the mathematical approach to field-particle duality are (i) to establish formal connections between the pole flows of evolution equations and the motions of particle systems with two-body interactions and (ii) to use these connections to transfer mathematical information from one system to the other. For example, solvable evolution equations can be employed as a tool for analyzing and classifying integrable many-body problems.[12] However, we can also look at related field and particle systems from a physical point of view and ask whether the mathematical connections between them have more than just a formal significance. In other words, is it possible to give these connections a physical interpretation? For the most part the answer is no; there is, in general, no *natural* map (at the classical level) between the field solution and its associated particle systems. However, for evolution equations with soliton solutions the situation is different. In this case, when the field is excited into its pure soliton modes, it is clearly trying to mimic the behavior of a particle system, and thus it seems physically reasonably to identify an associated pole flow with the center-of-mass motions of the individual solitons in an n-soliton solution. Such a "particle" interpretation would given quantitative insight into the nature of soliton interactions as well as have other advantages. For example, it can be used as a computational tool in the perturbation theory of the translational modes of free and interacting solitons, since this analysis is sometimes technically easier to carry out in a particle picture.[13,14] Thus, in the physical approach, we concentrate on the solitons and look for a *unique* particle system which has the same *dimensionality* as the evolution equation and which is also a *faithful* representation of the particular solution it comes from.

Now it is somewhat surprising that such a system exists, but it does and to see briefly how it can be obtained let us consider the soliton–soliton scattering state solutions of the SGE [Eq. 1.3)]. In the center-of-velocity frame extended to a map over a complex-space domain, $z = x + iy \in \mathbb{C}$, these are given by

$$\phi_{\rm ss}(z,t) = 4\tan^{-1}(u \sinh \gamma z / \cosh \gamma u t). \quad t \in \mathbb{R}, \tag{2.9}$$

where $u \in (0,1)$ is the common speed of the solitons and $\gamma = (1-u^2)^{-1/2}$ is the Lorentz factor. We now identify the poles of the corresponding (complex) Hamiltonian density [Eq. (1.4)] with the center of mass of the two "interacting" solitons on the real line, i.e., (2.9) with $y = 0$. These poles occur at the points $z_p(t) = x_p + iy_p(t)$, where $y_p(t) = (2n+1)\pi/2\gamma$, $n \in \mathbb{Z}$, and $\exp[\pm \gamma x_p(t)] = [\cosh \gamma u t + (\cosh^2 \gamma u t - u^2)^{1/2}]/u$, and have the following properties. First, the periodicity with respect to y can be factored out via an equivalence relation and thus we only need to consider the strip $y \in (0, \pi/\gamma)$. This gives us uniqueness. Secondly, within this strip there are only two poles, for all allowed values of the parameter u, so that the representation is faithful. Finally, $\dot{y}_p = 0$ for all t and hence the pole motions are parallel to the motions of the solitons on the real axis, i.e., dimensionality is preserved. The projection of these poles onto the real axis then gives us a unique, two-particle system in one real space dimension and is the representation we are looking for. The vector field of the system can be computed from $x_p(t)$, and the masses of the particles obtained from (1.4) and (2.9) in the limits $y \to 0$, $t \to \pm\infty$. It turns out, as we mentioned in the Introduction, that the vector field is Hamiltonian with the pair potential (1.5a). The details of these computations are given in Ref. 6.

The "philosophy" of the physical approach may be summarized in terms of the following proposition:

Proposition: A physically meaningful particle interpretation of soliton dynamics can be based on the pole flows of a suitable mass or energy density of the field provided that: (i) There exists a set of poles of the density which is a faithful representation of the soliton solution for all allowed values of the essential parameters of the latter [e.g., u in (2.9)]; (ii) the pole motions of the faithful set are parallel to the real axis for all t, i.e., the induced vector field is one-dimensional and thus preserves the dimensionality of the evolution equation; (iii) the chosen density leads to a well-defined mass spectrum for the particles (this will be explained in the next section). □

In what follows we shall establish this proposition for the solitons of the KdV equation. We start with (iii), i.e., the choice of a suitable mass density for the KdV field.

III. THE PARTICLE INTERPRETATION

Since the KdV is a completely integrable Hamiltonian system[15] and hence conservative, one approach to its derivation is via a Lagrangian. The advantage of this method of obtaining the equation is that it also leads to a well-defined Hamiltonian densiy. For example, the Lagrangian density

$$\mathscr{L} = \frac{1}{2}\phi_x\phi_t + \frac{1}{6}\phi_x^3 - \frac{1}{2}\phi_{xx}^2, \tag{3.1}$$

where ϕ is a potential for the KdV field u, i.e., $u = \phi_x$, leads to the equation

$$u_t + uu_x + u_{xxx} = 0 \tag{3.2}$$

and to the Hamiltonian density

$$\mathscr{H} = -\frac{1}{6}u^3 + \frac{1}{2}u_x^2. \tag{3.3}$$

Now in any physical application of (3.2) it is usual to identify $\mathscr{H}$ with the energy density of the system being described. Such an identification also leads to a natural association between particles and the poles of $\mathscr{H}$, since they both represent points at which the mass/energy density is infinite. The assumption that the poles of $\mathscr{H}$ move in accordance with the center-of-mass dynamics of the corresponding soliton solution then allows us to compute the mass spectrum of the pole particles from the energy spectrum of the solitons.

In order to implement such a procedure, it is first of all necessary to check whether it can be consistently applied to the KdV in general and to its soliton solutions in particular. There are two points to consider. First, since we are dealing with a nonrelativistic system, we would like both the field and particle equations to be invariant under Galilean transformations. Secondly, since the total energy of the system is given by the integral

$$E(u) = \int_{-\infty}^{\infty} \mathscr{H} dx, \tag{3.4}$$

where u is the soliton solution under consideration, this integral must converge in all Galilean frames. Now both these conditions would be satisfied if the KdV field transformed like a Galilean scalar. However, it is easy to check that (3.2) remains invariant under the transformation $x \to x - ct$, $t \to t$, and $u \to u - c$, showing that u transforms like a time-reversed velocity rather than a scalar. This implies that if $E(u)$ is finite in one frame of reference it will certainly diverge in any other. Consequently, we either have a finite energy integral together with a noncovariant equation of motion or else a covariant equation with a divergent energy functional. It is interesting to note that this "ambiguous" behavior of the Kdv is in direct contrast to that of the SGE, where a scalar field interpretation is consistent both with Lorentz covariance and an integrable Hamiltonian density. The Hamiltonian could thus be identified with the energy of the sine–Gordon solitons and hence used to obtain the mass and energy of the corresponding pole particles. Unfortunately, this straightforward method cannot be applied to the KdV.

We see two ways out of the difficulty. The first is to develop a "renormalization" technique, which enables us to cancel out the energy infinities introduced by frame changes, while the second is to abandon the Lagrangian approach altogether and adopt a more phenomenological and physically intuitive way of deriving the KdV. The second option is easier to implement and is the method generally employed by hydrodynamicists.[16] In their approach the basic field is assumed to be scalar and the field equations, such as , for example, the KdV, are derived as structural perturbations of the linear, unidirectional equation

$$u_t + cu_x = 0, \quad c \in \mathbb{R} > 0. \tag{3.5}$$

If we adopt this derivation, then the following argument leads to a consistent particle interpretation.

Consider the solitary wave solutions of (3.5) and, in particular, those of the form

$$u_{ns} = \sum_{i=1}^{n} a_i \operatorname{sech}^2 b_i(x - ct + \delta_i), \tag{3.6}$$

where the subscript on u refers to the fact that there are n individual waves in the profile and the a_i, b_i, $\delta_i \in \mathbb{R}$ with $a_i > 0$. These may be thought of as free-field solitons since (i) they do not interact with each other, i.e., the superposition principle applies, and (ii) there is no self-interaction, i.e., the speed is fixed and the a_i and b_i are free parameters. Now the form of (3.5) tells us that u is a locally conserved density. It follows, therefore, that, since both u_{ns} and its x derivative vanish as $|x| \to \infty$, the functional

$$m(u_{ns}) = \int_{-\infty}^{\infty} u_{ns} dx < \infty \tag{3.7}$$

is a constant of motion for the system of noninteracting solitons represented by the solution (3.6). This establishes the existence of an integrable conserved density for the field equation (3.5).

We now turn to the question of Galilean invariance. Since we have defined u to be a scalar, (3.5) will only be covariant under the Galilean transformations if the coefficient of u_x transforms like a velocity. In order for this to be physically meaningful, we have to assume that the system has a characteristic velocity, i.e., the solitons are local excitations of the field which move with a definite speed relative to a quiescent background. Thus, we have no control over the speed of the solitons other than by changing inertial frames. On the other hand, we do have a Galilean invariant interpretation in which u is a conserved scalar density.

If we adopt this interpretation, then we can identify u with the Newtonian mass density of the field. This has two immediate consequences. First, it gives us the required identification of the poles of u_{ns} with Newtonian particles and, second, it allows us to attribute definite masses to these particles via the conserved functional (3.7). This latter property can easily be checked by evaluating (3.7) for the n-soliton solution (3.6). We have, in this case,

$$m(u_{ns}) = \sum_{i=1}^{n} 2a_i |b_i|, \quad a_i \in \mathbb{R} > 0, \tag{3.8}$$

which shows that $m(u)$ has all the properties required of Newtonian inertia, i.e., it is finite, constant, Galilean-invariant, and additive.

We thus have a consistent association of classical particles with the free-field solitions (3.6). Our next step is to verify that the desirable features of this association are preserved in the transition to the KdV equation

$$u_t + cu_x + uu_x + u_{xxx} = 0 \tag{3.9}$$

considered as a perturbation of (3.5). [Note that (3.9) reduces to (3.2) under the transformation $x \to x - ct$, $t \to t$, $u \to u$]. Now it is obvious from the form of (3.9) that u retains the property of being a locally conserved density—this would not be true, for example, if the nonlinear term was u^2 rather than $(u^2)_x$. On the other hand, since the underlying structure of the field has been changed, the role played by c, the coefficient of u_x has to be redefined if we are to maintain Galilean invariance. Thus, consider the single-soliton solutions of (3.9), i.e.,

$$u_s = 3\alpha^2 \operatorname{sech}^2\{(\alpha/2)[x - (c + \alpha^2)t\,] + \delta\}, \quad \alpha, \delta \in \mathbb{R}, \tag{3.10}$$

Comparing this with the corresponding single-soliton solution in (3.6), we note that the amplitude factor a_i, the shape factor b_i, and the speed of the soliton are no larger independent quantities. As a consequence, c no longer represents the characteristic speed with which disturbances travel through the field. Rather, it acts as a lower bound for the speed of propagation of solitons. In other words, in any given frame the coefficient of u_x represents the minimum speed with which KdV solitons can be generated. However, this is also a Galilean invariant interpretation, with u transforming as a scalar, and by adopting it we preserve the identification of u with the Newtonian mass density of the KdV field. The functional (3.7) thus gives us the total mass of the solitons, and, since u is a locally conserved density of the KdV which is integrable for its soliton solutions, this total mass is a constant of the motion. Furthermore, we know that the asymtotic ($t \to \pm\infty$) form of the KdV n-soliton solution is given by (3.6), with $12b_i^2 = a_i$ and c replaced by $c + a_i/3$, and thus the mass functional is additive over the soliton masses as required. In fact from (3.8) we see that this total mass is given by

$$m(u_{ns}) = \sum_{i=1}^{n} m_i(u_s) = \sum_{i=1}^{n} 12|\alpha_i|, \tag{3.11}$$

where $3\alpha_i^2$ is the maximum height of the ith soliton. We have thus recovered the consistent pole-particle association that we had for the linear solitons, and we can proceed with the exercise of locating the poles and computing their orbits.

IV. THE POLES AND THEIR MOTIONS

We shall work in the fram in which the lower bound of the soliton speeds is zero, i.e., $c = 0$ in (3.9). Then, using the results of Hirota,[17] the n-soliton solution can be written in the form

$$u_{ns} = 12(\ln f)_{xx} = 12(ff_{xx} - f_x^2)/f^2, \tag{4.1}$$

where f is the determinant of the symmetric $n \times n$ matrix, $A = (a_{ij})$, whose elements are given by

$$a_{ij} = \delta_{ij} + \frac{2\sqrt{\alpha_i + \alpha_j}}{\alpha_i + \alpha_j} \exp\left(\frac{\xi_i + \xi_j}{2}\right),$$

$$\alpha_i \neq \alpha_j \quad \text{for } i \neq j, \tag{4.2}$$

with $\xi_i = \alpha_i x - \alpha_i^3 t + \delta_i$. The α_i and δ_i are the amplitude and phase factors, respectively, of the ith soliton. Note that this form of the solution shows, quite clearly, that the interactions between n solitons, $n \geqslant 2$, can be analyzed in terms of two-soliton forces determined by the invariants $2(\alpha_i\alpha_j)^{1/2}/(\alpha_i + \alpha_j)$. Thus, as in the case of the sine–Gordon equation, the essential features of the translational mode dynamics of KdV solitons are already contained in the two-soliton solutions. Hence, in this paper, we restrict ourselves to the cases $n = 1$ and $n = 2$.

From (4.1) we see that the poles of u_{ns} are given by the zeros of f provided that these zeros do not coincide with those of f_x. However, one can easily check, from the expressions for the matrix elements of A, that this coincidence of zeros does not occur and thus the problem of finding the poles of u_{ns} is reduced to the problem of finding the zeros of f.

The question now arises as to whether f has any isolated zeros. It is evident that, as long as u_{ns} is considered to be a map from $\mathbb{R}^2 \to \mathbb{R}$, then it has no poles and therefore f has no zeros. However, if we allow the space variable x to be complex, while still keeping the time t real, then u_{ns} becomes a map from $\mathbb{C}\times\mathbb{R}\to\mathbb{C}$ and the equation $f = 0$ has solutions. As we shall see, these solutions, parametrized in terms of t, lie on well-defined orbits in the complex plane and are isolated for almost all t. Furthermore, the faithful solutions have constant imaginary parts, and thus their trajectories are parallel to the real axis. This means, of course, that the dynamics of the faithful pole motions is essentially one-dimensional and identical to that of their projections on the real axis.

Thus, consider the single-soliton solutions u_s. In this case we have $f_s = 1 + \exp(\alpha x - \alpha^3 + \delta)$, $\alpha, \delta\in\mathbb{R}$, and, hence, in terms of $z = x + iy$, the poles of u_s occur at the points

$$z_n = x_n + iy_n = (\alpha^2 t - \delta/\alpha) + i(2n + 1)\pi/\alpha, \quad n\in\mathbb{Z}, \quad t\in\mathbb{R}. \tag{4.3}$$

We see that there is an infinite set of isolated poles whose orbits are parallel to the x axis for all time, i.e., $\dot{y}_n = 0 \forall n$ and t. The pole motions are therefore one-dimensional, and, since the x_n are independent of n, the projection onto the real axis, $\pi_t : z_n(t) \mapsto x_n(t)$, of the set $\{z_n\}$ is a single point $\{x_s\}(t)$ which moves with a constant speed α^2 in the direction of increasing x. Thus we have a faithful representation of the center-of-mass motion of the single soliton. Placing a particle of mass $12|\alpha|$ at the point $x_s(t)$ and allowing it to move with the point then completes the particle picture.

Remark: In the above analysis we have kept (δ/α) real. However, since we are working in a complex space domain, we are free to choose the initial position of the soliton anywhere in the z plane, i.e., to put (δ/α) equal to a complex constant. For the single soliton this only has a trivial influence on the results, as it merely produces a shift in the y_n values by the constant amount $\mathrm{Im}(\delta/\alpha)$. However, in the case of the two-soliton solutions the choice of initial positions is important, for, as we shall see below, it is only when the *imaginary* part of the *relative* phase satisfies a certain necessary condition that we obtain the faithful representation.

We turn now to the two-soliton solutions. Here we have

$$f_{2s} = 1 + \exp\xi_1 + \exp\xi_2 + a^2 \exp(\xi_1 + \xi_2), \tag{4.4}$$

where $a^2 = (\alpha_1 - \alpha_2)^2/(\alpha_1 + \alpha_2)^2 > 0$ and $\xi_{1,2}$ are as defined in (4.2). There is no loss of generality in treating a as a positive constant, so that if we redefine our phases to have $\xi_i = \alpha_i(x - \alpha_i^2 t + \delta_i) - \ln a$, then (4.4) can be written as

$$af_{2s} = a + \exp\xi_1 + \exp\xi_2 + a\exp(\xi_1 + \xi_2). \tag{4.5}$$

Before we solve for the zeroes of f_{2s} it is useful to reduce (4.5) to its minimal form, i.e., to eliminate all degrees of feedom which only have a trivial influence on the numbers and positions of the zeroes. First, we note that the speeds of the solitons in the solutions given by (4.4) are α_1^2 and α_2^2, respectively. Since the KdV is unidirectional and, in the form (3.2), has solitons moving from left to right, α_1 and α_2 are real. Thus, without loss of generality, we can take $\alpha_1 > \alpha_2 > 0$ and define $\rho = (\alpha_1/\alpha_2) > 1$, so that ρ^2 is the speed ratio. We now scale x and t via the substitutions $\alpha_2 x \to x$ and $\alpha_2^3 t \to t$, which, together with the redefinitions of the phases mentioned above, reduce $\xi_{1,2}$ to $\xi_1 = \rho(x - \rho^2 t + \delta_1)$ and $\xi_2 = (x - t + \delta_2)$. Furthermore, as only the relative phase is important, we can set $\delta_1 = 0$ corresponding to the faster soliton reaching the origin at $t = 0$. Then, in terms of z and $\delta_2\in\mathbb{C}$, the minimal form of $f_{2s} = 0$ is

$$a + e^{\rho(z - \rho^2 t)} + e^{(z - t + \delta_2)} + ae^{(\rho + 1)z - (\rho^3 + 1)t + \delta_2} = 0, \tag{4.6}$$

where $a = (\rho - 1)/(\rho + 1)$. The problem of finding the faithful representation can now be posed as follows: Identify those values of $\delta_2\in\mathbb{C}$ for which (i) (4.6) has two and only two distinct solution functions $z_1(\rho,t)$ and $z_2(\rho,t)$ for *all* $\rho\in(1,\infty)$ and almost all $t\in(-\infty,\infty)$ and (ii) the imaginary parts of z_1 and z_2 are independent of time.

Remarks: (i) Note that for $\rho\in\mathbb{Q}^* = \{p/q | p,q\in\mathbb{N}, p > q, p$ and q relatively prime$\}$, (4.6) is invariant under the translations $z \to z + i2\pi qn$, $n\in\mathbb{Z}$, i.e., z is periodic in the imaginary direction. Thus, when we talk of distinct solutions we mean modulo this periodicity. (ii) Almost all t is put in to take care of possible collisions. (iii) The second condition on $z_{1,2}$ means that the dynamics of the pole motions are identical to that of thier projections on the real axis as in the single soliton case. (iv) By identifying $\delta_2\in\mathbb{C}$ we mean only the imaginary part, since the real part corresponds to the phase of the slower soliton and is always a free choice.

Before we present our solution to this problem, let us briefly review the method used by Thickstun.[7] Let $\rho = p/q\in\mathbb{Q}^*$, $\delta_2\in\mathbb{R}^+$, $qw = (z - t + \delta_2)$, and $\tau = \rho[(\rho^2 - 1)t + \delta_2]\in\mathbb{R}$. The periodicity of z can be factored out by setting $\phi = \exp w$, and, for convenience, let $T = \exp\tau \Rightarrow T\in(0,\infty)$. Then (4.6) reduces to

$$(p - q)\phi^{p+q} + (p + q)(\phi^p + T\phi^q) + (p - q)T = 0, \tag{4.7}$$

which is a polynomial in ϕ, of degree $p + q$, with real coefficients. Thus, in any period strip $\Delta w = 2\pi i$, there are $p + q$ distinct poles for almost all $T\in(0,\infty)$. A qualitative description of the motions of these poles (in the w plane rather than the z plane) is given in Ref. 7, and we shall not dwell on them here. However, we wish to make three points about some of the general features of these solutions. First, since $p + q \geqslant 3$, it is obvious that the representation is not faithful. What is not so apparent is that it is also unstable; that is, in a neighborhood of any $\rho_L\in\mathbb{Q}^*$, $p + q$ has infinitely many values and is unbounded, e.g., consider the sequences defined by $\rho_n = (2n \pm 1)/n$ with $\rho_L = 2$. The same phenomenon occurs for rational sequences with irrational limits, e.g., $\rho_n = [(n + 1))/n]^n$, and thus it seems virtually impossible to analyze the pole motions for irrational ρ. Secondly, the $p + q$ solutions of (4.7) do contain the faithful poles, but, as we shall see below, only when $p + q$ is an even integer. However, since $\mathbb{Q}^*$ as a subset of $\mathbb{R}$ has measure zero, these faithful solutions are sparse in the set of all such solutions, i.e., for $\rho\in\mathbb{R}^+ > 1$. Thirdly, in general, the motions of the poles are two-dimensional, i.e., both real and imaginary parts are time-dependent, except in the asymptotic ($|t| \to \infty$) regions. Thus, for finite t, projection onto the real axis does not preserve the "poledynamics."

We now turn to our method of solving (4.6). Let $\delta_2 = i\delta$, $\delta \in \mathbb{R}$. Combined with our earlier choice of $\delta_1 = 0$, this corresponds to choosing $t = 0$ when both the solitons are at the origin of the x axis. Set $w = u + iv$, $\tau = r + is$, $u, v, r, s \in \mathbb{R}$, and consider the affine map $(z, t, \delta) \to (w, \tau)$ defined by

$$\begin{pmatrix} w \\ \tau \end{pmatrix} = \frac{1}{2}\begin{pmatrix} (\rho + 1) & -(\rho^3 + 1) \\ (\rho - 1) & -(\rho^3 - 1) \end{pmatrix}\begin{pmatrix} z \\ t \end{pmatrix} + \frac{1}{2}\begin{pmatrix} i\delta \\ -i\delta \end{pmatrix}. \tag{4.8}$$

This is essentially a map form $\mathbb{C} \times \mathbb{R} \times \mathbb{R} \to \mathbb{C} \times \mathbb{C}$, and it is easy to check that the matrix in (4.8) is nonsingular for all $\rho > 1$. Thus, if we recall that $z = x + iy$ and that t is *real*, then the inverse of (4.8), expressed as a map from $\mathbb{R}^4 \to \mathbb{R}^4$, takes the form

$$X = \begin{bmatrix} 2A_{2\times 2} & 0\ 0 \\ & 0\ 0 \\ 0\ 0 & 2A_{2\times 2} \\ 0\ 0 & \end{bmatrix} Y + 2|A|B, \quad ||\equiv \text{determinant}, \tag{4.9}$$

where A is the inverse of the matrix in (4.8), $X = (x,t,y,0)$, $Y = (u,r,v,s)$, and $B = (0,0,\rho^3\delta,\rho\delta)$. We now have the following lemma about the real quantities $y(t)$, $v(t)$, and $s(t)$.

Lemma 1: Fix $\rho > 1$ and $\delta \in \mathbb{R}$. Then (i) $(v - s)$ is constant if and only if $(v + s)$ is constant, (ii) v is constant if and only if s is constant, (iii) y is constant if and only if v is constant. □

Proof: The last equation of (4.9) can be written in the form $\rho\delta = \rho(v - s) - (v + s) \forall t$, from which (i) follows and (ii) follows from (i). From the last two equations of (4.9) we have $y = (v - s) - \delta$ and (iii) follows from (ii). ■

Having established necessary and sufficient conditions for the constancy of y, we now go back to (4.6). In terms of w and τ, (4.6) can be written in the form

$$e^{w}(\alpha \cosh w + \cosh \tau) = 0. \tag{4.10}$$

For finite x and t $e^{w} \neq 0$ and so the pole positions, in (w,τ) space, are given by the solutions of the two real equations

$$\alpha \cosh u \cos v + \cosh r \cos s = 0, \tag{4.11a}$$

$$\alpha \sinh u \sin v + \sinh r \sin s = 0, \tag{4.11b}$$

subject to the constraint

$$s = \alpha v - (\rho\delta/\rho + 1) \tag{4.11c}$$

which, as mentioned in Lemma 1 above, comes from (4.9).

To analyze the solutions of (4.11), let us think of these equations as defining a set of maps, $\{\phi_{\rho\sigma}\}$, each of which is a many-valued function from $v \in V_{\rho\delta} \subset \mathbb{R}$ to $(u, r, s) \in U_{\rho\delta} \subset \mathbb{R}^3$ for some fixed values of ρ and δ. The problem then is to specify the branches, domains and ranges of this set for all $\rho \in \mathbb{R}^+ > 1$ and $\delta \in \mathbb{R}$. Since the main purpose of this paper is to present and analyze the faithful maps, i.e., those $\phi_{\rho\delta}$ which lead to only two poles in the z plane for all $\rho > 1$, we shall not enter into a detailed discussion of the other solutions, i.e., those poles which form what we have previously referred to as the random or unstable component. Instead we list the salient features of the solutions of (4.11) in a sequence of propositions and supplement this information with some numerical examples.

Proposition 1: For $\rho = p/q \in \mathbb{Q}^*$ and any $\delta \in \mathbb{R}$, the map $\phi_{\rho\delta}$ is (i) periodic in v with period $\Delta v = \pi(p + q)$ and (ii) has $p + q$ distinct branches in a period strip Δv for almost all t. □

Proof: (i) $\Delta v = \pi(p + q) \Rightarrow \Delta s = \pi(p + q)$. If $p + q$ is even (odd), then $p - q$ is even (odd). Therefore, $\cos(v + \Delta v)/\cos(s + \Delta s) = \cos v/\cos s$ and similarly for the sines. (ii) Within a period strip, Eq. (4.10) may be rewritten as a polynomial of degree $p + q$, similar to (4.7), but with complex coefficients for $\delta \neq n\pi$. ■

Proposition 2: The relationship between the branches of $\phi_{\rho\delta}$ and the pole orbits in the z plane is one-to-one. □

Proof: This follows immediately from the linearity of (4.9). ■

Proposition 3: In a period strip of the z plane, $\Delta z = i2\pi q$, the $p + q$ pole orbits split into two distinct classes: (i) those which are parallel to the real axis, i.e., $\dot{y}(t) = 0\ \forall\ t$ and (ii) those which are not, i.e., $\ddot{y}(t) \neq 0\ \forall\ t$. □

Proof: There are three types of poles, doublets and singlets whose orbits are in class (i) and random poles with orbits in class (ii). We shall prove the existence of each of these types and give a brief description of their motions.

(*ia*) *Doublets or faithful poles*: For fixed $\rho = p/q \in \mathbb{Q}^*$, choose a value of δ from the sequence $\delta_m = [p - (2m + 1)q]\pi/p)$ defined over $m \in \mathbb{Z}$. Now, for fixed m, consider the sequence of points $\{v_{nm}\}$, defined by $v_{nm} = [(p + q)n + m + 1]\pi$, $n \in \mathbb{Z}$, and the function $v_{nm} \mapsto s_{nm} = [(p - q)n + m]\pi$. The graph of this function satisfies (4.11c) for all $n, m \in \mathbb{Z}$. Furthermore, $\cos v_{nm} = -\cos s_{nm} = \pm 1$ and $\sin v_{nm} = \sin s_{nm} = 0$, $m, n \in \mathbb{Z}$. Hence (4.11b) is identically satisfied and (4.11a) reduces to

$$\alpha \cosh u = \cosh r \quad \forall\ m, n \in \mathbb{Z}. \tag{4.12}$$

The solutions of (4.12) lie on two distinct, smooth curves in the (u, r) plane defined for $r \in (-\infty, \infty)$ and $u = \pm(a, \infty)$, $a = |\cosh^{-1}(1/\alpha)|$, the picture being similar to that of the hyperbola $u = \pm a(1 + r^2)^{1/2}$. Thus, $\phi_{\rho\delta}$, $\delta \in \{\delta_m\}$, maps the set $\{v_{nm}\}$ onto the set $\{s_{nm}\}$ and the two solution curves of (4.12). The net result, for fixed m, is a stratification of the (w,τ) space by planes parallel to its real subspace, each plane being located at definite values of v_{nm} and s_{nm} and containing an identical copy of the solution curves of (4.12). The corresponding picture in the z plane is a stratification by lines parallel to the x axis and located at the points $y_{nm} = (2np + 2m + 1)(\pi/\rho)$. Each line contains two nonintersecting pole orbits $\forall\ \rho \in \mathbb{Q}^*$, the pole motions along these orbits being independent of n and m, and, for fixed m, there is only one such line in each period strip $\Delta z = i2\pi q$. This then is the faithful representation for rational ρ.

Remark: $\{\delta_m\} = \{\delta_0\} \cap \{\delta_e,\}$ where δ_0 and δ_e are phases which lead to faithful poles when p and $q \neq 1$ are both odd and when one of them is even, respectively. $\{\delta_0\}$ contains zero, but $\{\delta_e\}$ does not. However, for the special cases in which $q = 1$, $0 \in \{\delta_m\}$. This explains why some of the solutions obtained in Ref. 7 contain faithful poles while the others do not.

(*ib*) *Singlets*: These come in two varieties, poles with regular motion and poles with singular motion. As in the case of the faithful poles, the maps $\phi_{\rho\delta}$ define a stratification of the (w, τ) space by planes parallel to the real subspace (u, r) and this translates into a stratification of the z plane by lines parallel to the x axis. Thus, for a fixed value of the phase, each period strip in the z plane contains a single pole whose motion is along some line $y(m, n, p, q) = \text{const}$, where $m, n \in \mathbb{Z}$ label the phase and the period strip, respectively. The general features of these singlets are as follows. Choose v and s so that $\cos v = \cos s = 0$. Then $\sin v = \pm \sin s$ and Eqs. (4.11a) and (4.11b) reduce to

$$\alpha \sinh u \pm \sinh r = 0, \tag{4.13}$$

defining two smooth curves $u_+(r)$ and $u_-(r)$ in the (u, r) plane. From (4.9), the transformations to the corresponding (x, t) planes are given by

$$x = u - r + t, \tag{4.14a}$$

$$\rho(\rho^2 - 1)t = (\rho - 1)u - (\rho + 1)r, \tag{4.14b}$$

and, from (4.13), it is easy to see that $u_+(r) + u_-(r) = 0$. Thus, if (x_+, t_+) and (x_-, t_-) label the (x,t) planes of $u_+(r)$ and $u_-(r)$, respectively, straightforward substitution into (4.14) leads to the relation

$$\rho\begin{pmatrix} x_+ \\ t_+ \end{pmatrix} = \begin{pmatrix} -(\rho^2+1) & (\rho^4+\rho^2+1) \\ -1 & (\rho^2+1) \end{pmatrix}\begin{pmatrix} x_- \\ t_- \end{pmatrix}, \tag{4.15}$$

so that the (x, t) trajectories of the poles corresponding to u_+ and u_- are connected by a nonsingular linear transformation. It follows that the velocities of the poles satisfy the equation

$$\dot{x}_+ = [(\rho^4 + \rho^2 + 1) - (\rho^2 + 1)\dot{x}_-]/ [(\rho^2 + 1) - \dot{x}_-], \tag{4.16}$$

implying that $(d\dot{x}_+ | d\dot{x}_-) < 0$, i.e., the accelerations are always in opposite directions. The occurence and motions of these poles are as follows.

Irregular singlets (u_-, x_-, t_-): The relevant sequences are $\delta_m = [p - (2m + 2)q](\pi/p)$, $v_{nm} = [(p + q)n + m + 3/2]\pi$ and $s_{nm} = [(p - q)n + m + \frac{1}{2}]\pi$. In the z plane the poles move along the lines $y = (pn + m + 1)(2\pi/\rho)$, and their common trajectory in the (x, t) plane has the following properties. It passes through the origin and in a neighborhood of $t = 0$ has the form

$$x(t) = -at^{1/3} + bt + O(t^{5/3}),$$

$$a,b > 0, \quad (a/b) < 0.8. \tag{4.17}$$

Thus, $\dot{x} \approx 0$ when $t = \pm t_0 = \pm (a/3b)^{3/2}$, $\dot{x} < 0$ for $|t| < t_0$ and $\dot{x} \to -\infty$ as $t \to 0$. On the other hand, as $|t| \to \infty$, we have $x(t) = t + \text{const}$ and therefore $\dot{x} \to 1$ as $|t| \to \infty$. Thus, in terms of the coordinates $(x, \dot{x}, t)$, the motion of the pole is $(-\infty, 1, -\infty) \to (x_0, 0, -t_0) \to (0, -\infty, 0) \to (-x_0, 0, t_0) \to (\infty, 1, \infty)$, where $x_0 \approx 4(1 - \alpha)/3(3 + \alpha^2)^{1/2}$. This motion is rather curious. The (x,t) trajectory is "s"-shaped with the middle leg of the s inflecting through the x axis at (0,0). The pole therefore bounces twice about the origin, and in between bounces, i.e., when $\dot{x} < 0$, it is subjected to unbounded accelerations and decelerations so that is passes through the origin with infinite negative speed, i.e., $\ddot{x} \propto -x^{-5}$, $\dot{x} \propto -x^{-2}$ in a neighborhood of (0,0).

Regular singlets (u_+, x_+, t_+): The sequences are $\delta_m = (2m - 1)q\pi/p$, $v_{nm} = [(p + q)n - m + 1/2]\pi$ and $s_{nm} = [(p + q)n - m + 1/2]\pi$, and in the z plane these poles move along the lines $y = (2pn - 2m + 1)(\pi/\rho)$. In the (x,t) plane their common trajectory looks very much like the curve $x = \tanh[t/(\rho^4 + \rho^2 + 1)]$ rotated anticlockwise through $\theta = \tan^{-1}\rho^2$, this approximation being exact in a neighborhood of the origin and giving the correct speeds in the asymptotic regions, i.e., in terms of the coordinates $(x, \dot{x}, t)$ the motion of the pole is $(-\infty, \rho^2, -\infty) \to (0, \rho^2 + 1, 0) \to (\infty, \rho^2, \infty)$. The motion is smooth and unidirectional with no stationary points or singularities of the speed. Note that, in terms of the original coordinates [Eq. (4.5)], $\dot{x} = 1 \to \dot{x} = \alpha_2^2$ and $\dot{x} = \rho^2 \to \dot{x} = \alpha_1^2$, i.e., asymptotically the irregular pole moves with the speed of the slower soliton, while the regular pole has the speed of the faster soliton. This completes the description of the singlets.

Remark: Note that, for $\rho \in \mathbb{Q}^*$, a necessary condition for the occurence of faithful poles and singlets is that $(p\delta/\pi)$ must be an integer.

Now the faithful poles and the singlets are the only ones which move parallel to the x axis, and the results given above can be used to prove a useful classification lemma.

Definition: A representation or multiplet is the set of poles which occur in a period strip. Thus, each multiplet is labelled by its phase δ and, as shown earlier, contains $p + q \geqslant 3$ poles.

Lemma 2: Let $p + q = N \in \mathbb{N} \geqslant 3$ and $(p\delta/\pi) = M \in \mathbb{Z}$; then the following are true. (a) If both N and M are even, the multiplet may contain faithful poles, but will not include any singlets. (b) If N is even and M is odd, one or both of the singlets may be present, but the faithful poles will not occur. (c) If both N and M are odd, the multiplet may contain faithful poles and/or one, but both, of the singlets. (d) If N is odd and M is even, then one of the singlets may be present, but there will be no faithful poles. □

Remark: This lemma explains why the results of Ref. 7 contain irregular poles, regular poles, and faithful poles in different representations.

(*ii*) *Random poles*: These are solutions of (4.11) which differ from the ones described above in two essential ways. First, v and s are not constant which implies that the pole motions in the z plane are two-dimensional. Secondly, the number of these poles in a given multiplet is a function of the speed ratio and the $\rho \in \mathbb{Q}^*$ lies between $p + q - 3$ and $p + q$. Now these solutions are difficult to analyze in general, and so we shall work with a specific example. Thus, consider functions $v(t)$, $s(t)$, $\neq n\pi/2$ at the same time, such that $\cos v$ and $\cos s$ have opposite signs. This ensures that (4.11a) and (4.11b) have solutions, and we can now rewrite these equations in the equivalent form

$$\sinh r = -f(\alpha, v, s)\sin v, \tag{4.18a}$$

$$\alpha \sinh u = f(\alpha, v, s)\sin s, \tag{4.18b}$$

where $f^2 + 1 = (1 - \alpha^2)(1 + \cos 2v)/(\cos 2v - \cos 2s)$. This introduces additional constraints on v and s, namely $1 + \cos 2v > 1 + \cos 2s > \alpha^2(1 + \cos 2v)$, which means that the solutions to (4.18) go through definite "windows" in (v, s) space. Let us now choose $p = 2$, $q = 1$, and $\delta = 3\pi/8$. It is easy to check that this choice of phase excludes faithful poles and singlets, and, since the multiplet contains three poles, we expect (4.18) to have three distinct solutions in each period interval on the line $v = 3s + (3\pi/4)$, the periods of v and s being 3π and π, respectively. From the constraints on v and s it follows, by a straightforward though tedious computation, that there are *two* windows per period interval and, for the interval $s\in[0,\pi]$, these are (i) $\pi/16 < s < \pi/8$ and (ii) $0.5411\pi \leqslant s < 0.5625\pi$. The first window contains two solutions, which in a neighborhood of $r = 0$ are $\sinh^2 u = 7.4 - (\mathrm{sgn}u)0.44r + 9.5r^2$, while the second window contains only one solution $u(r) + u(-r) = 0$ with $u'(0) \approx -3.237$ and $u''(\pm\infty) = -1$. Thus, as expected, all three poles in the multiplet are of the random type.

Remarks: (i) As mentioned earlier, these random poles have two-dimensional dynamics, i.e., $\ddot{y} \neq 0 \ \forall\ t$. The easiest way to see this is to compare the values of $\dot{y} = (1 - \alpha)\dot{v}$ as $|r| \rightarrow \infty$ and in a neighborhood of $r = 0$. (ii) If we had chosen $\delta = \pi/2$ in the above example, then the multiplet would have consisted of two faithful poles and a regular singlet.

This completes our somewhat lengthly demonstration of Proposition 3. ■

The information contained in the three propositions given above can be summarized as follows. (i) If $\rho = p/q \in \mathbb{Q}^*$, the pole-positions $z_j = x_j + iy_j$ are periodic in y with period $\Delta y = 2\pi q$. Thus the z plane is divided into period strips with each strip containing a multiplet of $p + q$ distinct poles. (ii) If $p + q > 3$, then the multiplet contains at least $p + q - 3$ random poles. This number is a "chaotic" function of the speed ratio and is unbounded when this ratio is irrational. (iii) If δ is chosen so that $p\delta/\pi$ is an integer, then the multiplet contains stable components, i.e., the doublet and the two singlets. As their names imply, the numbers of these poles are fixed and their motions are always parallel to the real axis. (iv) The maximum number of stable poles that can appear in a multiplet is three; thus the doublet can never occur together with *both* the singlets.

Now it is fairly clear that the random poles do not provide a useful model for a particle interpretation of the two interacting solitons and, hence, can be eliminated from further consideration. This leaves the faithful poles, i.e., the doublet, and the two singlets. The next step is to isolate the faithful poles and to analyze the dynamics of their motions. This is done in the following section.

V. DYNAMICS OF THE FAITHFUL REPRESENTATION

So far, our classification of the pole motions has been restricted to rational values of ρ. This means, of course, that we only have a sparse set of solutions, because, although the rationals are dense in $\mathbb{R}$, they are countable. Furthermore, there is still the possibility that the singlets plays some role in the particle representation, since, for even $p + q$, the phase can be chosen so that both singlets occur in the same multiplet. However, it turns out that the extension to irrational values of ρ also eliminates the singlets. This follows from the following proposition.

Proposition: For any $\rho\in\mathbb{R} > 1$, the multiplet will contain (a) faithful poles for the sequences $s = (n - m)\pi$, $v = s + (2m + 1)\pi$, and $\delta = (2m + 1)\pi - (2n + 1)(\pi/\rho)$, (b) regular singlets for the sequences $s = (2n - 2m + 1)/(\pi/2)$, $v = s + 2m\pi$, and $\delta = 2m\pi - (2n + 1)(\pi/\rho)$, (c) irregular singlets for the sequences $s = (2n - 2m + 1)/(\pi/2)$, $v = s + (2m + 1)\pi$, and $\delta = (2m + 1)\pi - (2n + 2)(\pi/\rho)$, where n, $n\in\mathbb{Z}$.□

Proof: Straightforward substitution into Eqs. (4.11) leads to the desired results. ■

Remark: Note that the pole orbits are now continuous functions of ρ and, hence, of the speed ratio.

Corollaries: (i) For irrational values of ρ, the doublet, regular singlet, and irregular singlet appear in different multiplets. Thus, the singlets are isolated from each other as well as from the doublet and, as a consequence, the latter is the only representation which is faithful for *all* values of ρ. (ii) In the z plane the motions of the faithful poles are along the lines $y = (2n + 1)\pi/\rho$, i.e., the z plane is divided into period strips $\Delta y = 2\pi/\rho$.

Having eliminated the singlets, we can now concentrate on analyzing and interpreting the dynamics of the faithful poles. Recalling the computations of the last section, we see, from (4.12), that the orbits of these poles in the (u, r) plane are given by $u = \pm g(r)$, where

$$g(r) = \ln[\cosh r + (\cosh^2 r - \alpha^2)^{1/2}] - \ln\alpha, \quad r\in\mathbb{R}. \tag{5.1}$$

It is fairly clear that $g(r)$ is an even, differentiable function of r which, since $0 < \alpha < 1$, is positive $\forall\ r\in\mathbb{R}$. The other properties of g that we shall need in our analysis can be summarized in terms of the following lemma.

Lemma: Given $g(r)$ as defined by (5.1), then (i) $g(r) > |r|$ and (ii) $g'(r)$ is a diffeomorphism from $(-\infty, \infty)$ to $(-1, 1)$, $\forall\ r\in\mathbb{R}$. □

Proof: (i) Since $\alpha < 1$, $\cosh^2 r - \alpha^2 > \sinh^2 r$. Therefore, $g(r) > \ln e^{|r|} - \ln\alpha > \ln e^{|r|} = |r|$. (ii) $g'(r) = \sinh r/(\cosh^2 r - \alpha^2)^{1/2} \Rightarrow |g'(r)| < 1\ \forall\ r\in\mathbb{R}$. Furthermore, g' is differentiable and we have $g''(r) = (1 - \alpha^2)\cosh r/(\cosh^2 r - \alpha^2)^{3/2} \Rightarrow 0 < g''(r) < \infty\ \forall\ r\in\mathbb{R}$. Hence g' is a diffeomorphism. ■

Now, from (4.9), the orbits of the faithful poles in the (x, t) plane are given, in parametrized form, by the equations

$$x_{2,1} = \pm ag(r) + br, \tag{5.2a}$$

$$t_{2,1} = \pm cg(r) + dr, \tag{5.2b}$$

where $a = (\rho^2 + \rho + 1)c$, $b = (\rho^2 - \rho + 1)d$, $c = (\rho^2 + \rho)^{-1}$, and $d = -(\rho^2 - \rho)^{-1}$. Note that, since $\rho\in(1, \infty)$, these coefficients satisfy the inequalities $1 < a < 3/2$, $-\infty < b < -1$, $0 < c < 1/2$, and $-\infty < d < 0$. Equations (5.2) are in computable form in the sense that, given a numerical value for ρ, the individual trajectories $\Gamma_1 = \{(x_1, t_1)\}$ and $\Gamma_2 = \{(x_2, t_2)\}$ may be calculated and plotted in the (x,t) plane by letting the parameter r run through all its values. The resulting graphs are hyperbolic curves, typical of a repulsive interaction, and their essential features are listed in the following theorems.

Theorem 1: In the (u, r) plane there are two and only two solutions for each $t \in \mathbb{R}$.

Proof: The pole orbits in the (u, r) plane are given by $u_{1,2} = \pm g(r)$ and the "curves" of constant t by the straight lines $L_t: u = (r/\alpha) + (t/c)$. Thus, the gradient of L_t is $\alpha^{-1} > 1$, whereas, from the lemma above, $|g'| < 1$. Hence, for each t, L_t intersects *each* of the curves $u_{2,1}$ in *one* and only one point, giving rise to the distinct solutions $(g(r_2), r_2, t)$ and $(-g(r_1), r_1, t)$, where $r_1 \neq r_2$ since $\alpha > 0$.

Theorem 2: In the (x, t) plane the pole orbits Γ_1 and Γ_2 are distinct for all values of $\rho \in (1, \infty)$, i.e., there is never any collision or overtaking.

Proof: From (5.2) and Theorem 1, the pole orbits are given by

$$x_2(t) = ag(r_2) + br_2 \quad \text{and} \quad x_1(t) = -ag(r_1) + br_1,$$

subject to the constraint $t = t_2 = t_1 \Rightarrow \alpha(g(r_2) + g(r_1)) = r_2 - r_1$. Therefore, $x_2 - x_1 = a[g(r_2) + g(r_1)] + b(r_2 - r_1) = 2[g(r_2) + g(r_1)]/(\rho + 1)$. Hence, using the lemma and Theorem 1, we have

$$x_2 - x_1 > 2(|r_2| + |r_1|)/(\rho + 1) > 0 \Rightarrow x_2 > x_1 \ \forall\ t \in \mathbb{R}. \quad ■$$

Theorem 3: $x_2(t)$ and $x_1(t)$ are monotonically increasing functions of t with closest approach at $t = 0$. □

Proof: Consider the constraint equations $t = t_2 = t_1$ and differentiate them with respect to t. This gives $(\cdot \equiv d/dt)$

$$\dot{r}_k = -\rho(\rho - 1)/[1 - (-1)^k \alpha g'(r_k)], \qquad (5.3a)$$

$$\ddot{r}_k = -(-1)^k g''(r_k)\dot{r}_k^3/\rho(\rho + 1), \qquad (5.3b)$$

where $k = 1, 2$. Then, since $\alpha < 1$ and, from the lemma, $|g'| < 1$ and $g'' > 0$, we have $-\rho(\rho^2 - 1) < 2\dot{r}_k < 0$, $\ddot{r}_1 < 0$, and $\ddot{r}_2 > 0$.

Differentiating (5.2a) with respect to t and using (5.3) and the constraint equations gives

$$\dot{x}_k = (\rho^2 + \rho + 1) + 2\dot{r}_k/(\rho - 1), \qquad (5.4a)$$

$$\ddot{x}_k = [2/(\rho - 1)]\ddot{r}_k. \qquad (5.4b)$$

Then, from (5.4a) and the inequality on $\dot{r}_k$, we have

$$\dot{x}_k > (\rho^2 + \rho + 1) - \rho(\rho + 1) = 1 \ \forall\ t \in \mathbb{R}.$$

Hence $x_k(t)$ are strictly increasing functions of t.

A stationary value of $(x_2 - x_1)$ occurs when $\dot{x}_2 = \dot{x}_1$, which, from (5.4a), $\Rightarrow \dot{r}_2 = \dot{r}_1$. Using (5.3a) translates this into the condition $g'(r_2) + g'(r_1) = 0 \Rightarrow r_2 + r_1 = 0$ since g' is odd and globally univalent. Substituting into the constraint equations, $t = t_2 = t_1$, then gives $t = 0$. Hence $(x_2 - x_1)$ has a unique stationary value at $t = 0$.

To check that it is a minimum, we consider $(\ddot{x}_2 - \ddot{x}_1)$. From (5.4b), $(\rho - 1)(\ddot{x}_2 - \ddot{x}_1) = 2(\ddot{r}_2 - \ddot{r}_1) > 0 \ \forall\ t \in \mathbb{R}$ since, as shown above, $\ddot{r}_1 < 0$ and $\ddot{r}_2 > 0 \ \forall\ t \in \mathbb{R}$. Hence the orbits $x_2(t)$ and $x_1(t)$ have their closest approach at $t = 0$. ■

Corollaries: (i) $x_2(t)$ and $x_1(t)$ are smooth repulsive orbits, i.e., $x_2 > x_1$ and $\ddot{x}_1 < 0 < \ddot{x}_2 \ \forall\ t \in \mathbb{R}$. (ii) $(x_2 - x_1)$ is a positive, even function of t. □

Theorem 4: In the asymptotic regions of the (x,t) plane, i.e., $|x|, |t| \to \infty$, the trajectories of the poles coincide with those of the peaks of the soliton waveforms. □

Proof: The behavior of the quantities $g(r)$, $x(r)$, and $t(r)$, as $|r| \to \infty$, can be computed from (5.1) and (5.2), and are as follows:

$$g(r) \sim |r| - \ln\alpha.$$

$$\rho(\rho - 1)t_k \sim (-1)^k \alpha(|r_k| - \ln\alpha) - r_k,$$

$$x_k \sim (-1)^k(|r_k| - \ln\alpha) - r_k + t,$$

where $k = 1, 2$ and $t = t_1 = t_2$. Recalling that $\rho = (\alpha_1|\alpha_2)$, where α_1^2 and α_2^2 are the speed of the faster and slower solitons respectively, and transforming back to original (x,t) coordinates [see Eq. (4.5)] then leads to the trajectories

$$\alpha_2 x_2^- \sim \alpha_2^3 t - \ln\alpha,$$

$$\alpha_1 x_1^- \sim \alpha_1^3 t + \ln\alpha, \quad r_k \to \infty, \ t \to -\infty,$$

$$\alpha_1 x_2^+ \sim \alpha_1^3 t - \ln\alpha,$$

$$\alpha_2 x_1^+ \sim \alpha_2^3 t + \ln\alpha, \quad r_k \to -\infty, \ t \to \infty.$$

Hence, asymptotically, the poles move at the same speed as the solitons. Furthermore, by comparison with Eq. (3.10), the phase shifts are $(\alpha_1|2)(x_2^+ - x_1^-) = -\ln\alpha > 0$ for the faster pole and $(\alpha_2|2)(x_1^+ - x_2^-) = \ln\alpha < 0$ for the slower pole. These are the same as those obtained from an asymptotic analysis of the two-soliton waveforms.[18] Combining these results with our choice of phase for the solitons then proves the theorem. The resulting "picture" is as follows: At $t = -\infty$, the pole on the trajectory Γ_2 starts out with the slower soliton and gradually accelerates to the speed of the faster soliton, which catches it up as $t \to \infty$. Similarly, the pole on the trajectory Γ_1 starts out with the faster soliton and decelerates smoothly to the speed of the lower soliton, catching it up as $t \to \infty$. ■

Remark: In Ref. 18, the authors work with the equation $u_t - 6uu_x + u_{xxx} = 0$. Hence, in order to compare their results with ours, it is first of all necessary to make the transformation $(u, x, t) \mapsto (-u/6, -x, -t)$. Apart from changing the amplitudes, this has the effect of reversing the phase shifts.

We are now in a position to use the pole motions to interpret and analyze the dynamics of the two-soliton solutions, and the first step in this analysis is to decide on the manner in which the interaction takes place. If the taller, and hence faster, soliton is initially behind the shorter one, then the usual description of the ensuing motion is that the faster soliton overtakes the slower one and then proceeds on its way, the only effect of the collision being to introduce a phase shift into the system. Thus, in this interpretation, the individual solitons maintain their identities right throughout the period of their effective interaction.[14(b)] This is obviously a zero-order approximation to the actual state of affairs since it assumes that the solitons behave like rigid bodies undergoing an impulsive type of interaction.

A more realistic description was given by Zabusky and Kruskal,[19] who, on the basis of numerical experiments, made the following observations about the interaction. When $\rho \gg 1$, then the faster soliton absorbs the slower one and reemits it later, i.e., an apparently straightforward case of overtaking. However, when the speeds of the solitons are of the same order, they interact as follows: As soon as the faster soliton comes reasonably close to the slower one, the latter starts to grow while the former begins to shrink. This process continues until the solitons interchange their roles, after which they separate. These observations were later confirmed by Lax[10] in his theoretical analysis of the evolution of the peaks of the two-soliton solutions. However, in addition to the two different kinds of behavior described above, Lax also found a third way, intermediate between the other two, in which the interaction takes place. Lax's results can be summarized as follows: (a) If $\rho^2 > 3$, then the waveform evolves with the formation of a single maximum during the period of interaction. (b) If $2.62 < \rho^2 < 3$, then there are intervals during which the waveform has three maxima. (c) If $\rho^2 < 2.62$, then the waveform has two maxima at all times. Lax refers to the first two as overtaking modes and to the third as an exchange mode. In all cases, during the initial stage of the interaction, the height of the smaller soliton increases, while that of the larger soliton decreases. The final state of the interaction is an exact time reversal of this process, i.e., the amplitude of the smaller soliton decreases, while that of the larger soliton increases till they teach their asymptotic values. Thus, if we identify the solitons by their asymptotic amplitudes, then these amplitudes are certainly not conserved during the interaction.

Now, although the Lax analysis clarifies the description given by Zabusky and Kruskal, the interpretation of the waveforms in (a) and (b) as overtaking modes is ambiguous. Since in case (a) the waveform has a single maximum during the period of strongest interaction, while in case (b) it has three maximam, one could equally well argue for an exchange of identities as the solitons lose themselves in the total profile. In fact this is the picture presented by the pole-motions. From Theorem 2 above, we see that the faithful poles neither collide with nor overtake each other; rather there is a smooth transfer of speed from the trailing to the leading pole, which eventually leads to an interchange of their asymptotic speeds. Thus, if we assume, as we did in our study on the SGE, that the poles can be identified with the center of mass of the individual solitons, then we are forced to conclude that KdV solitons *exchange*, rather than preserve, their identities during an interaction for *all* values of the speed ratio. Note, however, that as $\rho \to \infty$, $\alpha \to 1$, and $g(r) \to |r|$. Thus, both the phase shift and the distance of closest approach of the poles tends to zero, while the pole orbits Γ_1 and Γ_2 degenerate into the straight lines

$$x_1 = \alpha_1^2 t, \quad x_2 = \alpha_2^2 t, \quad t < -\epsilon < 0,$$

$$x_1 = \alpha_2^2 t, \quad x_2 = \alpha_1^2 t, \quad t > \epsilon > 0,$$

where ϵ is of order ρ^{-4} for $\rho \to \infty$. Therefore, even for moderately large speed ratios, the interaction becomes pointlike, and hence it is not surprising that the evolution of the waveform is interpreted as the faster soliton overtaking the slower one.

If we accept this "exchange" hypothesis, then the next step is to establish the mechanism by which it occurs. In our model, in which the KdV field is treated as a mass density, it can only take place by "convection" via the underlying field. That is, by a continuous transfer of mass from the trailing soliton to the underlying field and from the underlying field to the leading soliton, until the exchange of identities has been effectively completed, the process being similar to that of pushing air from one lobe to another of a corrugated rubber balloon. As a consequence, the forces acting on the solitons during the interaction are local in nature, rather than of the action-at-a-distance type, and are generated by the absorbtion and emission of mass, and hence momentum, by the underlying field. Apart from being correlated with the pole motions, another point in favor of this "exchange" interpretation is that it provides a natural explanation for the phase shift of the solitons. This is not the case with the "overtaking" description, in which the phase shift enters rather abruptly and its origin is obscure.

Having established a possible mechanism for the interaction, we can now compute the effective forces and potentials acting on the solitons. The calculation is a straightforward application of Newton's second law of motion, but it does have two nonstandard features. First, since the KdV is a unidirectional equation, the potentials, in the case of the two-soliton solution, will be one-sided. That is, the trailing soliton will always lose mass while the leading soliton always gains it. Secondly, because of these mass changes and the particular dependence of the soliton mass on its speed [Eq. (3.11)], we are forced to treat the poles as particles whose masses are not fixed, but vary as the square roots of their speeds. Now it is easy to check that the effect of such a power-law dependence of mass on speed, i.e., $m(v) \propto v^\beta$, $\beta > 0$, is to change the numerical factor in the expression for the kinetic energy from its fixed-mass value of half to $(\beta + 1)/(\beta + 2)$. In our case $\beta = \frac{1}{2}$, which means that a KdV pole paritcle has a kinetic energy of $3mv^2/5$. The potential energies $\phi_k(x_k)$ for the poles on the orbits Γ_1 and Γ_2 can now be obtained from the conservation equations

$$\phi_k(x_k) + \frac{3}{5} m_k \dot{x}_k^2 = \frac{36}{5} \alpha_k^3, \tag{5.5}$$

where $m_k = 12\sqrt{\dot{x}_k}$, $\phi_k \to 0$ as $x_k \to -\infty$, and the α_k are as defined before. Note the unidirectional nature of these potentials, i.e., $\phi(-\infty) \neq \phi(\infty)$, and also that they are generated by the underlying field and not by the solitons. Now we have not been able to obtain an explicit solution for $\phi(x)$, but numerical computations, supported by asymptotic analysis, show that it has the shape of a modulated, tanh "kink" with the bend at the *high-speed* end being drawn out over a distance proportional to α_2^{-1}, while the bend at the *low-speed* end is confined to a distance of the order of α_1^{-1}. Thus, the force acting on the faster soliton changes much more slowly with distance than the force acting on the slower soliton.

Another feature, which emphasizes the local nature of the forces acting on the solitons, is that $|\ddot{x}_k|$ do not attain their maximum values at the point of closest approach. A small α ($\rho \to 1$) analysis leads to the following results: (1) At the point of closest approach, $x_2 = -x_1 \approx 2^{-1} \times (2 - 5\alpha)\ln(2|\alpha)$. (2) At the points where $|\ddot{x}_k|$ attain their maximum values, $x_2 = -x_1 \approx 4^{-1}(5 - 8\alpha)\ln(2/\alpha)$. (3) In a neighborhood of these maximum points we have

$$\ddot{x}_k \propto (-1)^k \operatorname{sech}^2\{(\alpha/2)[4x_k - (-1)^k(5 - 8\alpha)\ln(2/\alpha)]\}.$$

Thus, from (1) and (2), we see that the trailing soliton has its maximum deceleration *before* the point of closest approach, while the leading soliton has its maximum acceleration at the corresponding symmetrical point *after* closest approach. This implies that the curve $[x_2(t) - x_1(t)]$ is fairly flat in a neighborhood of $t = 0$. Note further that since the forces acting on the solitons are proportional to $\dot{x}^{1/2}\ddot{x}$, the maximum values of these forces occur somewhat further along from the point of closest approach than the maximum values of $|\ddot{x}_k|$.

Finally we see, from the expressions for $\ddot{x}_k$ given in (3) above, that the range of the interaction is of the order of α^{-1}. Thus, solitons moving with nearly the same speeds interact with the underlying field over a large part of their trajectories, irrespective of the actual values of these speeds. This behavior is consistent with the exchange hypothesis since, from (1) above, as $\alpha \to 0$ the separation at closest approach tends to infinity and hence the convection process has to take place over larger and larger distances. On the other hand, numerical computations show that as α increases the range begins to decrease more rapidly than α^{-1}. Thus, for all but the smallest values of α, the forces acting on the solitons have a relatively short range.

VI. CONCLUDING REMARKS

In this paper, we have shown how methods developed for the SGE can be extended to the KdV equation to obtain a consistent particle interpretation of free and interacting Korteweg–de Vries solitons. The main assumptions behind the interpretation are, first, that the KdV field can be treated as a Newtonian mass density and, second, that the faithful poles of the soliton modes of this mass density represent particles which move in step with the center of mass of the individual solitons in the solution. Thus, the dynamics of the translational modes of the solitons can be computed from the orbits of these faithful poles.

Working explicitly with the one- and two-soliton solutions, we have (1) shown that such a faithful representation exists for all values of the speed ratio, (2) isolated the faithful poles, and (3) analyzed both the kinematics and dynamics of the corresponding pole motions in some considerable detail. However, the results are not as nice as those we obtained for the SGE. Apart from the presence of a background of random poles and singlets, the special properties of the KdV and its solitons has meant the representative particles are forced to have variable masses and to interact via a mediating field rather than directly. Furthermore, the potentials are unidirectional and, at the mathematical level, the dynamics is obtained in parametric form and not explicitly as was the case for the SGE.

The last shortcoming is due to our lack of success in unravelling the analytic structure of the implicit equation.

$$g(r) = \ln[\cosh r + (\cosh^2 r - \alpha^2)^{1/2}] - \ln\alpha$$

for *all* values of $\alpha \in (0,1)$. This is not a trivial task, since the object of the exercise is to obtain explicit, and *recognizable*, closed-form expressions for the orbits, accelerations, forces, etc. For example, in a recent paper[20] Hagedorn and Rafelski analyzed a similar equation

$$g(r) = \ln[(1 + \alpha)g(r) + (\alpha - r)] - \ln\alpha,$$

known as the "boot strap equation," and obtained an explicit integral representation of $g(r)$ for the case $\alpha = 1$. However, although their analysis is fairly complicated, the advantage gained is numerical rather than analytical. That is, it merely leads to a representation which may be directly used for making numerical calculations of high precision, rather than to a representation which can be used for obtaining explicit, closed-form expressions. Thus, applying a similar procedure to our equation does not lead to useful results, and the problem still remains. On the other hand, we have obtained explicit formulae for those cases in which α is close to its extreme values (i.e., 0 and 1).

Apart from the defects mentioned above, most of which are intrinsic to the KdV, the pole-particle picture is both interesting and informative. It gives us both qualitative and quantitative information on the forces acting on the solitons and their reactions to these forces, and also provides a natural explanation of the phase shifts that occur as a result of these interactions. The significant information in this case is that the KdV forces are repulsive, generated locally, and have a relatively short range and that the solitons exchange their identities during collision.

Now both in this paper and in our study of the SGE, we have concentrated on the soliton solutions because of their physical importance and their obvious particlelike behavior. However, if one looks at the question from a purely mathematical angle, then it is possible to associate classical particle motions with other solutions as well, for example, the rational and elliptic solutions of the KdV.[3,4]

We have presented a detailed analysis of this approach and have come to the conclusion that the established equivalences (between these solutions and their corresponding many-body problems) merely have a formal significance. This is based on the following observations. First, at the classical level, there are no compelling physical reasons for requiring particle representations of solution other than the soliton solutions. Second, the pole flows of these other solutions are, in general, generated, *not* by the Hamiltonians of the associated particle problems, but, rather, by other constants of motion whose physical meaning is obscure. Third, as a consequence of the above, the equivalent many-body problems are not unique. Another point worth mentioning, in the case of the KdV, is that the rational and elliptic solutions have a relatively simple pole structure, whereas, as we have seen, the pole structure of the soliton solutions can be extremely complex.

Finally, we see two ways in which our analysis can be improved. First, there is the difficult problem of obtaining explict solutions to the parametric equations of motion and, second, there is the question of reconciling the pole picture with the specific interaction regimes found by Lax in his analysis of the waveforms.[10] We think that the singlets may have something to do with the latter.

[1]A. C. Bryan, C. R. Haines, and A. E. G. Stuart, Lett. Math. Phys. **2**, 445 (1978).

[2]A. C. Bryan, C. R. Haines, and A. E. G. Stuart, Nuovo Cimento B **58**, 1 (1980).

[3] H. Airault, H. P. McKean, and J. Moser, Comm. Pure Appl. Math. **30**, 95 (1977).

[4] D. V. Choodnovsky and G. V. Choodnovsky, Nuovo Cimento B **40**, 339 (1977).

[5] H. Hancock, *Theory of Elliptic Functions* (Dover, New York, 1958).

[6] G. Bowtell and A. E. G. Stuart, Phys. Rev. D **15**, 3580 (1977).

[7] W. R. Thickstun, J. Math. Anal. Appl. **55**, 335 (1976).

[8] M. D. Kruskal, Lect. Appl. Math. **15**, 61 (1974).

[9] G. Bowtell and A. E. G. Stuart, Prog. Theor. Phys. **64**, 68 (1980).

[10] P. D. Lax, Comm. Pure Appl. Math. **21**, 467 (1968).

[11] J. Moser, Advan. Math. **16**, 197 (1975).

[12] F. Calogero, Nuovo Cimento B **43**, 177 (1978).

[13] G. Bowtell and A. E. G. Stuart, report, City University, London, U. K. Department of Mathematics, (1980).

[14] (a) M. B. Fogel, S. E. Trullinger, A. R. Bishop, and J. A. Krumhansl, Phys. Rev. Lett. **36**, 1411 (1976); **37**, 314 (E) (1976). (b) D. J. Daup and A. C. Newell, Proc. Roy. Soc., London A **361**, 413 (1978).

[15] V. E. Zakharov and L. D. Faddeev, Funct. Anal. Appl. **5**, 280 (1971).

[16] T. B. Benjamin, J. L. Bona, and J. J. Mahony, Phil. Trans. Roy. Soc. London **272**, 47 (1972).

[17] R. Hirota, Phys. Rev. Lett. **27**, 1192 (1971).

[18] M. Wadati and M. Toda, J. Phys. Soc. Jpn. **32**, 1403 (1972).

[19] N. J. Zabusky and M. D. Kruskal, Phys. Rev. Lett. **15**, 240 (1965).

[20] R. Hagedorn and J. Rafelski, Commun. Math. Phys. **83**, 563 (1982).

Group-theoretical analysis of the sine-Gordon equation as a relativistic dynamical system

L. Martínez Alonso
Departamento de Métodos Matemáticos de la Física, Facultad de Ciencias Físicas, Universidad Complutense, Madrid-3, Spain

(Received 22 June 1982; accepted for publication 10 September 1982)

The inverse scattering transform method is used in order to achieve a group-theoretical characterization of the sine-Gordon equation considered as a relativistic dynamical system. The action of the Poincaré group is formulated in terms of scattering data variables and the three basic components of the sine-Gordon field, namely, solitons, breathers and radiation are identified with simple canonical realizations of the Poincaré group. In particular, the nonlinear flow associated with the one-parameter group of pure Lorentz transformations is integrated by means of the inverse scattering transform method.

PACS numbers: 11.30.Cp, 11.30.Na

1. INTRODUCTION

The most characteristic property of solitons is that they are objects which in many respects resemble particles. But to what extent can this analogy be formulated in a precise mathematical way? Since Wigner's work[1] we have learned to describe particles mathematically by means of realizations of invariance Lie groups. Thus, quantum particles are associated with projective unitary representations[1,2] on Hilbert spaces and classical particles with canonical realizations[3,4] on phase spaces. The two main invariance Lie groups of physical theories which provide us with specific mathematical models of the concept of particle[5] are the Galilei and the Poincaré groups. Recently, we have analyzed[6] the nonlinear Schrödinger equation considered as a Galilean-invariant dynamical system and we have verified that from the group-theoretical point of view the solitons of this equation are classical Galilean particles. The fundamental tool used in our analysis was the inverse scattering transform method which enabled us to reduce the realization of the Galilei group to a direct product of well-known Galilean-invariant dynamical systems.

The aim of the present paper is to study the sine-Gordon equation

$$\phi_{tt} - \phi_{xx} + \frac{m^3}{\sqrt{g}} \sin\left(\frac{\sqrt{g}}{m}\phi\right) = 0, \quad m, g > 0, \tag{1.1}$$

as a relativistic dynamical systems according to Wigner's group-theoretical point of view. Here the invariance group is the Poincaré group G in two-dimensional space-time whose elements are of the form

$$(b,a,\sigma) = \exp(-b\hat{H})\exp(a\hat{P})\exp(\sigma\hat{K}), \quad b,a,\sigma \in \mathbb{R}, \tag{1.2}$$

where $\hat{H}$, $\hat{P}$, and $\hat{K}$ are the generators of time translations, space translations, and pure Lorentz transformations respectively. Given one element $(b,a,\sigma) \in G$, the sine-Gordon equation is invariant under the transformation

$$\phi(t,x) \rightarrow \phi'(t',x') = \phi(t,x), \tag{1.3}$$

$$t' = \gamma(t + vx) + b, \quad x' = \gamma(x + vt) + a,$$

where $\gamma = \cosh\sigma$ and $v = \tanh\sigma$. If Eq. (1.1) is formulated as an infinite-dimensional Hamiltonian system then the action (1.3) determines a canonical realization (CR) of G. At first sight this CR seems to be quite different from the usual ones like those describing relativistic particles or the free Klein-Gordon equation. Nevertheless, the inverse scattering transform method[7-9] for solving (1.1) provides a picture of this dynamical system which can be analyzed in terms of simple relativistic systems.[10]

Following Wigner's ideas we consider the CR of the Poincaré group as the fundamental mathematical object associated with the sine-Gordon equation and consequently we look for its decomposition as a direct product of simple CR's. This decomposition is attained by means of the inverse scattering transform. Scattering data variables define an appropriate local coordinate system for the phase space of (1.1) in which the generating functionals H, P, and K of the CR of G take a simple form. We notice that the expressions for H and P were already calculated by Takhtadzhyan and Faddeev.[10] However, the derivation of the corresponding expression for the generating functional K of pure Lorentz transformations is in our knowledge new and it constitutes the main contribution of the present paper since it leads us to a complete characterization of the Poincaré group realization in terms of scattering data variables.

It is well known that the inverse scattering transform reveals the existence of three basic independent components of the sine-Gordon field, namely, solitons, breathers and radiation. Our analysis provides the following group-theoretical characterization of both the decomposition and the basic components of the sine-Gordon field.

(1) The CR of the Poincaré group G associated with (1.1) is locally equivalent to a direct product of three distinct CR's $R_{\rm I}$, $R_{\rm II}$, and $R_{\rm III}$ which represent the solitons, the breathers, and the radiation component, respectively.

(2) The realization $R_{\rm I}$ is a direct product of a finite number of copies of the Poincaré group CR describing the elementary classical relativistic particle of mass 8 m^3/g. Consequently, from the group-theoretical point of view, solitons of the sine-Gordon equation are relativistic classical particles.

(3) The realization $R_{\rm II}$ is also a finite direct product of the form $R_B \otimes \cdots \otimes R_B$ where R_B is a nonelementary CR of

the Poincaré group. Moreover, it is found that R_B cannot be interpreted as a composite system of relativistic particles if the "manifest covariant" and the Poisson bracket involution conditions for particle positions are imposed. However, the explicit form of R_B turns to be very simple. It may be formulated in terms of four coordinates, two of them transforming as position and momentum observables and the other two remaining invariant under the action of G of the dynamical trajectories. In this way, the realization R_B which represents the breather solution of (1.1) describes a nonelementary relativistic particle with an internal structure. In the center-of-mass frame this internal structure evolves periodically in time.

(4) The realization $R_{\rm III}$ is the infinite-dimensional CR of the Poincaré group describing the Klein-Gordon field of mass m.

Incidentally, the expression of the generating functional K in terms of scattering data variables allows us to apply the inverse scattering transform method to its associated Hamiltonian flow which is given by

$$\phi_{tt} - x^2\phi_{xx} - x\phi_x + x^2\frac{m^3}{\sqrt{g}}\sin\left(\frac{\sqrt{g}}{m}\phi\right) = 0. \tag{1.4}$$

This equation has the interesting property that independently of the initial conditions solitons reduce to step functions centered at the origin as $t \to \pm\infty$. Moreover, there are solutions of (1.4) which suggest processes of creation and annihilation of soliton-antisoliton pairs.

2. THE CANONICAL REALIZATION OF THE POINCARÉ GROUP

The Hamiltonian formulation of Eq. (1.1) is obtained by considering an infinite-dimensional phase space V consisting of pairs $(\phi(x),\pi(x))$ of regular real functions such that

$$\lim_{|x|\to\infty}\phi(x) = 0 \quad \left(\operatorname{mod}\frac{m}{\sqrt{g}}2\pi\right), \quad \lim_{|x|\to\infty}\pi(x) = 0. \tag{2.1}$$

The symplectic structure is defined by means of the following Poisson bracket operation

$$\{F_1, F_2\} = \int_{-\infty}^{\infty}\left(\frac{\delta F_1}{\delta\phi(x)}\frac{\delta F_2}{\delta\pi(x)} - \frac{\delta F_1}{\delta\pi(x)}\frac{\delta F_2}{\delta\phi(x)}\right)dx. \tag{2.2}$$

Then we can express Eq. (1.1) as the Hamiltonian system

$$\partial_t\phi(x) = \{\phi(x),H\}, \quad \partial_t\pi(x) = \{\pi(x),H\}, \tag{2.3}$$

where the Hamiltonian functional H is

$$H = \frac{1}{2}\int_{-\infty}^{\infty}\left\{\pi^2 + \phi_x^2 + \frac{2m^4}{g}\left[1 - \cos\left(\frac{\sqrt{g}}{m}\phi\right)\right]\right\}dx. \tag{2.4}$$

As a consequence the evolution law $U(t):\phi(0,x),\pi(0,x)) \to (\phi(t,x),\pi(t,x))$ associated with the sine-Gordon equation is a one-parameter group of canonical transformations over V.

Let us consider the action (1.3) of the Poincaré group G on the solutions of the sine-Gordon equation. It determines a correspondence between initial data which defines a realization $g \to R(g)$ of G on the phase space V given by

$$(R(g)\phi)(x) = (U(\tilde{t})\phi)(\tilde{x}),$$

$$(R(g))\pi(x) = \gamma[(U(\tilde{t})\pi)(\tilde{x}) - v(\partial_x(U(\tilde{t})\phi))(\tilde{x})], \tag{2.5}$$

where

$$\tilde{t} = -\gamma(b + v(x-a)), \quad \tilde{x} = \gamma(x + vb - a). \tag{2.6}$$

Observe that the appearance of the evolution map U in the expressions (2.5) implies that $R(g)$ is a nonlinear realization of the Poincaré group. Moreover, from (2.5) one deduces at once that the generators of $R(g)$ are Hamiltonian fields on V. That is, for each element $\hat{A}$ in the Lie algebra of G there is a functional $A[\phi,\pi]$ such that

$$\frac{d}{da}\bigg|_{a=0} R(\exp(a\hat{A}))\begin{pmatrix}\phi(\mathrm{x})\\ \pi(\mathrm{x})\end{pmatrix} = \begin{pmatrix}\{\phi(x), A\}\\ \{\pi(x), A\}\end{pmatrix}. \tag{2.7}$$

The functional associated with the generator $\hat{H}$ of time translations is the Hamiltonian H defined in (2.4) and the two other basic generators of G are represented by the functionals

$$P = -\int_{-\infty}^{\infty}\phi_x\pi\,dx,$$

$$K = -\frac{1}{2}\int_{-\infty}^{\infty}x\left\{\pi^2 + \phi_x^2 + \frac{2m^4}{g}\times\left[1 - \cos\left(\frac{\sqrt{g}}{m}\phi\right)\right]\right\}dx. \tag{2.8}$$

They satisfy the Poisson bracket relations

$$\{H,P\} = 0, \quad \{K,H\} = -P, \quad \{K,P\} = -H, \tag{2.9}$$

which characterize the Lie algebra structure of the Poincaré group. In this way, the action (2.6) of the Poincaré group is an infinite-dimensional CR.

3. THE INVERSE SCATTERING TRANSFORM OF THE GROUP REALIZATION

A. Scattering data variables

Let us perform the transformation

$$u(t,x) = \frac{\sqrt{g}}{m}\phi\left(\frac{t}{m},\frac{x}{m}\right), \tag{3.1}$$

which reduces the sine-Gordon equation (1.1) to the usual form

$$u_{tt} - u_{xx} + \sin u = 0. \tag{3.2}$$

In order to describe the inverse scattering transform of (3.2) we follow the method due to Zakharov-Takhtadzhyan-Faddeev[9] based on the spectral problem

$$L(\lambda,u,w)F(\lambda,x) = 0, \quad L(\lambda,u,w) \equiv J\partial_x + A + \frac{1}{\lambda}B - \lambda, \tag{3.3}$$

where $F = F(\lambda,x)$ is a 2×2 matrix, and J, A, and B are 2×2 matrices defined by

$$J = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}, \quad A = \frac{i}{4}\begin{pmatrix}0 & w\\ w & 0\end{pmatrix},$$

$$B = \frac{1}{16}\begin{pmatrix}\exp(iu) & 0\\ 0 & \exp(-iu)\end{pmatrix}, \tag{3.4}$$

$$w = u_x + u_t.$$

Let us consider the solutions $F_\pm$ of (3.3) which satisfy the asymptotic conditions

$$F_+(\lambda,x) \underset{x\to+\infty}{\to} E(\lambda,x), \quad F_-(\lambda,x) \underset{x\to-\infty}{\to} E(\lambda,x), \tag{3.5}$$

where

$$E(\lambda,x) = (e(\lambda,x),e^*(\lambda,x)),$$

$$e(\lambda,x) = \exp\left[i\left(\lambda - \frac{1}{16\lambda}\right)x\right]\begin{pmatrix}1\\ i\end{pmatrix}. \tag{3.6}$$

The essential information about scattering data variables is related with the transition matrix $T(\lambda)$ defined by the relation

$$F_+(\lambda,x) = F_-(\lambda,x)T(\lambda), \quad \lambda\in\mathbb{R} - \{0\}. \tag{3.7}$$

This matrix turns out to be of the form

$$T = \begin{pmatrix} a & -b^* \\ b & a^* \end{pmatrix}, \tag{3.8}$$

where a and b are two complex-valued functions depending on λ and satisfying the properties

$$|a|^2 + |b|^2 = 1, \quad a(\lambda) = a(-\lambda)^*, \quad b(\lambda) = b(-\lambda)^*. \tag{3.9}$$

The function $a(\lambda)$, the first column $f_1(\lambda,x)$ of the matrix $F_+(\lambda,x)$, and the second column $g_2(\lambda,x)$ of the matrix $F_-(\lambda,x)$ may be analytically continued to the half-plane $\mathrm{Im}k > 0$. Moreover,

$$\lim_{|\lambda|\to\infty} a(\lambda) = 1. \tag{3.10}$$

It follows that for each zero λ_k of $a(\lambda)$ a complex number b_k exists verifying

$$f_1(\lambda_k,x) = b_k\, g_2(\lambda_k,x). \tag{3.11}$$

We shall assume that the following two conditions are satisfied

(i) $a(\lambda)$ has no zeros on the real axis,
(ii) the zeros of $a(\lambda)$ are simple zeros.

In this case, the relevant set of scattering data which determines the potential functions u and w uniquely is given by

$$\{\lambda_k, m_k, r(\lambda)\}, \quad k = 1,\ldots,N, \quad \lambda > 0, \tag{3.12}$$

where

$$m_k = -ib_k\left(\frac{\partial a}{\partial\lambda}(\lambda_k)\right)^{-1}, \quad r(\lambda) = \frac{b(\lambda)}{a(\lambda)}. \tag{3.13}$$

The zeros λ_k and their corresponding coefficients m_k are distributed symmetrically with respect to the imaginary axis. As a consequence the discrete part of the scattering data may be divided into two different subsets $\{\lambda_j,m_j : j = 1,\ldots,N_1\}$ and $\{\lambda_l,m_l : l = 1,\ldots,N_2\}$ $(N_1 + 2N_2 = N)$ according to the following two cases

$$(1)\ \lambda_j = -\lambda_j^* = ik_j, \quad k_j > 0, \tag{3.14}$$

$$(2)\ \mathrm{Re}\,\lambda_l > 0, \quad \mathrm{Im}\,\lambda_l > 0. \tag{3.15}$$

The Poisson bracket relations among the scattering data were calculated by Faddeev and Takhtadzhyan[10] who introduced the following set of variables

$$\{\tilde q_j, \tilde p_j, \eta_l, \xi_l, \phi_l, \theta_l, \phi(\lambda), \rho(\lambda)\}, \tag{3.16}$$

$$j = 1,\ldots,N_1, \quad l = 1,\ldots,N_2, \quad \lambda > 0,$$

where

$$\tilde q_j = 8\ln|c_j|, \quad \tilde p_j = \ln k_j, \quad \left(c_j \equiv m_j \frac{\partial a}{\partial\lambda}(\lambda_j)\right), \tag{3.17a}$$

$$\eta_l = 4\ln|d_l|, \quad \xi_l = 4\ln|\lambda_l|, \quad \left(d_l \equiv m_l \frac{\partial a}{\partial\lambda}(\lambda_l)\right), \tag{3.17b}$$

$$\phi_l = -16\arg d_l, \quad \theta_l = \arg\lambda_l, \tag{3.17c}$$

$$\phi(\lambda) = -\arg b(\lambda), \quad \rho(\lambda) = -\frac{8}{\pi\lambda}\ln|a(\lambda)|. \tag{3.17d}$$

Observe that the determination of the scattering data (3.12) from the variables (3.16) requires the additional specification of the following signs

$$\epsilon_j = \mathrm{sgn}(-im_j), \quad j = 1,\ldots,N_1. \tag{3.18}$$

According to the results of Ref. 10 we have that the Poisson bracket relations between two of the variables (3.16) are zero except for the following ones

$$\{\tilde q_j, \tilde p_{j'}\} = \frac{g}{m^2}\delta_{jj'}, \quad \{\eta_l,\xi_{l'}\} = \{\phi_l,\theta_{l'}\} = \frac{g}{m^2}\delta_{ll'}, \tag{3.19a}$$

$$\{\phi(\lambda), \rho(\lambda')\} = \frac{g}{m^2}\delta(\lambda - \lambda'). \tag{3.19b}$$

We notice that because of (2.2) and (3.1) our Poisson bracket definition differs from the one used in Ref. 10 by a factor g/m^2.

B. Poincaré generators in terms of scattering data variables

As it was shown by Faddeev and Takhtadzhyan,[10] the functionals H and P associated with the generators of time and space translations may be expressed in terms of scattering data variables as follows

$$\begin{aligned} H = \frac{m^3}{g}\Bigg[& \sum_j (e^{-\tilde p_j} + 16e^{\tilde p_j}) \\ & + \sum_l 2\sin\theta_l(e^{-\xi_l/4} + 16e^{\xi_l/4}) \\ & + \int_0^\infty \left(\frac{1}{8\lambda} + 2\lambda\right)\rho(\lambda)\,d\lambda\Bigg], \end{aligned} \tag{3.20}$$

$$\begin{aligned} P = \frac{m^3}{g}\Bigg[& \sum_j (e^{-\tilde p_j} - 16e^{\tilde p_j}) \\ & + \sum_l 2\sin\theta_l(e^{-\xi_l/4} - 16e^{\xi_l/4}) \\ & + \int_0^\infty \left(\frac{1}{8\lambda} - 2\lambda\right)\rho(\lambda)\,d\lambda\Bigg]. \end{aligned} \tag{3.21}$$

We are going to prove that the functional K associated with the generator of pure Lorentz transformations also admits a representation of this kind which is given by

$$K = \frac{m^2}{g}\left[\sum_j \tilde q_j + \sum_l 4\eta_l + \int_0^\infty \lambda\frac{\partial\phi(\lambda)}{\partial\lambda}\rho(\lambda)\,d\lambda\right]. \tag{3.22}$$

In order to derive this expression it is convenient to introduce the following notation convention. Given a function $F = F[\phi,\pi]$ we will denote the variation of F under pure Lorentz transformations by

$$\delta F \equiv \frac{d}{d\sigma}\bigg|_{\sigma=0} F[R(\exp(\sigma\hat K))\phi, R(\exp(\sigma\hat K))\pi]. \tag{3.23}$$

Because of the canonical character of the realization of the Poincaré group it follows that

$$\delta F = \{F,K\}. \tag{3.24}$$

This relation enables us to calculate the derivatives of K with respect to the scattering data variables by calculating the variations of these variables under pure Lorentz transformations. Firstly, we have that the variations of the functions u and w arising in the spectral problem (3.3) are

$$\delta u = x(u_x - w), \quad \delta w = x(\sin u - w_x) - w, \tag{3.25}$$

so it is easy to see that

$$\begin{aligned}\delta L &= \frac{d}{d\sigma}\bigg|_{\sigma=0} L(\lambda,u,w)\\ &= -A - x(A_x + 2[B,J])\\ &\quad - \frac{x}{\lambda}(-B_x + 2(AJB - BJA)).\end{aligned} \tag{3.26}$$

On the other hand, as was found by Zakharov, Takhtadzhyan, and Faddeev,[9] the sine-Gordon equation (3.2) takes the following form in terms of the matrices A and B:

$$A_t = A_x + 2 + [B,J], \quad B_t = -B_x + 2(AJB - BJA). \tag{3.27}$$

Therefore, for a solution F of (3.3) we have

$$\begin{aligned}\delta LF &= -(A + x\partial_t L)F = -(A - xL\partial_t)F\\ &= -AF + L(xF_t) - JF_t,\end{aligned} \tag{3.28}$$

where the identity $xL = Lx - J$ has been used. Moreover, it is known that[9,11]

$$F_t = F_x - (2/\lambda)JBF, \tag{3.29}$$

and this implies

$$JF_t + AF = \left(L + \frac{1}{\lambda}B + \lambda\right)F = L\left(F + \lambda\frac{\partial F}{\partial\lambda}\right). \tag{3.30}$$

From (3.28)–(3.30) and taking into account that $\delta LF + L\delta F = 0$ we obtain

$$L(\delta F + xF_x - \lambda F_\lambda - 2(x/\lambda)JBF - F) = 0. \tag{3.31}$$

This equation together with the asymptotic conditions (3.5) leads us to the following expression for the variation of the solution F_+

$$\delta F_+ = -x\partial_x F_+ + \lambda\partial_\lambda F_+ + 2(x/\lambda)JBF_+. \tag{3.32}$$

Now it is plain to calculate the variations of the spectral data. One gets

$$\begin{aligned}&\partial\lambda_k = -\lambda_k, \quad \delta b_k = 0,\\ &\delta a(\lambda) = \lambda\frac{da(\lambda)}{d\lambda}, \quad \delta b(\lambda) = \lambda\frac{db(\lambda)}{d\lambda},\end{aligned} \tag{3.33}$$

or equivalently, in terms of the variables (3.17)

$$\delta\tilde{q}_j = 0, \quad \delta\tilde{p}_j = -1,$$

$$\delta\eta_l = \delta\phi_l = \delta\theta_l = 0, \quad \delta\xi_l = -4, \tag{3.34a}$$

$$\delta\phi(\lambda) = \lambda\frac{d\phi(\lambda)}{d\lambda}, \quad \delta\rho(\lambda) = \rho(\lambda) + \lambda\frac{d\rho(\lambda)}{d\lambda}. \tag{3.34b}$$

Consequently, according to (3.24) and the Poisson bracket relations (3.19) we deduce from (3.34) that

$$\frac{\partial K}{\partial\tilde{p}_j} = \frac{\partial K}{\partial\xi_l} = \frac{\partial K}{\partial\theta_l} = \frac{\partial K}{\partial\phi_l} = 0,$$

$$\frac{\partial K}{\partial\tilde{q}_j} = \frac{1}{4}\frac{\partial K}{\partial\eta_l} = \frac{m^2}{g}, \tag{3.35a}$$

$$\frac{\delta K}{\delta\rho(\lambda)} = \frac{m^2}{g}\lambda\frac{d\phi(\lambda)}{d\lambda},$$

$$\frac{\delta K}{\delta\phi(\lambda)} = -\frac{m^2}{g}\left(\rho(\lambda) + \lambda\frac{d\rho(\lambda)}{d\lambda}\right). \tag{3.35b}$$

These relations imply the identity (3.21).

The inverse scattering theory for the spectral problem (3.5) shows[9,10] that the set (3.16) of scattering data determines the functions u and w of (3.3) uniquely. Then, we may consider this set of variables as a local coordinate system for the phase space V. Therefore, Eqs. (3.20) and (3.21) expressing the generators of the Poincaré group CR in terms of this local coordinate system, define the group realization completely. As we shall see in the next section, the scattering data variables provide us with a picture of the Poincaré group CR which can be analyzed in terms of simple CR's.

4. ANALYSIS OF THE GROUP REALIZATION

A. Physical variables

The group realization R takes a more appropriate form when the following scattering data variables are introduced

$$q_j = -\frac{1}{m}\tilde{q}_j(e^{-\tilde{p}_j} + 16e^{\tilde{p}_j})^{-1},$$

$$p_j = \frac{m^3}{g}(e^{-\tilde{p}_j} - 16e^{\tilde{p}_j}), \tag{4.1}$$

$$Q_l = -\frac{2}{m}\eta_l\,[\sin\theta_l(e^{-\xi_l/4} + 16e^{\xi_l/4})]^{-1}, \tag{4.2a}$$

$$P_l = \frac{2m^3}{g}\sin\theta_l(e^{-\xi_l/4} - 16e^{\xi_l/4}),$$

$$\rho_l = \frac{m^2}{g}\phi_l - \frac{Q_l P_l}{\tan\theta_l}, \tag{4.2b}$$

$$q(k) = \phi(\lambda(k)), \quad p(k) = -\frac{m^2}{g}\frac{\partial\lambda}{\partial k}\rho(\lambda(k)),$$

$$\left[k \equiv m\left(\frac{1}{8\lambda} - 2\lambda\right)\right]. \tag{4.3}$$

One shows easily that the expressions (3.20)–(3.22) for the generators of the group realization reduce to

$$\begin{aligned}H = \sum_j (p_j^2 + M^2)^{1/2} + \sum_l (P_l^2 + M(\theta_l)^2)^{1/2}\\ + \int_{-\infty}^{\infty}(k^2 + m^2)^{1/2}p(k)\,dk,\end{aligned} \tag{4.4}$$

$$P = \sum_j p_j + \sum_l P_l + \int_{-\infty}^{\infty} kp(k)\,dk, \tag{4.5}$$

$$\begin{aligned}K = -\sum_j (p_j^2 + M^2)^{1/2}q_j - \sum_l (P_l^2 + M(\theta_l)^2)^{1/2}Q_l\\ - \int_{-\infty}^{\infty}(k^2 + m^2)^{1/2}\frac{\partial q(k)}{\partial k}p(k)\,dk,\end{aligned} \tag{4.6}$$

where

$$M = \frac{8m^3}{g}, \quad M(\theta_l) = 2M \sin \theta_l. \tag{4.7}$$

The phase space V may be described locally by means of the following set of scattering data variables

$$\{q_j, p_j, Q_l, P_l, \rho_l, \theta_l, q(k), p(k)\}, \tag{4.8}$$
$$j = 1,\dots,N_1, \quad l = 1,\dots,N_2, \quad k \in \mathbb{R}.$$

Moreover, from (3.19) and (4.1)–(4.3) we have the Poisson bracket relations

$$\{q_j, p_{j'}\} = \delta_{jj'}, \quad \{Q_l, P_{l'}\} = \{\rho_l, \theta_{l'}\} = \delta_{ll'}, \tag{4.9a}$$
$$\{q(k), p(k')\} = \delta(k - k'). \tag{4.9b}$$

All other Poisson brackets between two variables of the set (4.8) vanish. The continuous part of this set is represented by the two real functions $(q(k), p(k))$. Observe that $q(k)$ is an angle variable and the $p(k)$ is positive and vanishes at infinity. Therefore, if we introduce the functions

$$\tilde{\phi}(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} \frac{dk}{(2\omega)^{1/2}} \times (a(k)e^{ikx} + a^*(k)e^{-ikx}), \tag{4.10a}$$

$$\tilde{\pi}(x) = i(2\pi)^{-1/2} \int_{-\infty}^{\infty} dk \left(\frac{\omega}{2}\right)^{1/2} \times (a^*(k)e^{-ikx} - a(k)e^{ikx}), \tag{4.10b}$$

where

$$\omega = \omega(k) = (k^2 + m^2)^{1/2},$$
$$a(k) = p(k)^{1/2} \exp(-iq(k)), \tag{4.11}$$

then there is a bijective correspondence between the pairs of functions $(q(k), p(k))$ and $(\tilde{\phi}(x), \tilde{\pi}(x))$. Hence, we may define a local coordinate system for the phase space V by means of the correspondence

$$(\phi(x), \pi(x)) \rightarrow (q_j, p_j, Q_l, P_l, \rho_l, \theta_l, \tilde{\phi}(x), \tilde{\pi}(x)). \tag{4.12}$$

We now note that according to (4.9), (4.10), and (4.11) the Poisson bracket operation (2.2) adopts the form

$$\{F_1, F_2\} = \sum_j \left(\frac{\partial F_1}{\partial q_j}\frac{\partial F_2}{\partial p_j} - \frac{\partial F_1}{\partial p_j}\frac{\partial F_2}{\partial q_j}\right) + \sum_l \left[\left(\frac{\partial F_1}{\partial Q_l}\frac{\partial F_2}{\partial P_l} - \frac{\partial F_1}{\partial P_l}\frac{\partial F_2}{\partial Q_l}\right) + \left(\frac{\partial F_1}{\partial \rho_l}\frac{\partial F_2}{\partial \theta_l} - \frac{\partial F_1}{\partial \theta_l}\frac{\partial F_2}{\partial \rho_l}\right)\right] + \int_{-\infty}^{\infty} \left(\frac{\delta F_1}{\delta \tilde{\phi}(x)}\frac{\delta F_2}{\delta \tilde{\pi}(x)} - \frac{\delta F_1}{\delta \tilde{\pi}(x)}\frac{\delta F_2}{\delta \tilde{\phi}(x)}\right) dx. \tag{4.13}$$

Moreover, from (4.4)–(4.6) and (4.10), (4.11) the generators of the group realization can be written as

$$H = \sum_j (p_j^2 + M^2)^{1/2} + \sum_l (P_l^2 + M(\theta_l))^{1/2} + \frac{1}{2}\int_{-\infty}^{\infty} (\tilde{\pi}^2 + \tilde{\phi}_x^2 + m^2\tilde{\phi}^2)\, dx, \tag{4.14a}$$

$$P = \sum_j p_j + \sum_l P_l - \int_{-\infty}^{\infty} \tilde{\pi}\tilde{\phi}_x\, dx, \tag{4.14b}$$

$$K = -\sum_j (p_j^2 + M^2)^{1/2} q_j - \sum_l (P_l^2 + M(\theta_l)^2)^{1/2} Q_l - \frac{1}{2}\int_{-\infty}^{\infty} x(\tilde{\pi}^2 + \tilde{\phi}_x^2 + m^2\tilde{\phi}^2)\, dx. \tag{4.14c}$$

B. Decomposition of the group realization

The local coordinate system (4.12) allows us to analyze the properties of the sine-Gordon equation (1.1) considered as a relativistic invariant dynamical system. This is so due to the structure of the expressions (4.14) of the generators of the group realization R. It exhibits the fact that R is locally equivalent to a direct product of three distinct CR's of the Poincaré group

$$R = R_{\mathrm{I}} \otimes R_{\mathrm{II}} \otimes R_{\mathrm{III}}. \tag{4.15}$$

The first factor R_{I} is a direct product of N_1 identical CR's

$$R_{\mathrm{I}} = R_s \overset{N_1}{\otimes \cdots \otimes} R_s, \tag{4.16}$$

where R_s acts on the two-dimensional phase space $\mathbb{R}^2$ with canonically conjugate variables (q, p) and such that the generators of R_S are the functions

$$H_s = (p^2 + M^2)^{1/2}, \quad P_s = p, \quad K_s = -(p^2 + M^2)^{1/2} q. \tag{4.17}$$

This means that R_s is the well-known Poincaré group CR describing the free elementary particle of mass M in classical relativistic mechanics in two-dimensional space-time. Under the action R_S the coordinate p transforms as a momentum observable and the coordinate q satisfies the "manifest covariance" condition[4]

$$q'(t') = \gamma(q(t) + vt) + a, \quad t' = \gamma(t + vq(t)) + b. \tag{4.18}$$

That is to say, each point (t,q) on the trajectory of the particle transforms as an "event" in two-dimensional space-time. Therefore, we have that R_{I} describes a system of N_1 free elementary relativistic particles of mass M.

The second factor R_{II} is also a direct product of the form

$$R_{\mathrm{II}} = R_B \overset{N_2}{\otimes \cdots \otimes} R_B, \tag{4.19}$$

where R_B acts on a four-dimensional phase space with two pairs of canonically conjugate variables (Q,P) and (ρ,θ). The geometry of this phase space is better described by using the coordinate $\phi = (g/m^2)(\rho + QP/\tan\theta)$ instead of the coordinate ρ [see (3.17c) and (4.2b)]. In this way, the points of the phase space are specified by four coordinates (Q,P,ϕ,θ) in such a form that

$$Q,P \in \mathbb{R}, \quad \phi \in \mathbb{R}\ (\mathrm{mod}\ 32\pi), \quad 0 < \theta < \pi/2. \tag{4.20}$$

The generators of R_B are the functions

$$H_B = (P^2 + 4M^2 \sin^2\theta)^{1/2}, \quad P_B = P,$$
$$K_B = -(P^2 + 4M^2 \sin^2\theta)^{1/2} Q. \tag{4.21}$$

The question arises: What kind of relativistic system does R_B describe? It is not any elementary one since for two-dimensional space-time the elementary classical systems of the Poincaré group have a two-dimensional phase space. Then one is tempted to think of R_B as describing a system of two interacting relativistic particles. This is particularly appealing as the mass of the system is given by $2M \sin\theta$. Nevertheless, since the formulation of the "zero interaction theorem" by Currie–Jordan–Sudarshan[12] it is well-known that severe limitations arise for the description of relativistic dynamics within the framework of the canonical realizations of the Poincaré group.[13] In particular, for the two-dimensional space-time case, there is only one possible model of a relativistic system of interacting particles satisfying the conditions

(i) The coordinates q_i describing the particle positions verify the "manifest covariant" condition.

(ii) The Poisson brackets $\{q_i, q_j\}$ vanish.

This model consists of relativistic particles moving in a constant external field.[14] Clearly, our canonical realization R_B is not of this kind. Indeed, the equations of the motion determined by the Hamiltonian H_B are

$$\dot{Q} = PH_B^{-1}, \quad \dot{P} = \dot{\theta} = 0, \quad \dot{\phi} = \frac{g}{m^2} H_B (\tan\theta)^{-1}. \tag{4.22}$$

Then due to the condition (4.20) for ϕ it is deduced that in the center-of-mass frame (i.e., $P = 0$) the trajectories are periodic with a period given by

$$T = 2\pi/m \cos\theta. \tag{4.23}$$

Therefore, under the assumption of the conditions (i) and (ii) it is not possible to interprete R_B as describing a system of two interacting relativistic elementary particles. In part C of this section we will see that R_B represents the breather solution of the sine-Gordon equation. This solution is sometimes interpreted as a bound soliton-antisoliton pair.[15] However, the above analysis shows that this interpretation has not a clear group-theoretical basis. On the other hand, the analysis of the transformation properties of the phase space points under R_B shows a very simple structure; the coordinate P behaves as a momentum observable and Q, ρ, and θ verify

$$Q'(t') = \gamma(Q(t) + vt) + a, \quad t' = \gamma(t + vQ(t)) + b, \tag{4.24}$$

$$\rho'(t') = \rho(t), \quad \theta'(t') = \theta(t). \tag{4.25}$$

The derivation of the transformation law for ρ is rather involved and it is performed in the Appendix. Observe that in what concerns to the coordinates Q and P the system looks like a relativistic particle. We will refer to R_B as the "pulsating nonelementary relativistic particle." Thus, we have that the factor $R_{\rm II}$ of the descomposition (4.15) describes a system of N_2 noninteracting particles of this kind.

We notice that the numbers N_1 and N_2 of subsystems which appear in the descompositions (4.16) and (4.19) of $R_{\rm I}$ and $R_{\rm II}$ respectively are only locally constant in the phase space V. That is to say, for every point $(\phi = \phi(x), \pi = \pi(x))$ of V there is a neighborhood in which N_1 and N_2 remain constant but they are not constant on the whole of V.

The third factor $R_{\rm III}$ in the descomposition (4.15) is an infinite-dimensional CR of the Poincaré group. It acts on a phase space $\tilde{V}$ whose elements are pairs $(\tilde{\phi}(x), \tilde{\pi}(x))$ of real functions which vanish as $|x| \to \infty$. The symplectic structure on $\tilde{V}$ is determined in the usual form [see (2.2)]. The generators of $R_{\rm III}$ are the functionals

$$H_{\rm III} = \frac{1}{2}\int_{-\infty}^{\infty} (\tilde{\pi}^2 + \tilde{\phi}_x^2 + m^2\tilde{\phi}^2)\, dx,$$

$$P_{\rm III} = -\int_{-\infty}^{\infty} \tilde{\pi}\tilde{\phi}_x \, dx,$$

$$K_{\rm III} = -\frac{1}{2}\int_{-\infty}^{\infty} x(\tilde{\pi}^2 + \tilde{\phi}_x^2 + m^2\tilde{\phi}^2)\, dx. \tag{4.26}$$

Evidently, $R_{\rm III}$ describes the realization of the Poincaré group associated with the Klein–Gordon field equation with mass m

$$\tilde{\phi}_{tt} - \tilde{\phi}_{xx} + m^2\tilde{\phi}^2 = 0. \tag{4.27}$$

C. Solitons and breathers as relativistic particles

The inverse scattering transform theory for the spectral problem (3.3) provides explicit solutions of the sine-Gordon equation for the case in which the reflection coefficient $r(\lambda)$ [see (3.12) and (3.13)] vanishes. If the operator $L(\lambda, u, w)$ defines a set $\{\lambda_k, m_k, r(\lambda) = 0; k = 1,...,N\}$ of scattering data, then the function $u(x)$ is determined by[9]

$$u(x) = -2i \ln \frac{\det(\mathbb{1} + U)}{\det(\mathbb{1} - U)}, \tag{4.28}$$

where U denotes the $N \times N$ matrix

$$U_{kk'} = \frac{im_k}{\lambda_k + \lambda_{k'}} \exp\left\{\left[\left(\lambda_k - \frac{1}{16\lambda_k}\right) + \left(\lambda_{k'} - \frac{1}{16\lambda_{k'}}\right)\right] x\right\}. \tag{4.29}$$

In terms of the local coordinate system (4.12) the condition $r(\lambda) = 0$ is equivalent to $\tilde{\phi}(x) = \tilde{\pi}(x) = 0$. That is to say, this condition determines a region in the phase space V on which the group realization reduces to a direct product $R = R_{\rm I} \otimes R_{\rm II}$ of finite-dimensional CR's describing a system of classical relativistic particles. Therefore, we have that these elements of the phase space admit two distinct representations, either as a pair of field functions $(\phi(x), \pi(x))$ (wave representation) or as a point $(q_j, p_j, Q_l, P_l, \rho_l, \theta_l; j = 1,...,N; l = 1,...,N_2)$ in a finite-dimensional phase space (particle representation). By using (4.28) it is easy to relate both representations explicitly in the simplest cases. Here, we will indicate three examples. Firstly, we consider the case in which $N_1 = 1$, and $N_2 = 0$, then the group realization R reduces to the realization R_S of the elementary relativistic particle and Eq. (4.28) yields the following wave representation of the free particle motion

$$\phi_s(t,x) = \frac{4m}{\sqrt{g}} \tan^{-1}\left[\exp\left(\epsilon \frac{m}{M} H_s (x - q(t))\right)\right]. \tag{4.30}$$

Here ϵ stands for the sign (3.18) which, as mentioned above, must be incorporated to the set of scattering data in order to define the inverse scattering transform uniquely. The solution (4.30) is the so-called soliton ($\epsilon = 1$) or antisoliton ($\epsilon = -1$) solution of the sine-Gordon equation. Clearly, the present analysis proves that from the group theoretical point of view solitons are elementary classical relativistic particles. If we take $N_1 = 0$ and $N_2 = 1$ then the group realization R reduces to the realization R_B of the pulsating nonelementary relativistic particle. Its associated wave representation provided by (4.28) is the breather solution of the sine-Gordon equation

$$\phi_B(t,x) = \frac{4m}{\sqrt{g}} \tan^{-1}\left[\tan\theta \frac{\sin((m/2M)((m^2/g)\phi(t) - (P/\tan\theta)x))}{\cosh((m/2M)H_B(x - Q(t)))}\right]. \tag{4.31}$$

The dynamical behavior of this wave is a beautiful picture of the pulsating nature of the underlying nonelementary relativistic particle.[15] Finally, let us consider the case $N_1 = 2$ and $N_2 = 0$ with signs $\epsilon_1 = 1$ and $\epsilon_2 = -1$. Now $R = R_S \otimes R_S$ and then the group realization describes a noninteracting system of two elementary relativistic particles. From (4.28) one derives the following wave representation of this system

$$\phi_{s\bar{s}}(t,x) = \frac{4m}{\sqrt{g}} \tan^{-1}\left[c \frac{\sinh((m/2M)(h_1(x - q_1(t)) - h_2(x - q_2(t))))}{\cosh((m/2M)(h_1(x - q_1(t)) + h_2(x - q_2(t))))}\right],$$

$$h_i = (p_i^2 + M^2)^{1/2}, \quad c = \left|\frac{(h_1 + h_2) - (p_1 + p_2)}{(h_1 - h_2) - (p_1 - p_2)}\right|. \tag{4.32}$$

This is the soliton-antisoliton solution of the sine-Gordon equation. It follows easily from (4.32) that when the time goes to $-\infty$ the solution $\phi_{s\bar{s}}$ appears as the superposition of a soliton and an antisoliton with particle trajectories

$$q_1^{(\text{in})}(t) = q_1(t) + \frac{M}{2mh_1} \ln\left|1 - \frac{4M^2}{s}\right|, \tag{4.33a}$$

$$q_2^{(\text{in})}(t) = q_2(t) - \frac{M}{2mh_2} \ln\left|1 - \frac{4M^2}{s}\right|, \tag{4.33b}$$

$$(s \equiv (h_1 + h_2)^2 - (p_1 + p_2)^2),$$

respectively. Analogously, when $t \to +\infty$ $\phi_{s\bar{s}}$ is the superposition of a soliton and an antisoliton with particle trajectories

$$q_1^{(\text{out})}(t) = q_1(t) - \frac{M}{2mh_1} \ln\left|1 - \frac{4M^2}{s}\right|, \tag{4.34a}$$

$$q_2^{(\text{out})}(t) = q_2(t) + \frac{M}{2mh_2} \ln\left|1 - \frac{4M^2}{s}\right|. \tag{4.34b}$$

The "in" and "out" asymptotes do not coincide and it shows the presence of an interaction process between the solitons. In this way, when the dynamics is looked at from the point of view of the asymptotic soliton components the wave representation of this system of two noninteracting relativistic particles exhibits an interactinglike behavior.

5. FURTHER RESULTS

One interesting application which emerges from the decomposition of the generator K of pure Lorentz transformations in terms of scattering data variables is the integration of the Hamiltonian system

$$\partial_t \phi(x) = \{\phi(x), K\}, \quad \partial_t \pi(x) = \{\pi(x), K\}, \tag{5.1}$$

by means of the inverse scattering transform method. The corresponding field equation is the following nonlinear equation with x-dependent coefficients

$$\phi_{tt} - x^2\phi_{xx} - x\phi_x + x^2 \frac{m^3}{\sqrt{g}} \sin\left(\frac{\sqrt{g}}{m}\phi\right) = 0. \tag{5.2}$$

The evolution law of the coordinates (4.12) derives easily from (4.14c) and it takes the form

$$\dot{q}_j = -p_j q_j (p_j^2 + M^2)^{-1/2}, \quad \dot{p}_j = (p_j^2 + M^2)^{1/2}, \tag{5.3a}$$

$$\dot{Q}_l = -P_l Q_l (P_l^2 + M(\theta_l)^2)^{-1/2}, \quad \dot{P}_l = (P_l^2 + M(\theta_l)^2)^{1/2}, \tag{5.3b}$$

$$\dot{\rho}_l = -2M^2 \sin(2\theta_l) Q_l (P_l^2 + M(\theta_l))^{-1/2}, \quad \dot{\theta}_l = 0, \tag{5.3c}$$

$$\tilde{\phi}_{tt} - x^2\tilde{\phi}_{xx} - x\tilde{\phi}_x + m^2x^2\tilde{\phi} = 0. \tag{5.3d}$$

In particular, this yields

$$q_j(t) = \frac{q_j(0)\cosh\delta_j}{\cosh(t - \delta_j)}, \quad p_j(t) = x_{0j} e^{\delta_j} \sinh(t - \delta_j), \tag{5.4}$$

where

$$\delta_j = \ln\left|\frac{M}{x_{0j}}\right|, \quad x_{0j} = p_j(0) + (p_j(0)^2 + M^2)^{1/2}. \tag{5.5}$$

Analogous expressions are found for the functions $Q_1(t)$ and $P_1(t)$. We observe that independently of the initial conditions the following asymptotic properties are verified

$$q_j(t) \underset{t\to\pm\infty}{\longrightarrow} 0, \quad |p_j(t)| \underset{t\to\pm\infty}{\longrightarrow} \infty. \tag{5.6}$$

The trajectory $q_j(t)$ describes a bounded motion in which the particle leaves the origin at $t = -\infty$, reaches a maximum distance $q_j(0)\cosh\delta_j$ at the time $t = \delta_j$, and then returns, arriving to the origin at $t = +\infty$. This feature of the dynamics determines a curious behavior for the soliton solutions of (5.2). The one-soliton solution ($N_1 = 1, N_2 = 0, \tilde{\phi} = \tilde{\pi} = 0$) is given by

$$\phi_s(t,x) = \frac{4m}{\sqrt{g}} \tan^{-1}\left[\exp\left(\epsilon \frac{m}{M} H_s(t)(x - q(t))\right)\right]. \tag{5.7}$$

Here $H_S = (p^2 + M^2)^{1/2}$ depends on t and consequently the profile of ϕ_s deforms during the motion. Note that despite $H_S \to \infty$ as $t \to \pm\infty$ the product $H_S q$ is a constant of the motion since the Hamiltonian K of (5.2) reduces in this case to $-H_S q$. Therefore

$$\phi_s(t,x) \underset{t\to\pm\infty}{\longrightarrow} \frac{2\pi m}{\sqrt{g}} \theta(\epsilon x), \tag{5.8}$$

where $\theta(x)$ is the step function. It is also easy to analyze the breather and the soliton-antisoliton solutions of (5.2). To do it we only need to introduce the new expressions for the motion of the variables (Q,P,θ,ϕ) and (q_1, p_1, q_2, p_2) in Eq. (4.31) for ϕ_B and (4.32) for $\phi_{s\bar{s}}$, respectively. In this way, one finds that both $\phi_B(t,x)$ and $\phi_{s\bar{s}}(t,x)$ vanish as $t \to \pm\infty$. It is worth noting that the analysis of $\phi_{s\bar{s}}$ suggests that this solution describes a process in which a soliton-antisoliton pair is created at $t = -\infty$ and subsequently annihilated at

$t = +\infty$.

Finally we notice that under the change of variables

$$t' = -x \sinh t, \quad x' = x \cosh t, \tag{5.9}$$

Eq. (5.2) transforms into the sine-Gordon equation (1.1). Indeed, given a solution $\phi'(t,x)$ of the sine-Gordon equation then according to (2.5) it follows that $\phi(t,x) \equiv \phi'(t',x')$ is the image of $\phi'(0,x)$ under the pure Lorentz transformation $\exp(tK)$.

ACKNOWLEDGMENTS

I wish to thank L. Abellanas, A. F. Rañada, and A. Galindo for helpful discussions during the course of this work. The financial support of the Comisión Asesora de Investigación Científica y Técnica is also acknowledged.

APPENDIX

In order to prove Eq. (4.25) for the transformation law of the variable ρ it is convenient to express the Poincaré group realization R_B as follows

$$\begin{aligned} R_B(g) &= R_B(e^{-b\hat{H}} e^{a\hat{P}} e^{\sigma\hat{K}}) \\ &= R_B(e^{-b\hat{H}}) R_B(e^{a\hat{P}}) R_B(e^{\sigma\hat{K}}). \end{aligned} \tag{A1}$$

From the form (4.21) of the generators of R_B one deduces that under the action of $R_B(g)$ the energy H_B and the momentum P transform as

$$H'_B = \gamma(H_B + vP), \quad P' = \gamma(P + vH_B). \tag{A2}$$

The generator $K_B = -H_B Q$ and the variables ϕ and θ remain invariant under $R_B(\exp(\sigma\hat{K}))$ since they have a null Poisson bracket with K_B. Therefore

$$R_B(e^{\sigma\hat{K}})(Q,P,\phi,\theta) = (QH_B/H'_B, P', \phi, \theta). \tag{A3}$$

By means of the Poisson bracket relations among the variables Q, P, ϕ, and θ it is easy to complete the calculation of the action R_B on the phase space. Thus, it follows that

$$\begin{aligned} &R_B(g)(Q,P,\phi,\theta) \\ &= R_B(e^{-b\hat{H}})\left(\frac{QH_B}{H'_B} + a, P', \phi + a\frac{g}{m^2}\frac{P'}{\tan\theta}, \theta\right) \\ &= \Big((QH_B - bP')H'^{-1}_B + a, P', \phi \\ &\quad + \frac{g}{m^2 \tan\theta}(aP' - bH'_B), \theta\Big). \end{aligned} \tag{A4}$$

The evolution law of the variable ϕ and (A4) imply that the transformed variable ϕ' satisfies

$$\begin{aligned} \phi'(t') &= \phi'(0) + \frac{g}{m^2 \tan\theta} t' H'_B \\ &= \phi(0) + \frac{g}{m^2 \tan\theta}(aP' + (t' - b)H'_B). \end{aligned} \tag{A5}$$

Now, we notice that according to (4.24) and (A2) the two-component vectors $(t' - b, Q'(t') - a)$ and (H'_B, P') are the images under a pure Lorentz transformation of $(t, Q(t))$ and (H_B, P), respectively. As a consequence

$$(t' - b)H'_B = (Q'(t') - a)P' + tH_B - Q(t)P. \tag{A6}$$

Hence (A5) takes the form

$$\phi'(t') = \phi(t) + \frac{g}{m^2 \tan\theta}(Q'(t')P'(t') - Q(t)P(t)), \tag{A7}$$

and this leads at once to the equality $\rho'(t') = \rho(t)$.

---

[1] E. P. Wigner, Ann. Math. (N.Y.) **40**, 149 (1939).

[2] A. S. Whigtman, Rev. Mod. Phys. **34**, 845 (1962).

[3] I. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).

[4] E. C. G. Sudarshan and N. Mukunda, *Classical Dynamics: A modern perspective* (Wiley, New York, 1974).

[5] L. Martínez Alonso, J. Math. Phys. **18**, 1577 (1977); J. Math. Phys. **20**, 219 (1979).

[6] L. Martínez Alonso, "The nonlinear Schrödinger equation as a Galilean-invariant dynamical system," J. Math. Phys. **23**, 1518 (1982).

[7] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Phys. Rev. Lett. **30**, 1262 (1973); Phys. Rev. Lett. **31**, 125 (1973).

[8] G. L. Lamb, Phys. Rev. A **9**, 422 (1974).

[9] V. E. Zakharov, L. A. Takhtadzhyan, and L. D. Faddeev, Sov. Phys. Dokl. **19**, 842 (1975).

[10] L. A. Takhtadzhyan and L. D. Faddeev, Theor. Math. Phys. **21**, 1046 (1974).

[11] We notice that Eq. (9) of Ref. 9 contains a misprint. It should read $\psi_t = \psi_x - (2/\lambda)JH\psi$.

[12] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963).

[13] *The Theory of Action-at-a-Distance in Relativistic Particle Dynamics—A Reprint Collection*, edited by E. H. Kerner, International Science Review Series II (Gordon and Breach, New York, 1972).

[14] R. N. Hill, J. Math. Phys. **8**, 1756 (1967).

[15] See for instance G. Eilenberger, *Solitons, Springer Series in Solid-State Sciencies*, Vol. 19 (Springer-Verlag, Berlin, 1981).

Lorentz transformation properties of nonconserved charges

Carl M. Bender and David A. Williams
Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 4 November 1981; accepted for publication 6 August 1982)

If a current $j^\mu(x)$ is conserved, then the charge Q defined as the three-space integral of $j^0(x)$ is a Lorentz scalar. In this paper we investigate the Lorentz transformation properties of Q assuming that $j^\mu(x)$ is not conserved; that is, $\partial_\mu j^\mu(x) \neq 0$. We find that Q, which now depends on time, transforms as the infinite direct sum of the spin-0 parts of the $(0,2j+1)$ finite-dimensional tensor representations of the Lorentz group plus the spin-0 component of an infinite-dimensional indecomposable representation of the Lorentz group, $[(0,1)\rightarrow(1,0)]:[(0,1)\rightarrow(1,0)] \oplus \Sigma_{j=0\oplus}^{\infty}(0,2j+1)$, where we are using the Gel'fand, Milnos, and Shapiro notation. More simply, if $T^{\mu_1\mu_2\cdots\mu_n}(x)$ is an n-index traceless, symmetric tensor density and $S(x)$ is a density transforming as the spin-0 component of the infinite-dimensional representation, we find that Q transforms as the infinite sum $S(x)+T(x)+T^{00}(x)+T^{0000}(x)+T^{000000}(x)+\cdots$.

PACS numbers: 11.30.Cp, 11.40. − q, 02.20. + b

1. INTRODUCTION

It is well known that a conserved vector current density $j^\mu(x)$ gives rise to a time-dependent charge Q. A conserved current is one that satisfies

$$\partial_\mu j^\mu(x)=0. \tag{1.1}$$

The charge Q is defined as the space integral of the 0 component of the current

$$Q=\int d_3xj^0(\mathbf{x},t). \tag{1.2}$$

It is a simple theorem that if $j^\mu(x)$ transforms as a vector density and Eq. (1.1) holds, then the charge Q is a Lorentz scalar. To prove this theorem we use the assumed infinitesimal Lorentz transformation properties of $j^\mu(x)$

$$\frac{1}{i}[j^0(x),J^{0k}]=(x^k\partial^0-x^0\partial^k)j_0(x)+j^k(x), \tag{1.3a}$$

$$\frac{1}{i}[j^l(x),J^{0k}]=(x^k\partial^0-x^0\partial^k)j^l(x)+\delta^{kl}j^0(x), \tag{1.3b}$$

where J^{0k} is the generator of infinitesimal Lorentz transformations. Communting Q in (1.2) with J^{0k} and using (1.3a) gives

$$\frac{1}{i}[Q,J^{0k}]=\int d_3x\,x^k\partial^0j^0(x)-\int d_3xx^0\partial^kj^0(x)+\int d_3xj^k(x). \tag{1.4}$$

The second integral in (1.4) vanishes because it is a total divergence. The third integral can be rewritten as follows

$$\int d_3xj^k=\int d_3x\delta^{kl}j^l=\int d_3x(\partial_l x^k)j^l=-\int d_3xx^k\partial_lj^l,$$

where in the last term we have integrated by parts. Thus the first and third terms of (1.4) combine to give

$$-\int d_3xx^k\partial_\mu j^\mu(x),$$

which vanishes if the current $j^\mu(x)$ satisfies (1.1) (is conserved). The vanishing of the commutator in (1.4) implies that Q is a scalar. (In the above argument we have used the metric $x^k=x_k, x^0=-x_0$.)

It is trivial to show that if (1.1) holds, then the charge Q is time independent.

In this paper we answer the question how does the charge Q in (1.2) transform under the Lorentz group if the current is not conserved? [For example, suppose that $j^\mu(x)$ were the axial current in quantum field theory and Q the axial charge.]

The answer to this question is that Q transforms as an infinite direct sum of finite-dimensional, irreducible tensor representations of the Lorentz group plus an infinite-dimensional indecomposable (noncompletely reducible) representation of the Lorentz group. Let $T(x)$, $T^{\mu_1}(x)$, $T^{\mu_1\mu_2}(x)$, $T^{\mu_1\mu_2\mu_3}(x)$, $T^{\mu_1\mu_2\mu_3\mu_4}(x)$,$\cdots$ be a sequence of totally symmetric, traceless tensor densities. These objects transform as the irreducible representations (0,1),(0,2),(0,3),(0,4),(0,5),..., where we are using the notation of Gel'fand, Minlos, and Shapiro.[1] The infinite-dimensional representation required in the direct sum transforms as the $[(0,1)\rightarrow(1,0)]$ representation. (We will explain this notation shortly.) We will show that the charge Q transforms as the direct sum

$$[(0,1)\rightarrow(1,0)]\oplus\sum_{j=0\oplus}^{\infty}(0,2j+1). \tag{1.5}$$

Equivalently, we will find explicit formulas for the tensors $T^{\mu_1\cdots\mu_n}(x)$ and for $S(x)$ such that

$$Q=S(x)+T(x)+T^{00}(x)+T^{0000}(x)+\cdots, \tag{1.6}$$

where $S(x)$ is a density transforming as the spin-0 component of the $[(0,1)\rightarrow(1,0)]$ representation of the Lorentz group.

We find that the infinite series in (1.6) is interesting in two respects. First, this infinite series must be interpreted using a summation procedure. (Any summation procedure such as Borel summation or Euler summation is sufficient.) Second, each of the tensor densities in the series depends on space as well as time. Nevertheless, in summing this series all of the space dependence cancels, leaving an object $Q(t)$ which depends only on time.[2]

0022-2488/83/040990-06$02.50

We do not have a physical interpretation[3] for the tensors $T^{\mu_1\mu_2}, T^{\mu_1\mu_2\mu_3\mu_4},...$ or for S. We find that reimposing current conservation forces each of these tensors and S to vanish. Thus, the series truncates, leaving only the term T which becomes time independent and equal to the scalar charge.

This paper is organized as follows: In Sec. 2 we describe the notation used in this paper. Section 3 gives the solution to a simpler model problem in which we determine the transformation properties of the space integral of a scalar field $\phi(x)$. Finally, in Sec. 4 we give the solution to the main problem of this paper and explicitly establish the results of (1.5) and (1.6).

2. NOTATION

To label the irreducible representations of the Lorentz group, we use the notation of Ref. 1. Each such representation is labeled by a pair of numbers (l_0, l_1). l_0 is the lowest spin component in the representation, and if the representation is finite dimensional, $|l_1| - 1$ is the highest spin in the representation. For tensor representations, l_0 and l_1 are integers. For example, for a scalar S there is only one component and it has spin 0. Therefore, the lowest spin in this representation is 0, which gives $l_0 = 0$. The highest spin in this representation is also 0, so $l_1 = 1$.

As another example, consider a vector S^μ. This transforms as the (0,2) representation of the Lorentz group. The lowest spin component is S^0, which transforms as a rotational scalar (spin 0). The highest spin component is S^i which is a vector under the rotation group (spin 1).

In Ref. 1, it is shown that an irreducible representation consists of a sequence of spin components (representations of the rotation group) which runs by increments of one from the lowest spin to the highest spin. Each spin component occurs exactly once in this sequence.

To illustrate this, consider the case of a symmetric, traceless tensor $S^{\mu_1\cdots\mu_n}$. This transforms irreducibly as the $(0, n+1)$ representation of the Lorentz group. It contains the sequence of spin components spin 0, spin 1,...,spin n. For example, when $n = 2$ the spin-0 component is S^{00}, the spin-1 component is S^{0i}, and the spin-2 component is $S^{ij} - \delta^{ij}S^{00}/3$.

To characterize the trasformation properties of tensor densities (quantum fields which are functions of space and time) we examine the infinitesimal Lorentz transformations. A field which transforms irreducibly under the Lorentz group can be represented as a sequence of spin components $S_{l_0}^{a_1a_2\cdots a_{l_0}}(x), S_{l_0+1}^{a_1a_2\cdots a_{l_0+1}}(x),..., S_{|l_1|-1}^{a_1a_2\cdots a_{|l_1|-1}}(x)$, where $S_n^{a_1\cdots a_n}(x)$ is a totally symmetric, traceless object transforming irreducibly as the spin-n representation of the rotation group. For example, for the traceless, symmetric tensor density $T^{\mu\nu}(x)$, $S_0(x) = T^{00}(x), S_1^a(x) = T^{0a}(x), S_2^{ab}(x) = T^{ab}(x) - \frac{1}{3}\delta^{ab}T^{00}(x)$. The numbers l_0 and l_1 uniquely characterize the infinitesimal transformation laws of the field[4]

$$\frac{1}{i}\,[S_N^{a_1\cdots a_N}, J^{0k}] = (x^k\partial^0 - x^0\partial^k)S_N^{a_1\cdots a_N} + (l_1 - 1 - N)\frac{(N+1)^2 - l_0^2}{(N+1)^2}S_{N+1}^{a_1\cdots a_Nk} + \frac{l_0l_1}{iN(N+1)}\sum_{i=1}^{N}\epsilon^{a_ikq}S_N^{a_1\cdots\bar{a}_i\cdots a_Nq}$$
$$+\frac{N+l_1}{2N+1}\left\{\sum_{i=1}^{N}\delta^{a_ik}S_{N-1}^{a_1\cdots\bar{a}_i\cdots a_N} - \frac{1}{2N-1}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\delta^{a_ia_j}S_{N-1}^{a_1\cdots\bar{a}_i\cdots\bar{a}_j\cdots a_Nk}\right\}, \tag{2.1}$$

where an index with a bar over it indicates that the index is absent. This formula is crucial because given the spin components S_N, it enables us to identify the numbers l_0 and l_1.

In addition to irreducible representations of the Lorentz group, there exist indecomposable representations, that is, representations which cannot be expressed as a direct sum of irreducible representations. The simplest class of such representations are those called operator irreducible. An operator irreducible representation consists of two representations (l_0, l_1) and (l_1, l_0) "glued" together. l_0 and l_1 must both be integers. Here we assume that both are positive and that $l_0 < l_1$. Thus (l_0, l_1) is finite dimensional and (l_1, l_0) is infinite dimensional. There are two possible structures for an operator irreducible representation. In one case, the finite-dimensional part is an invariant subspace, but terms from the infinite-dimensional part mix into the finite-dimensional subspace. This representation is denoted $[(l_0,l_1)\leftarrow(l_1,l_0)]$. The other case, denoted $[(l_0,l_1)\rightarrow(l_1,l_0)]$, occurs when the infinite-dimensional part forms an invariant subspace, and terms from the finite-dimensional part mix into terms of the infinite-dimensional part. An operator irreducible representation of the second type occurs in the decomposition of Q in Sec. 4. A more detailed discussion of indecomposible representations is found in Ref. 5.

3. A MODEL PROBLEM

In this section we solve a problem which is slightly simpler than determining the transformation properties of charges from nonconserved currents. Instead, we examine the transformation properties of the "charge" $P(t)$ obtained by integrating a scalar field $\phi(x)$:

$$P(t) = \int d_3x\phi(\mathbf{x}, t). \tag{3.1}$$

The general technique we will follow is outlined as an illustrative example in a paper by Bender and Griffiths.[5] The idea is that $P(t)$ is a rotational scalar. If we commute it with the generator J^{0k}, the result after removing the orbital part $(x^k\partial^0 - x^0\partial^k)P(t)$ will be a spin-1 object P^k. Note that even though $P(t)$ in (3.1) is not a function of the space coordinate x^k, the angular momentum generator J^{0k} introduces a space dependence in P^k. Commuting P^k with the generator J^{0l} and again removing the orbital part $(x^l\partial^0 - x^0\partial^l)P^k$ gives a spin-2 object P^{kl} and a new spin-0 object, P_1. If P_1 were a multiple α of P, then there would be only one spin-0 object generated by repeated infinitesimal Lorentz transformations. This would allow us to argue that P is the spin-0 component of an irreducible representation of the Lorentz group. Moreover, from the number α, we could compute l_1. (For this case we know that $l_0 = 0$.) In fact, P_1 is not a multiple of P. Rather, it is a new spin-0 object. Commuting P_1 twice with the generator of Lorentz transformations produces yet another spin-object P_2, which is linearly independent of P and P_1. Repeating this process produces an infinite sequence of linearly independent spin-0 objects $P, P_1 P_2, \cdots$.

The analysis now consists of two parts. First, we conduct an infinite linear combination of the P's using an arbitrary set of undetermined coefficients β_n

$$A = P + \sum_{n=1}^{\infty} \beta_n P_n. \tag{3.2}$$

We determine the coefficients β_n by demanding that two successive commutations of A with the generator of Lorentz transformations give a spin-0 component which is a numerical multiple α of A. In effect, α plays the role of an eigenvalue for a Schrödinger-like difference equation. The corresponding eigenvector is the set of coefficients β_n. We find that there are in fact an infinite number of solutions to this eigenvalue problem, each solution for A being the spin-0 component of some irreducible representation of the Lorentz group. We label each solution A by the corresponding value of l_1 determined from α.

Second, we select from the set of A's an infinite set A_m which in linear combination reproduce the original quantity P. That is, we determine a set of coefficients γ_m such that

$$P = \sum_{m=1}^{\infty} \gamma_m A_m. \tag{3.3}$$

This represents a formal solution to the problem. It expresses the transformation properties of P as a direct sum of irreducible represenations of the Lorentz group. We believe that this decomposition is unique.

Griffiths[6] has suggested to us that to facilitate the formal procedure just described it is useful to introduce a multiply-indexed Kronecker δ function $D^{2n}_{a_1 a_2 \cdots a_{2n}}$ which is defined by

$$D^{2n}_{a_1 \cdots a_{2n}} = \frac{1}{2n!} \sum_{\substack{\text{all permutations} \\ \text{of } a_1 a_2 \cdots a_{2n}}} \delta_{a_1 a_2} \delta_{a_3 a_4} \cdots \delta_{a_{2n-1} a_{2n}}. \tag{3.4}$$

This object has a number of useful properties observed by Griffiths. Two that are useful for our calculation are

$$D^{n+1}_{a_1 \cdots a_n k} = \frac{1}{n} \sum_{i=1}^{n} \delta_{k a_i} D^{n-1}_{a_1 \cdots \bar{a}_i \cdots a_n} \tag{3.5}$$

and

$$\delta_{ik} D^{n+2}_{a_1 \cdots a_n ik} = \frac{n+3}{n+1} D^{n}_{a_1 \cdots a_n}. \tag{3.6}$$

We have found that the possible spin-0 terms are most simply expressed in terms of a quantity $R^{n}_{a_1 \cdots a_n}$ which is defined by

$$R^{n}_{a_1 \cdots a_n}(\mathbf{x}, t) = (\partial^0)^n \sum_{p=0}^{n} (-1)^p \binom{n}{p} x_{a_1} \cdots x_{a_p} \times \int d_3 y y_{a_{p+1}} \cdots y_{a_n} \phi(\mathbf{y}, t). \tag{3.7}$$

(Note that for $n = 0$, this is the quantity P that we are investigating.) The most general spin-0 term is then

$$M(\mathbf{x}, t) = \sum_{n=0}^{\infty} \beta_n D^{2n}_{a_1 \cdots a_{2n}} R^{2n}_{a_1 \cdots a_{2n}}(\mathbf{x}, t), \tag{3.8}$$

where β_n are arbitrary coefficients.

Commuting this with J^{0k} and removing the orbital part gives, after much combinatorics, integration by parts, and use of identity (3.5), the resulting spin-1 object which we call $M^k(\mathbf{x}, t)$

$$M^k(\mathbf{x}, t) = \sum_{n=0}^{\infty} [\beta_n - (2n+2)^2 \beta_{n+1}] \times D^{2n+2}_{a_1 \cdots a_{2n+1} k} R^{2n+1}_{a_1 \cdots a_{2n+1}}. \tag{3.9}$$

[M^k is the analog of P^k in the discussion following (3.1).] Commuting again with J^{0l} and removing the orbital piece gives a spin-2 and a spin-0 object. The spin-0 object is

$$\frac{1}{3} \delta^{kl} \sum_{n=0}^{\infty} [(2n+1)(2n+2)^2(2n+3)\beta_{n+1} - (8n^2 + 8n + 3)\beta_n + \beta_{n-1}] D^{2n}_{a_1 \cdots a_{2n}} R^{2n}_{a_1 \cdots a_{2n}}, \tag{3.10}$$

where $\beta_{-1} = 0$.

Demanding that the spin-0 object in (3.10) be a numerical multiple α of the original spin-0 expression in (3.8) gives a difference equation satisfied by the coefficients β_n

$$\frac{1}{3} [(2n+1)(2n+2)^2(2n+3)\beta_{n+1} - (8n^2 + 8n + 3)\beta_n + \beta_{n-1}] = \alpha \beta_n. \tag{3.11}$$

This is a second-order eigenvalue equation in which α plays the role of the eigenvalue. However, unlike conventional eigenvalue problems, there is no boundary condition at $n = \infty$ which excludes a growing solution and therby determines a set of eigenvalues. This is because the possible asymptotic behaviors of β_n for large n are

$$\beta_n \sim \frac{C_1 + C_2 \log n}{(2n+1)!} \quad (n \to \infty), \tag{3.12}$$

C_1 and C_2 are arbitrary constants. There is no justification for excluding any solutions because the sum in (3.8) converges for all solutions. Thus, the eigenspectrum is continuous.

If β_n satisfies the difference equation in (3.11), then M in (3.8) is the spin-0 part of an *irreducible* representation. Using (2.1) we find that this representation is labeled by $(0, l_1)$, where $l_1 = \sqrt{(3\alpha + 1)}$.

Next we solve (3.11) in terms of a generating function $f(z)$. We define

$$f(z) = \sum_{j=0}^{\infty} (2j)! \beta_j z^j. \tag{3.13}$$

The asymptotic behavior in (3.12) implies that this series converges for all $|z| < 1$. Thus $f(z)$ is analytic in the unit circle in the complex-z plane. We see that $f(z)$ satisfies the differential equation

$$f''(z)+\frac{5z-3}{2z(z-1)}f'(z)+\frac{2(z-1)-l_1^2}{4z(z-1)^2}f(z)=0. \tag{3.14}$$

It is not hard to recognize that this differential equation has 3 regular singular points and is therefore a transformed hypergeometric equation.[7] The general solution to (3.14) is a linear combination of

$$(z-1)^{l_1/2}F\left(1+\frac{l_1}{2},\frac{1}{2}+\frac{l_1}{2};\frac{3}{2};z\right)$$

and

$$z^{-1/2}(z-1)^{l_1/2}F\left(\frac{l_1}{2},\frac{1}{2}+\frac{l_1}{2};\frac{1}{2};z\right).$$

Choosing the solution which is analytic at the origin and recognizing that the hypergeometric function can be written in terms of elementary functions gives the following expression for $f(z)$:

$$f(z)=\begin{cases}\dfrac{1}{\sqrt{z}}\log\dfrac{1+\sqrt{z}}{1-\sqrt{z}} & (l_1=0),\\[2ex] \dfrac{1}{\sqrt{z}}\left[\left(\dfrac{1-\sqrt{z}}{1+\sqrt{z}}\right)^{l_1/2}-\left(\dfrac{1+\sqrt{z}}{1-\sqrt{z}}\right)^{l_1/2}\right] & (l_1\neq 0),\end{cases} \tag{3.15}$$

where we neglect an overall multiplicative constant. Note that there is a special class of positive l_1's for which $f(z)$ in (3.15) is single valued; namely, l_1 an even integer $\geqslant 2$. $M(\mathbf{x},t)$ is the spin-0 component of a finite-dimensional representation whenever l_1 is a nonzero integer.

The final step is to find a set of l_1's such that summing over the irreducible representations $M(\mathbf{x},t)$ with an appropriate set of coefficients γ reproduces the original expression $P(t)$ in (3.1):

$$\sum_{l_1}\gamma_{l_1}M_{l_1}(\mathbf{x},t)=P(t). \tag{3.16}$$

Remembering that $M_{l_1}(\mathbf{x},t)$ also contains an infinite sum, we interchange orders of summation and find that (3.16) is true if and only if the following identity holds for all z:

$$\gamma_0\frac{1}{\sqrt{z}}\log\frac{1+\sqrt{z}}{1-\sqrt{z}}+\sum_{l_1\neq 0}\gamma_{l_1}\frac{1}{\sqrt{z}}\times\left[\left(\frac{1-\sqrt{z}}{1+\sqrt{z}}\right)^{l_1/2}-\left(\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)^{l_1/2}\right]=1. \tag{3.17}$$

A set of γ for which this identity holds is

$$\gamma_{l_1}=\begin{cases}(-1)^{l_1/2} & l_1 \text{ an even integer } \geqslant 2,\\ 0 & \text{otherwise.}\end{cases} \tag{3.18}$$

However, for this choice the sum in (3.17) must be performed using a summation procedure such as Euler or Borel summation.[8]

To conclude, we have shown algebraically that

$$P(t)=-M_2+M_4-M_6+\cdots \tag{3.19}$$

and group theoretically that $P(t)$ transforms as the 0-spin component of the direct sum of an infinite number of finite-dimensional, irreducible, tensor representations of the Lorentz group:

$$(0,2)\oplus(0,4)\oplus(0,6)\oplus\cdots. \tag{3.20}$$

It is interesting that while each of the M's in (3.19) depends on space and time, the identity in (3.17) implies that all space dependence in the right side of (3.19) cancels. Equation (3.20) implies that $P(t)$ is the sum of the spin-0 components of totally symmetric, traceless, odd-rank tensor densities

$$P(t)=T^{0}(\mathbf{x},t)+T^{000}(\mathbf{x},t)+T^{00000}(\mathbf{x},t)+\cdots. \tag{3.21}$$

4. TRANSFORMATION PROPERTIES OF $Q(t)=\int d_3xj^0(\mathbf{x},t)$

To determine the transformation properties of

$$Q(t)=\int d_3xj^0(\mathbf{x},t), \tag{4.1}$$

where $j^0(x)$ is the zeroth component of a current which is not locally conserved, we follow exactly the same procedure as in the last section. We begin by constructing a class of quantities $U^{n}_{a_1\cdots a_n}$ analogous to $R^{n}_{a_1\cdots a_n}$ in (3.7)

$$U^{n}_{a_1\cdots a_n}(\mathbf{x},t)=(\partial^0)^{n-1}\sum_{p=0}^{n}(-1)^{p+1}\binom{n}{p}x_{a_1\cdots}x_{a_p}\times\int d_3y y_{a_{p+1}}\cdots y_{a_n}(\partial_\mu j^\mu)(\mathbf{y},t). \tag{4.2}$$

Note that U^0 is just $Q(t)$ in (4.1). Next, we construct a set of quantities $N(\mathbf{x},t)$ which are analogous to $M(\mathbf{x},t)$ in (3.8)

$$N(\mathbf{x},t)=\sum_{n=0}^{\infty}\epsilon_n D^{2n}_{a_1\cdots a_{2n}}U^{2n}_{a_1\cdots a_{2n}}(\mathbf{x},t), \tag{4.3}$$

where the coefficients ϵ_n are analogous to the coefficients β_n.

Now we impose the condition that $N(\mathbf{x},t)$ be the spin-0 component of an irreducible representation of the Lorentz group. We commute $N(\mathbf{x},t)$ with J^{0k}, the generator of infinitesimal Lorentz transformations, remove the orbital part, and repeat the process. We then demand that the new spin-0 object that arises be a numerical multiple α of the original spin-0 quantity $N(\mathbf{x},t)$. This gives a difference equation satisfied by the coefficients ϵ_n which is analogous to the difference equation in (3.11):

$$\tfrac{1}{3}[(2n+3)(2n+2)(2n+1)2n\epsilon_{n+1}-(8n^2+4n)\epsilon_n+\epsilon_{n-1}]=\alpha\epsilon_n, \tag{4.4}$$

where $\epsilon_{-1}=0$. The possible asymptotic behaviors of ϵ_n for large n are

$$\epsilon_n\sim\frac{C_1+C_2n}{(2n+1)!}\quad(n\rightarrow\infty), \tag{4.5}$$

where C_1 and C_2 are arbitrary constants. Again we observe that there is no constraint on the possible choices for α. As before, we conclude from (2.1) that $N(\mathbf{x},t)$ is the spin-0 component $(0,l_1)$ where $l_1=\sqrt{3\alpha+1}$.

To solve (4.4), we introduce a generating function $g(z)$ defined by:

$$g(z)=\sum_{j=0}^{\infty}(2j+1)!\epsilon_j z^j. \tag{4.6}$$

We find that $g(z)$ satisfies the differential equation

$$g''(z) + \frac{1}{2}\frac{7z+1}{z(z-1)}g'(z) + \frac{1}{4}\frac{6z^2+(1-l_1^2)z+2}{z^2(z-1)^2}g(z) = 0. \tag{4.7}$$

For the special case $l_1^2 = 1$ ($\alpha = 0$) there is an additional solution $g(z)$ satisfying the algebraic equation

$$(1-z)^2 g(z) = g(0). \tag{4.8}$$

Equation (4.7) can be transformed into a hypergeometric differential equation. The general solution to (4.7) is a linear combination of

$$z(z-1)^{-3/2+l_1/2}F\left(1+\frac{l_1}{2},\frac{1}{2}+\frac{l_1}{2};\frac{3}{2};z\right)$$

and

$$z^{1/2}(z-1)^{-3/2+l_1/2}F\left(\frac{l_1}{2},\frac{1}{2}+\frac{l_1}{2};\frac{1}{2};z\right).$$

Because of the asymptotic behavior in (4.5) we know that we want the solution for $g(z)$ that is analytic at the origin. Neglecting an overall multiplicative constant we find that

$$g(z) = \begin{cases} \sqrt{z}(1-z)^{-3/2}\log\dfrac{1+\sqrt{z}}{1-\sqrt{z}} & (l_1 = 0), \\ \dfrac{1}{(1-z)^2} & (l_1 = 1), \\ \sqrt{z}(1-z)^{-3/2}\left[\left(\dfrac{1-\sqrt{z}}{1+\sqrt{z}}\right)^{l_1/2} - \left(\dfrac{1+\sqrt{z}}{1-\sqrt{z}}\right)^{l_1/2}\right] & (l_1 \neq 0). \end{cases} \tag{4.9}$$

Note that for $l_1 = 1$ there are two linearly independent solutions. In this case, the general solution for $g(z)$ is

$$g(z) = \frac{C_1 + C_2 z}{(1-z)^2},$$

where C_1 and C_2 are arbitrary constants. However, only one linear combination leads to a solution for $N(\mathbf{x},t)$ that is the spin component of the (0,1) representation. A second linearly independent solution for $g(z)$ gives rise to the [(0,1)→(1,0)] indecomposable representation of the Lorentz group. We will label the spin-0 component of the $(0,l_1)$ representation by $l_1{:}N_{l_1}(\mathbf{x},t)$. We let $N_1(\mathbf{x},t)$ be the spin-0 component of the irreducible representation (0,1). The coefficients ϵ_n for $N_1(\mathbf{x},t)$ in the sum in (4.3) are generated by

$$g(z) = \frac{1+z}{(1-z)^2}.$$

Let $N_1'(\mathbf{x},t)$ denote the spin-0 component of the [(0,1)→(1,0)] representation. For $N_1'(\mathbf{x},t)$, the coefficients ϵ_n are generated by

$$g(z) = \frac{z}{(1-z)^2}.$$

As in Sec. 3, there is a special set of positive values of l_1 for which $g(z)$ is single valued; namely l_1 an odd integer $\geqslant 1$. Except for the indecomposable representation just discussed $N(\mathbf{x},t)$ is the spin-0 component of a finite-dimensional, irreducible representation whenever l_1 is a nonzero integer.

Finally, we must find a set of l_1's so that when we sum over the irreducible representations $N(\mathbf{x},t)$ multiplied by an appropriate set of coefficients λ_{l_1} we obtain $Q(t)$ in (4.1):

$$\lambda_1' N_1'(\mathbf{x},t) + \sum_{l_1}\lambda_{l_1}N_{l_1}(\mathbf{x},t) = Q(t). \tag{4.10}$$

This is the analog of equation (3.16).

Interchanging the order of sums in (4.10) and in $N(\mathbf{x},t)$ we obtain the identity that ensures the validity of (4.10):

$$\lambda_0\sqrt{z}(1-z)^{-3/2}\log\frac{1+\sqrt{z}}{1-\sqrt{z}} + \lambda_1\frac{1+z}{(1-z)^2} + \lambda_1'\frac{z}{(1-z)^2} + \sum_{l_1\neq 0,1}\lambda_{l_1}\sqrt{z}(1-z)^{-3/2} \times\left[\left(\frac{1-\sqrt{z}}{1+\sqrt{z}}\right)^{l_1/2} - \left(\frac{1+\sqrt{z}}{1-\sqrt{z}}\right)^{l_1/2}\right] = 1. \tag{4.11}$$

A set of λ which satisfies this identity is:

$$\lambda_{l_1} = \begin{cases} 1 & (l_1 = 1), \\ (l_1 - \frac{1}{2})(-1)^{(l_1+1)/2} & (l_1 \text{ an odd integer } \geqslant 3), \\ 0 & \text{otherwise,} \end{cases} \tag{4.12a}$$

and

$$\lambda_1' = -1. \tag{4.12b}$$

As in Sec. 3, the sum in (4.11) must be performed using a summation procedure such as Euler or Borel.

We have thus shown algebraically that

$$Q(t) = N_1(\mathbf{x},t) - N_1'(\mathbf{x},t) + \sum_{n=1}^{\infty}(2n+\tfrac{1}{2})(-1)^{n+1}N_{2n+1}(\mathbf{x},t). \tag{4.13}$$

We also see that $Q(t)$ transforms as the 0-spin component of an infinite direct sum of finite-dimensional, irreducible, tensor representations of the Lorentz group plus the 0-spin component of an infinite-dimensional indecomposable representation:

$$[(0,1)\rightarrow(1,0)] \oplus (0,1) \oplus (0,3) \oplus (0,5) \oplus \cdots. \tag{4.14}$$

Another way to state this result is that $Q(t)$ is the sum of the spin-0 components of totally symmetric, traceless, even-rank tensor densities plus a density $S(\mathbf{x},t)$ that transforms as the spin-0 component of the $[(0,1)\rightarrow(1,0)]$ representation of the Lorentz group [in fact, $S(\mathbf{x},t) = -N_1'(\mathbf{x},t)$]

$$Q(t) = S(x,t) + T(x,t) + T^{00}(x,t) + T^{0000}(x,t) + \cdots. \tag{4.15}$$

Note that this representation of $Q(t)$ reduces to a time-independent scalar in the case of a conserved current. If $\partial_\mu j^\mu(x) = 0$, then

$$U^n_{a_1 \cdots a_n} = 0, \quad (n \geqslant 1). \tag{4.16}$$

Since for $l_1 \neq 1$, the generating function $g(z)$ is 0 at $z = 0$, it follows that ϵ_0, the coefficient of U^0 in $N_{l_1}(\mathbf{x},t)$, vanishes. Thus $\partial_\mu j^\mu(x) = 0$ implies that

$$N_{l_1}(\mathbf{x},t) = 0 \quad (l_1 \neq 1). \tag{4.17a}$$

Similarly,

$$N_1'(\mathbf{x},t) = 0 \tag{4.17b}$$

and

$$N_1(\mathbf{x},t) = U^0(\mathbf{x},t). \tag{4.17c}$$

Thus, the decomposition in (4.13) reduces to the single term $N_1(\mathbf{x},t)$, which is a Lorentz scalar, in the case of a conserved current.

ACKNOWLEDGMENTS

We are greatly indebted to D. J. Griffiths for providing us with helpful ideas and inspiration. We also thank the U.S. Department of Energy for partial financial support.

---

[1] I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and Their Applications* (MacMillan, New York, 1963).

[2] The problem of time-dependent charges has been investigated extensively [see e.g. S. Coleman, J. Math. Phys.7, 787 (1966), E. Gal-Ezer and H. Reeh, Commun. Math. Phys. **43**, 137 (1975); J. Bros. D. Buchholz and V. Glaser, Commun. Math. Phys. **50**, 11 (1976)]. It is known that defining such charges in quantum field theories presents a serious problem because they commute with translations. In this paper we merely assume that where necessary all space integrals of fields converge because the fields vanish rapidly in the spacelike direction. We hope that this paper may possibly give a hint as to how in quantum field theory to construct rigorously a time-dependent charge.

[3] It is striking that a nonconserved charge implies the existence of an infinite sequence of quantum fields whose spins are spaced by two. Such a sequence is reminiscent of a Regge trajectory.

[4] C. M. Bender and D. J. Griffiths, J. Math. Phys. **12**, 2151 (1971).

[5] C. M. Bender and D. J. Griffiths, Phys. Rev. D **2**, 317 (1970).

[6] D. J. Griffiths, private communications.

[7] See A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 1, Chap. II.

[8] A description of these summation procedures can be found in C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw–Hill, New York, 1978) or E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed. (Cambridge U.P., Cambridge, 1927). It is rigorously true that (3.17) is Euler and Borel summable. However, while we believe that the coefficients γ in (3.18) are unique we have not proved it.

Friedmann universe and superluminal Dirac particles

Sushil K. Srivastava
Government College, Port Blair 744104, India

(Received 29 September 1981; accepted for publication 8 January 1982)

Superluminal Dirac particles are considered in a Friedmann universe. It is concluded that if such particles were produced at the epoch of the big bang (as assumed by Narlikar and Sudarshan) and also that they survived up to the present time, their metamass should be less than 8.77×10^{-54} g. It is also concluded that the wavelength of these primordial tachyons (superluminal particles surviving to present time) would be greater than 4.56×10^{-23} cm. In this paper, expressions for various physical parameters such as probability density and energy and rate of dissipation of energy for these particles are derived.

PACS numbers: 14.80.Pb

1. INTRODUCTION

In recent years, there has been continuing interest in research on tachyons (superluminal particles) in the background of general relativity and relativistic cosmology. In 1976, Narlikar and Sudharshan[1] published a paper in which they assumed that such particles would have been produced at or just after the epoch of the big bang along with other fundamental particles of ordinary matter. In their paper, they have discussed many features of primordial tachyons taking the Friedmann model of the universe, having zero space curvature. Recently this author[2] also has done a similar problem in the background of Robertson–Walker cosmology taking the model of positive curvature. In these papers spinless tachyons, following the Klein–Gordon equation, have been considered.

In the present paper, we have considered spin-$\frac{1}{2}$ tachyons obeying the Dirac equation and have discussed various features of such particles in the background of Robertson–Walker cosmology taking the model of positive, negative as well as zero curvature. We have also analyzed the characteristics of primordial spin-$\frac{1}{2}$ tachyons in the models containing different perfect fluids such as dust, radiation, super-dense matter and nonrelativistic matter.

We consider the Robertson–Walker line element

$$ds^2 = dt^2 - [R^2(t)/(1+kr^2/4)^2](dx^2+dy^2+dz^2)\,, \tag{1.1}$$

where $R(t)$ is the scale factor, k is the space curvature having values $+1$, -1, adn 0 and $r^2 = x^2+y^2+z^2$.

Under coordinate transformations

$$T=\int_0^t \frac{dt}{R(t)}\,,\quad X=\int_0^x \frac{dx}{1+kr^2/4}\,,$$
$$Y=\int_0^y \frac{dy}{1+kr^2/4}\,,\quad Z=\int_0^z \frac{dz}{1+kr^2/4}\,, \tag{1.2}$$

the line element (1.1) is written as

$$ds^2 = S^2(T)(dT^2 - dX^2 - dY^2 - dZ^2)\,, \tag{1.3}$$

where $S(T) = R(t)$.

In Sec. 2, we describe the Dirac theory. In Sec. 3, we solve the Dirac equation for a tachyon using the method of Wentzel–Kramers–Brioullin (hereafter called WKB) approximation. Also we have tested the validity of WKB solutions. We get two solutions. It is found that one solution is valid before the age of one second of this universe. The other solution is valid for time beyond one second.

In Sec. 4, we discuss the probability density of the primordial spin-$\frac{1}{2}$ tachyon. It is found that the probability density decreases with time in both cases when $t<1$ second as well as $t>1$ second. We also find that probability density also decreases if the metamass of the tachyon increases.

In Sec. 5, we have derived expressions for energy and rate of dissipation of energy of the tachyon in the medium of different perfect fluids. It is found that dissipation of energy of a tachyon is fastest in the medium of super dense matter and slowest in the medium of dust.

In the last section, we show that if a primordial spin-$\frac{1}{2}$ tachyon survives up to the present epoch, its metamass should be less than 8.77×10^{-54} g and its wavelength should be greater than 4.56×10^{-23} cm.

2. DIRAC THEORY

We choose an orthonormal tetrad field $e_a^\alpha(x)$ such that

$$e_a^\alpha e_b^\beta g_{\alpha\beta} = \eta_{ab}\,, \tag{2.1}$$

where $g_{\alpha\beta}$ is the metric tensor given by the line element (1.3) and $\eta_{ab} = \mathrm{diag}(-1,-1,-1,+1)$. Indices $\alpha, \beta, \cdots$ run from 1 to 4 and $a, b\cdots = 1\cdots 4$.

The Dirac equation for a primordial spin-$\frac{1}{2}$ tachyon in curved space–time takes the form

$$i\hbar\gamma^\mu\psi_{;\mu} - m\psi = 0\,, \tag{2.2}$$

where

$$m = i\Omega \quad \text{(Ref. 3)} \tag{2.3}$$

$$\psi_{;\mu} = \psi_{,\mu} + \Gamma_\mu\,;\quad \Gamma_\mu = \tfrac{1}{4}\, e^\alpha_{a;\mu} e_{ab}\gamma^b\gamma^a\,,$$
$$\gamma^\mu = e_a^\mu\gamma^a,\quad \gamma^{(a}\gamma^{b)} = \eta^{ab},\quad \gamma^{(\mu}\gamma^{\nu)} = g^{\mu\nu}\,. \tag{2.4}$$

Standard γ^a-matrices[4] are given as

$$\gamma^{\hat a} = \begin{bmatrix} 0 & \sigma^{\hat a} \\ \sigma^{\hat a} & 0 \end{bmatrix},\quad \gamma^{(4)} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{2.5}$$

Indices $\hat a, \hat b\cdots = 1,2,3$. Semicolon (;) denotes covariant derivative and comma (,) denotes partial derivative.

The Dirac current j^α is defined by means of $\bar\psi = \psi^\dagger\gamma^{(4)}$ as

$$j^\alpha = \bar\psi\gamma^\alpha\psi = \bar\psi e_a^\alpha\gamma^a\psi\,; \tag{2.6}$$

0022-2488/83/040996-05$02.50

j^α is divergence free, hence

$$j^\alpha_{;\alpha} = 0 \,.$$

This enables us a hypersurface-independent normalization on a hypersurface σ with timelike normal vector u^α using the integral (where $d^{\,3}V$ is the invariant volume element of σ, $d\sigma^2 = dX^2 + dY^2 + dZ^2$)

$$\int_\sigma j^\alpha u_\alpha d^{\,3}V \,.$$

3. SOLUTIONS OF DIRAC EQUATION

Here we introduce a procedure characterized as a conformal method[5] such that the metric tensor $g_{\alpha\beta}$ in (1.3) is related to the Minkowski metric $\tilde{g}_{\alpha\beta}$ as

$$g_{\alpha\beta} = S^2 \tilde{g}_{\alpha\beta} \tag{3.1}$$

and correspondingly a tetrad field

$$\tilde{e}^\alpha_a = S e^\alpha_a \,. \tag{3.2}$$

For a given congruence of world lines of particles in the underlying manifold describing a particle, the corresponding currents are related according to

$$\tilde{j}^\alpha = S^4 j^\alpha \,. \tag{3.3}$$

From Eqs. (2.6), (3.1), and (3.2), Dirac fields in the two space–times are connected by

$$\phi = S^{3/2} \psi, \quad \phi^\dagger = S^{3/2} \psi^\dagger \,, \tag{3.4}$$

where ψ is the solution of the Dirac equation (2.2), $\psi^\dagger$ and $\phi^\dagger$ are Hermitian conjugates of ψ and ϕ, respectively. Adjustment of tetrads along the coordinate lines yields

$$e^\alpha_a = (1/S)\delta^\alpha_a; \quad \tilde{e}^\alpha_a = \delta^\alpha_a \,. \tag{3.5}$$

Hence the Dirac equation (2.2) takes the form

$$i\hbar \delta^\alpha_a \gamma^a \tilde{\psi}_{,\alpha} + \tfrac{3}{2} i\hbar (\dot{S}/S) \gamma^4 \tilde{\psi} - Sm\tilde{\psi} = 0 \,. \tag{3.6}$$

The corresponding equation for ϕ is, because of Eq. (3.4),

$$i\hbar \delta^\alpha_a \gamma^a \phi_{,\alpha} - Sm\phi = 0 \,. \tag{3.7}$$

This equation can be written as

$$(i\hbar \gamma^a \partial_a - Sm)\phi = 0 \,, \tag{3.8}$$

$$\partial_a \equiv \{\partial_X, \partial_Y, \partial_Z, \partial_T\} \,,$$

$$\partial_X \equiv \frac{\partial}{\partial X} \,, \quad \text{etc.}$$

Now applying the operator $(-i\hbar\gamma^a \partial_a - Sm)$ on Eq. (3.8) we have the squared equation

$$[-\hbar^2 \eta^{ab} \partial_a \partial_b + i\hbar^2 \gamma^{(4)} \dot{S} m - S^2 m^2]\phi = 0 \tag{3.9}$$

(dot over the variable denotes d/dT).

We solve Eq. (3.9) with the ansatz

$$\phi = f(T) \Gamma \exp(-ik_{\hat{a}} x^{\hat{a}}) \,, \tag{3.10}$$

where $x^{\hat{a}} = (X, Y, Z)$ and the constant spinor Γ is defined by

$$\gamma^{(4)} \Gamma = \epsilon \Gamma, \quad \epsilon = \pm 1 \,. \tag{3.11}$$

For the standard representation of the γ^a, Γ becomes[3]

$$\Gamma^{\epsilon=1}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Gamma^{\epsilon=1}_{-1} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

$$\Gamma^{\epsilon=-1}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \Gamma^{\epsilon=-1}_{-1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \tag{3.12}$$

and Eq. (3.9) reduces to a differential equation for $f(T)$ as

$$\frac{d^{\,2}f}{dT^2} + \left(\frac{S^2 m^2}{\hbar^2} - i\epsilon m \dot{S} + K^2\right) f = 0 \,, \tag{3.13}$$

where $k^2 = k_1^2 + k_2^2 + k_3^2$. (Here k does not denote space curvature). But from Eq. (2.3) $m^2 = -\Omega^2$, hence we have

$$\frac{d^{\,2}f}{dT^2} + \left(-\frac{S^2\Omega^2}{\hbar^2} + \epsilon\Omega\dot{S} + k^2\right) f = 0 \,. \tag{3.14}$$

This differential can be written in one of the two forms

$$\frac{d^{\,2}f}{dT^2} + \eta^2 f = 0 \quad \text{for } \eta^2 > 0 \,, \tag{3.15}$$

where

$$\eta^2 = -S^2\Omega^2/\hbar^2 + \epsilon\Omega\dot{S} + K^2 \tag{3.16}$$

and

$$\frac{d^{\,2}f}{dT^2} - \eta'^2 f = 0 \quad \text{for } \eta'^2 > 0 \,, \tag{3.17}$$

where

$$\eta'^2 = S^2\Omega^2/\hbar^2 - \epsilon\Omega\dot{S} - K^2 \,. \tag{3.18}$$

We obtain solutions of differential equations (3.15) and (3.18) by the method of WKB approximation as[6]

$$f(T) = A\eta^{-1/2} \exp\left(\pm i \int_0^T \eta \, dT\right) \quad \text{for } \eta^2 > 0 \tag{3.19}$$

and

$$f(T) = B\eta'^{-1/2} \exp\left(\pm i \int_0^T \eta' \, dT\right) \quad \text{for } \eta'^2 > 0 \,, \tag{3.20}$$

respectively.

From Einstein's field equations[7] with a suitable equation of state $p = (\gamma - 1)\rho$ (where p is the pressure, γ is the ratio of specific heats at constant pressure and constant volume, and ρ is the energy density) we have $S(t) = t^q$, where possible values of q are $\frac{1}{3}$, $\frac{2}{5}$, $\frac{1}{2}$, and $\frac{2}{3}$ for superdense matter, nonrelativistic matter, radiation and dust, respectively.

Now we shall test the validity of WKB solutions obtained above. For this test we shall see whether $|(d\eta/dT)/2\eta^2|$ or $|(d\eta'/dT)/2\eta'^2|$ is less than unity.

From Eq. (3.16) we find that

$$\left|\frac{d\eta/dT}{2\eta^2}\right| = \left|\frac{-(\Omega^2/\hbar^2)qt^{3q-1} + \frac{1}{2}\epsilon\Omega q(2q-1)t^{3q-2}}{2[-(\Omega^2/\hbar^2)t^{2q} + \epsilon q\Omega t^{2q-1} + k^2]^{3/2}}\right| \,. \tag{3.21}$$

Equation (3.21) shows that $|(d\eta/dT)/2\eta^2| < 1$ up to $t < 1$; hence, solution (3.19) is valid before one second.

From Eq. (3.18) we find that

$$\left|\frac{d\eta'/dT}{2\eta'^2}\right| = \left|\frac{(\Omega^2/\hbar^2)qt^{3q-1} - \frac{1}{2}\epsilon\Omega q(2q-1)t^{3q-2}}{2[(\Omega^2/\hbar^2)t^{2q} - \epsilon\Omega q t^{2q-1} - k^2]^{3/2}}\right| \,. \tag{3.22}$$

Equation (3.22) shows that $|(d\eta'/dT)/2\eta'^2| < 1$ when $t > 1$. It means that beyond one second, solution (3.20) is valid. Connecting Eqs. (3.4), (3.10), and (3.11) we have the solution of the Dirac equation (2.2) as

$$\psi = A\Gamma\eta^{-1/2}S^{-3/2}\exp\left(-ik_{\hat a}x^{\hat a} \pm i\int_0^T \eta\, dT\right) \quad (3.23)$$

up to t less than one second.

Similarly connecting Eqs. (3.4), (3.10), and (3.20), we have the solution of the Dirac equation (2.2) as

$$\psi = B\Gamma\eta'^{-1/2}S^{-3/2}\exp\left(-ik_{\hat a}x^{\hat a} \pm \int_0^T \eta'\, dT\right) \quad (3.24)$$

beyond t equal to one second.

Both these solutions show damping of the tachyon wave function with the passage of time.

4. PROBABILITY DENSITY OF PRIMORDIAL SPIN-$\frac{1}{2}$ TACHYON

The probability density of spin-$\frac{1}{2}$ tachyon is defined as

$$P = \psi^\dagger\psi\,. \quad (4.1)$$

From Eq. (3.23) we have

$$P_1 = \frac{A^2}{t^{3q}[-(\Omega^2/\hbar^2)t^{2q} + \epsilon\Omega q t^{2q-1} + k^2]^{1/2}} \quad (4.2)$$

when $t < 1$ second.

From Eq. (3.24) we have

$$P_2 = \frac{B^2}{t^{3q}[(\Omega^2/\hbar^2)t^{2q} - \epsilon\Omega q t^{2q-1} - k^2]^{1/2}} \quad (4.3)$$

when $t > 1$ second.

Both expressions (4.2) and (4.3) for probability density show that the probability density of a spin-$\frac{1}{2}$ tachyon decreases with time. From Eq. (4.2) we find that decay of the probability density of an antitachyon is faster than for a tachyon provided that for any value of t before one second

$$k^2 > \frac{\Omega^2}{\hbar^2}t^{2q} + \Omega q t^{2q-1}\,. \quad (4.4)$$

Here $\epsilon = 1$ corresponds to a tachyon whereas $\epsilon = -1$ corresponds to an antitachyon. Now we discuss decay of probability density beyond one second taking different values of q for different perfect fluids, because it is difficult to decide what perfect fluid was present up to the time when this universe was one second old.

$$\text{If } q = \tfrac{1}{3}, \quad P_2 = \frac{B^2}{[(\Omega^2/\hbar^2)t^{8/3} - \frac{1}{3}\epsilon\Omega t^{5/3} - k^2t^4]^{1/2}}\,. \quad (4.5)$$

$$\text{If } q = \tfrac{2}{5}, \quad P_2 = \frac{B^2}{[(\Omega^2/\hbar^2)t^{16/5} - \frac{2}{5}\epsilon\Omega t^{11/5} - k^2t^{12/5}]^{1/2}}\,. \quad (4.6)$$

$$\text{If } q = \tfrac{1}{2}, \quad P_2 = \frac{B^2}{[(\Omega^2/\hbar^2)t^{4} - \frac{1}{2}\epsilon\Omega t^{3} - k^2t^3]^{1/2}}\,. \quad (4.7)$$

$$\text{If } q = \tfrac{2}{3}, \quad P_2 = \frac{B^2}{[(\Omega^2/\hbar^2)t^{16/3} - \frac{2}{3}\epsilon\Omega t^{13/3} - k^2t^4]^{1/2}}\,. \quad (4.8)$$

From the above computation of P_2, we find that in all cases, decay of probability density of an antitachyon is faster than that for a tachyon beyond one second also. Above expressions for P_2 also show that the metamass of the tachyon also helps in the decay of P_2. This shows that "heavy" tachyons decay faster.

5. ENERGY OF PRIMORDIAL SPIN-$\frac{1}{2}$ TACHYON

Like Audretsch and Schafer,[5] we define the energy density as

$$\rho = T_{\alpha\beta}u^\alpha u^\beta = \frac{i}{2S}\left[\phi^\dagger\frac{\partial\phi}{\partial T} + \frac{\partial\phi^\dagger}{\partial T}\phi\right], \quad (5.1)$$

where u^α is the four-velocity satisfying $u^\alpha u_\alpha = +1$ in the case of tachyons.

As in the previous section, here also we are not interested in the case when $t < 1$ second. Hence we consider the solution (3.20) only. From Eqs. (3.10) and (3.20) we have

$$\rho = -(B^2/R)[\eta'^{3/2} + \tfrac{1}{2}i\bar\eta'^2]\,,$$

but $\eta'^{-2} \approx (\hbar^2/\Omega^2t^{2q})[1 + 2\epsilon q\hbar^2t^{-1}/\Omega + 2\hbar^2k^2t^{-2q}/\Omega]$ which is a very small quantity for a particular value of t, and also this quantity goes on decreasing with the passage of time. Hence,

$$\rho = -(B^2/R)\eta'^{3/2}\,. \quad (5.2)$$

Substituting the value of η' from Eq. (3.18) we have

$$\rho = -\frac{B^2\Omega^{3/2}t^{3q/2}}{\hbar^{3/2}R}\left[1 - \frac{3\epsilon q\hbar}{2\Omega}t^{-1} - \frac{3\hbar k^2}{2\Omega^2}t^{-2q}\right]. \quad (5.3)$$

Hence the energy of tachyon is given by

$$E = \int \rho\, d^3V = \int \rho\, dX\, dY\, dZ = \int \rho R^2\, dR\, d\omega\,, \quad (5.4)$$

where $d\omega = \sin\theta\, d\theta\, d\Phi$ gives the solid angle.

Now connecting Eqs. (5.3) and (5.4) we have

$$\begin{aligned}
E &= -\frac{4\pi B^2\Omega^{3/2}}{\hbar^{3/2}}\int t^{3q/2}\left[1 - \frac{3\epsilon q\hbar^2}{2\Omega}t^{-1} - \frac{3\hbar^2k^2}{2\Omega^2}t^{-2q}\right]R\, dR \\
&= \frac{4\pi B^2\Omega^{3/2}q}{\hbar^{3/2}}\int\left[\frac{3\hbar^2k^2}{2\Omega^2}t^{3q/2-1} + \frac{3\epsilon q\hbar^2}{2\Omega}t^{7q/2-2} - t^{7q/2-1}\right]dt \\
&= \frac{4\pi B^2\Omega^{3/2}q}{\hbar^{3/2}}\left[\frac{\hbar^2k^2}{q\Omega^2}t^{3q/2} + \frac{3\epsilon q\hbar^2}{2\Omega(7q/2-1)}t^{7q/2-1} - \frac{2t^{7q/2}}{7q}\right] \\
&= -\frac{8\pi B^2\Omega^{3/2}}{7\hbar^{3/2}}t^{7q/2} + 4\pi B^2\Omega^{3/2}q\hbar^{1/2} \\
&\quad\times\left[\frac{k^2}{q\Omega^2}t^{3q/2} + \frac{3\epsilon q t^{7q/2-1}}{\Omega(7q-2)}\right].
\end{aligned} \quad (5.5)$$

For any value of $t > 1$ second

$$\frac{8\pi B^2\Omega^{3/2}t^{7q/2}}{7\hbar^{3/2}} \gg 4\pi B^2\Omega^{3/2}q\hbar^{1/2} \times\left[\frac{k^2}{q\Omega^2}t^{3q/2} + \frac{3\epsilon q t^{7q/2-1}}{\Omega(7q-2)}\right].$$

Hence we have expression for energy of primordial spin-$\frac{1}{2}$ tachyon:

$$E \approx -8\pi B^2\Omega^{3/2}t^{7q/2}/7\hbar^{3/2}\,. \quad (5.6)$$

For our convenience we can choose constant B such that $8\pi B^2/7 = 1$; hence

$$E \approx -\Omega^{3/2} t^{7q/2}/\hbar^{3/2}. \tag{5.7}$$

Now we are in a position to get the expression for energy of a primordial spin-$\frac{1}{2}$ tachyon in different perfect fluids substituting different values of q.

(a) In the case of superdense matter $q = \frac{1}{3}$; hence,

$$E \approx -\Omega^{3/2} t^{7/6}/\hbar^{3/2}. \tag{5.8}$$

(b) In the case of nonrelativistic matter, $q = \frac{2}{5}$; hence,

$$E \approx -\Omega^{3/2} t^{7/5}/\hbar^{3/2}. \tag{5.9}$$

(c) In the case of radiation, $q = \frac{1}{2}$; hence,

$$E \approx -\Omega^{3/2} t^{7/4}/\hbar^{3/2}. \tag{5.10}$$

(d) In the case of dust, $q = \frac{2}{3}$; hence,

$$E \approx -\Omega^{3/2} t^{7/3}/\hbar^{3/2}. \tag{5.11}$$

From above expressions for E, we note that energy of a tachyon decreases rapidly with time. Order of energy decay is given as

$$E_d < E_r < E_n < E_s \tag{5.12}$$

(d, r, n, and s denote dust, radiation, nonrelativistic matter, and superdense matter, respectively).

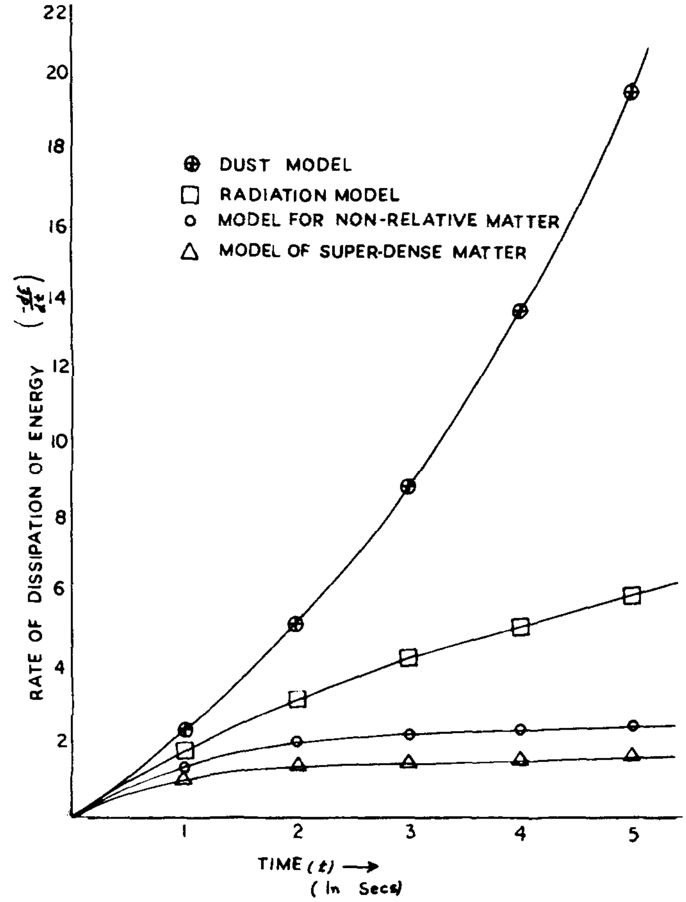


FIG. 1. Graph of rate of dissipation of energy versus time is plotted. As a result a comparative study of dissipation of energy in different models is done.

It is also interesting to note that energy decay of "heavy" tachyons is very fast.

We can have the rate of dissipation of energy in all four cases given above as

(a) In the case of superdense matter

$$-\frac{dE}{dt} = \frac{7}{6} \frac{\Omega^{3/2}}{\hbar^{3/2}} t^{1/6}. \tag{5.13}$$

(b) In the case of nonrelativistic matter

$$-\frac{dE}{dt} = \frac{7}{5} \frac{\Omega^{3/2}}{\hbar^{3/2}} t^{2/5}. \tag{5.14}$$

(c) In the case of radiation

$$-\frac{dE}{dt} = \frac{7}{4} \frac{\Omega^{3/2}}{\hbar^{3/2}} t^{3/4}. \tag{5.15}$$

(d) In the case of dust

$$-\frac{dE}{dt} = \frac{7}{3} \frac{\Omega^{3/2}}{\hbar^{3/2}} t^{4/3}. \tag{5.16}$$

We can find the comparison of rate of dissipation of energy of a tachyon of given mass in Fig. 1.

6. PRIMORDIAL SPIN-$\frac{1}{2}$ TACHYON IN THE PRESENT DAY UNIVERSE

In this section, we shall deal with the question of survival of spin-$\frac{1}{2}$ tachyons up to the present epoch.

From the age of 10^8 years up to the present epoch, it is expected that nonrelativistic matter dominates in the universe.[8] Hence, for the present epoch t_0 we consider the expression for energy,

$$|E| \approx + \frac{\Omega^{3/2} t_0^{7/5} c^2}{\hbar^{3/2}} \tag{6.1}$$

(in cgs units).

We know that experiments to detect tachyons have failed so far. Hence we may expect that energy of a primordial tachyon surviving up to the present epoch would be even less than the energy of an electron. Therefore, we have

$$\Omega^3 t_0^{14/5} c^4/\hbar^3 < m_e^2 c^4, \tag{6.2}$$

where m_e is the mass of an electron and $t_0 = 3.2 \times 10^{17}$ sec (present age of the universe). Equation (6.2) yields

$$\begin{aligned}\Omega &\lesssim \hbar \left[\frac{m_e^2}{t_0^{14/5}}\right]^{1/3} \\ &= 1.05 \times 10^{-27} \times \left[\frac{(9.11)^2 \times 10^{-56}}{(3.2)^{14/5} \times 10^{238/5}}\right]^{1/3} \\ &= 8.77 \times 10^{-54}\ \text{g}.\end{aligned}$$

We also know that

$$E^2 - c^2p^2 = -\Omega^2 c^4;$$

hence,

$$p^2 = \Omega^2 c^2 + E^2/c^2. \tag{6.3}$$

Connecting Eqs. (6.1) and (6.3) we have

$$
\begin{aligned}
p^2 &= \Omega^2 c^2 + \frac{\Omega^3 c^2 t_0^{14/5}}{\hbar^3} \\
&= \frac{c^2 \Omega^3 t_0^{14/5}}{\hbar^3}\left[1 + \frac{\hbar^3}{\Omega t_0^{14/5}}\right].
\end{aligned}
$$

It gives

$$
\begin{aligned}
p &= \frac{c\Omega^{3/2} t_0^{7/5}}{\hbar^{3/2}}\left[1 + \frac{\hbar^3}{2\Omega t_0^{14/5}}\right] \\
&\approx \frac{c\Omega^{3/2} t_0^{7/5}}{\hbar^{3/2}}. \qquad (6.4)
\end{aligned}
$$

The wavelength of a tachyon is given by

$$
\begin{aligned}
\lambda &= \hbar/p \gtrsim \frac{\hbar^{5/2}}{c\Omega^{3/2} t_0^{7/5}} \\
&= \frac{(1.05)^{5/2} \times 10^{-135/2}}{3 \times 10^{10} \times (8.77)^{3/2} \times 10^{-81} \times (3.2)^{7/5} \times 10^{119/5}} \\
&= 4.56 \times 10^{-23}\ \text{cm}.
\end{aligned}
$$

Thus we find that if a primordial spin-$\frac{1}{2}$ tachyon survives up to the present epoch, its metamass would be less than 8.77×10^{-54} g and its wavelength would be greater than 4.56×10^{-23} cm. This consequence agrees with Narlikar and Sudarshan's result[1] that a primordial tachyon, surviving up to the present epoch, should be lighter than an electron. The Compton wavelength of such tachyons would be greater than 4×10^{15} cm.

[1] J. V. Narlikar and E. C. G. Sudarshan, Mon. Not. R. Astron. Soc. **175**, 105 (1976).

[2] S. K. Srivastava, J. Math. Phys. **23**, 1981 (1982).

[3] O. M. P. Bilanuik, V. K. Deshpande, and E. C. G. Sudarshan, Am. J. Phys. **30**, 718 (1962).

[4] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).

[5] J. Andretsch and G. Schafer, J. Phys. A: Math. Gen. **11** (8), 1583 (1978).

[6] L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill, New York, 1968).

[7] M. P. Pathak, Ph.D. thesis, Gorakhpur University, Gorakhpur (1975).

[8] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).

Quasisteady arterial blood flow

Paul Gordon
Department of Mathematical Sciences, Purdue University Calumet, Hammond, Indiana

(Received 1 December 1981; accepted for publication 12 March 1982)

Fluid flow through a bifurcating system of elastic tubes is modeled with the inviscid compressible time-dependent equations of fluid flow. Because of the physical scale of the problem (the length of the tube is large relative to its diameter), quasisteady analysis can be applied to the system of hyperbolic first order partial differential equations. Several simplified models are thereby obtained. Although similar in some respects to the usual incompressible models, the quasisteady models are seen to be significantly different. For example, the sound speed of the fluid remains an important parameter in the boundary condition. Numerical solutions are obtained for various problems. Results are presented for the one-layer problem, which was studied in some depth, and a three-layer problem, in which the layers are connected by bifurcation boundary conditions. The solutions are shown to be physically reasonable. Of special interest is the clear display of the mechanism in the model by which the elastic wall converts the pulsatile flow to a much more continuous outflow.

PACS numbers: 87.45.Hw, 47.40.Dc

I. INTRODUCTION

In 1970 the author collaborated on a study[1] in which the compressible equations were used to model arterial blood flow. It was shown that the retention of the finite sound speed did in fact produce a sensible set of equations. However, resolution of the time step restriction, caused by the sound speed, required appropriate quasisteady analysis. In Ref. 1 we attempted to apply the quasisteady features of the problem through the numerics. This was not completely successful and consequently a full numerical solution was not obtained.

Since that time the author has worked on the problem of analytically applying quasisteady analysis to hyperbolic systems of partial differential equations. This effort seems to have been successful in the physical problem of large-scale weather prediction.[2,3] The purpose of the present paper is to apply similar analysis to the time-dependent Euler equations for the physical problem corresponding to arterial blood flow.

Womersley's work[4] initiated a large and extended effort in terms of modeling pulsatile blood flow. For the most part this work falls into two categories.

(i) The first attempts to obtain a complete description of all the flow variables: The starting point usually is the incompressible viscous Navier–Stokes equations, although non-Newtonian fluids have also been considered in this framework. This kind of analysis is described thoroughly by McDonald[5]; see also Refs. 6 and 7. The complexity of the mathematics has thus far precluded three-dimensional studies or analysis of an entire arterial tree.

(ii) The second category pursues wave analysis and attempts to calculate propagation speeds and "reflective" points in the system. Often an entire system is modeled. See, for example, Refs. 8–10.

Our reasons for using an inviscid, compressible model were discussed in Ref. 1 and are briefly as follows.

(1) There is both theoretical and experimental evidence to support the hypothesis that pulsatile arterial flow in large mammals is, except for thin regions near the wall, plug flow.[11–14] Also, our numerical work with viscous flows indicates that the spatial and time scales of the physical problems preclude the possibility of viscous effects, except for thin regions near the wall.[15,16]

(2) The incompressible equations are not consistent with the compressible equations. The difficulty lies in the fact that the assumption of incompressibility is not applied consistently to the sound speed that exists in the compressible model. This fact was established analytically in the case of asymptotic stability.[17]

Quasisteady analysis attempts to simultaneously apply scale assumptions (rather than an incompressibility assumption) to both the partial differential equations and the boundary conditions. It will be seen that the resulting models are not entirely dissimilar to the incompressible models, but at the same time there are important differences.

In Sec. II the complete time-dependent problem is specified, including all equations used at the boundary. In Sec. III the quasisteady analysis is applied to both the partial differential equations and the boundary conditions. Several systems are obtained. The final set of equations involves the calculation of variables in only one space dimension. Results of the calculations are discussed in Sec. IV and conclusions are summarized in Sec. V.

II. TIME-DEPENDENT INVISCID COMPRESSIBLE MODEL

Assuming an equation of state of the form

$$P - P_0 = c^2(\rho - \rho_0), \tag{1.1}$$

where c is the sound speed (taken to be constant), the energy equation uncouples from the system. The remaining equations of the inviscid compressible Euler equations in cylindrical coordinates (assuming angular symmetry) are

conservation of mass:

$$\rho_t + w\rho_r + u\rho_z = -\rho(w_r + u_z + w/r); \tag{1.2}$$

conservation of momentum (r):

$$w_t + ww_r + uw_z = -\frac{1}{\rho}P_r; \tag{1.3}$$

0022-2488/83/041001-10$02.50

conservation of momentum (z):

$$u_t + wu_r + uu_z = -\frac{1}{\rho} P_z \,, \tag{1.4}$$

where ρ = density, P = pressure, w = radial velocity (m/sec), u = horizontal velocity (m/sec), r = radial coordinate, and z = horizontal coordinate. If we now take $\rho \equiv \rho_0$ wherever it appears as a coefficient, then from Eq. (1.1), the system can be taken as

$$P_t + wP_r + uP_z = -P_0(w_r + u_z + w/r) \,, \tag{2.1}$$

$$w_t + ww_r + uw_z = -\frac{c^2}{P_0} P_r \,, \tag{2.2}$$

$$u_t + wu_r + uu_z = -\frac{c^2}{P_0} P_z \,. \tag{2.3}$$

The physical situation we intend to model is shown schematically in Fig. 1. Symmetrical bifurcations are assumed so that we need consider only one segment leaving each bifurcation. Equations (2) are intended to be valid at interior points of each region. Additional boundary conditions are then required as follows.

(i) Inlet conditions (presumably from the heart) need to be specified at location H.

(ii) Outlet or downstream conditions (at, for example, location 0) need to be specified so as to reflect the physical assumptions being made in regard to the remainder of the system.

(iii) In each segment conditions must be specified at the outer wall and at the center line.

(iv) Conditions connecting the segments need to be specified at each bifurcation (locations B).

As in Ref. 1, the number of required boundary conditions is determined by considering the characteristic surfaces of Eqs. (2) as follows.

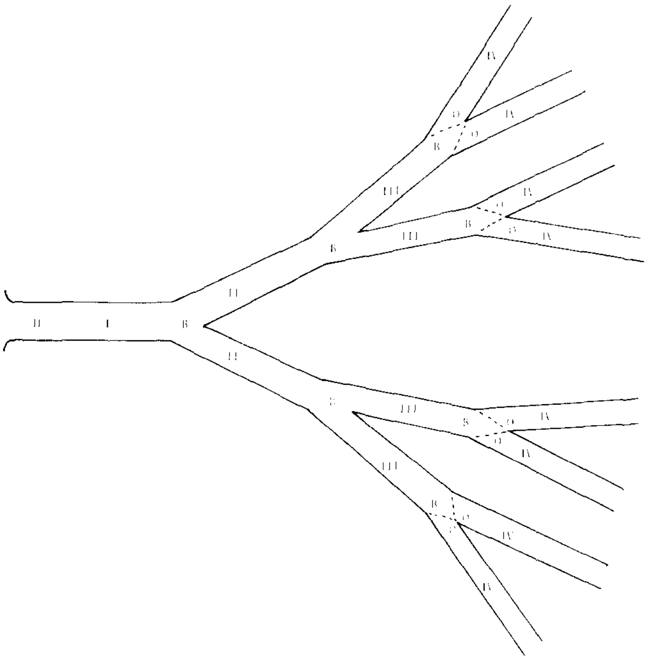


FIG. 1. Physical configuration.

> At any boundary point, a boundary condition is specified for each characteristic that enters the region of computation. The remaining characteristic variables are utilized to obtain the other variables. (3)

(See Ref. 1 for a more thorough discussion of the procedure).

The following transformation allows the elastic wall to be more easily incorporated into each segment of the system:

$$\begin{aligned} \tau &= t \,, \\ \eta &= z \,, \\ \zeta &= r/R \,, \end{aligned} \tag{4.1}$$

where $R = R\,(t,z)$ represents the position of the wall. In each segment the configuration is as shown in Fig. 2. The equations then take the form

$$\begin{aligned} &P_\tau + \alpha P_\zeta + uP_\eta \\ &\quad = -P_0(\zeta_r w_\zeta + \zeta_z u_\zeta + u_\eta + w/r) \,, \\ &w_\tau + \alpha w_\zeta + uw_\eta = -\left(\frac{c^2}{P_0}\zeta_r\right)P_\zeta \,, \\ &u_\tau + \alpha u_\zeta + uu_\eta = -\frac{c^2}{P_0}(P_\eta + \zeta_z\, P_\zeta) \,, \end{aligned} \tag{4.2}$$

where

$$\alpha = \frac{1}{R}\,[(w - \zeta R_z u) - \zeta R_t] \,.$$

In matrix form, Eqs. (4.2) become

$$V_\tau + AV_\zeta + BV_\eta = -F \,, \tag{5.1}$$

$$A = \begin{pmatrix} \alpha & P_0\,\zeta_r & P_0\,\zeta_z \\ (c^2/P_0)\,\zeta_r & \alpha & 0 \\ (c^2/P_0)\,\zeta_z & 0 & \alpha \end{pmatrix}, \quad B = \begin{pmatrix} u & 0 & P_0 \\ 0 & u & 0 \\ c^2/P_0 & 0 & u \end{pmatrix},$$

$$F = \begin{pmatrix} P_0(w/r) \\ 0 \\ 0 \end{pmatrix}, \quad V = \begin{pmatrix} P \\ w \\ u \end{pmatrix}.$$

The eigenvalues λ of A are

$$\lambda = \alpha, \alpha \pm c(\zeta_r^2 + \zeta_z^2)^{1/2} \,, \tag{5.2}$$

and the eigenvalues of B are

$$\lambda = u, u \pm c \,. \tag{5.3}$$

The characteristic equations corresponding to (5.2) are

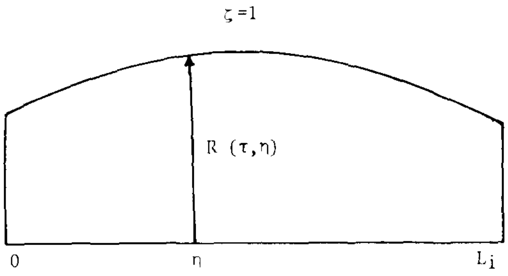


FIG. 2. Coordinate system.

$$(\zeta_z w - \zeta_r u)_\tau = -\alpha(\zeta_z w - \zeta_r u)_\zeta + G_1 , \tag{6.1}$$

$$(\tilde{c}P + P_0\,\zeta_r w + P_0\,\zeta_z u)_\tau$$
$$= -(\alpha + \tilde{c})(\tilde{c}P + P_0\,\zeta_r w + P_0\,\zeta_z u)_\zeta + G_2 , \tag{6.2}$$

$$(\tilde{c}P - P_0\,\zeta_r w - P_0\,\zeta_z u)_\tau$$
$$= -(\alpha - \tilde{c})(\tilde{c}P - P_0\,\zeta_r w - P_0\,\zeta_z u)_\zeta + G_3 , \tag{6.3}$$

where the G_i represent the remaining terms in the equation and $\tilde{c} = c(\zeta_r^2 + \zeta_z^2)^{1/2}$.

The characteristic equations corresponding to (5.3) are

$$w_\tau = -uw_\eta + G_4 , \tag{6.4}$$

$$(cP + P_0 u)_\tau = -(u + c)(cP + P_0 u)_\eta + G_5 , \tag{6.5}$$

$$(cP - P_0 u)_\tau = -(u - c)(cP - P_0 u)_\eta + G_6 . \tag{6.6}$$

Boundary conditions need to be specified at $\eta = 0$ and $\eta = L_i$ and at $\zeta = 0$ and $\zeta = 1$.

(i) $\eta = 0$: Suppose this is the inlet position. (The bifurcation points will be discussed below). Assuming $|u| \ll c$, from Eq. (5.3) we see that the characteristic surface corresponding to $u + c$ enters the region, while that corresponding to $u - c$ leaves the region. This then is the situation.

(a) If $u > 0$, two boundary conditions are needed. We chose

$$u = f_1(t,\xi) , \tag{7.1}$$

$$w = f_2(t,\xi) , \tag{7.2}$$

where f_1 and f_2 are specified functions. In addition the characteristic equation (6.6) is used.

(b) If $u < 0$, one boundary condition is needed. We chose Eq. (7.1). The additional characteristic equations are (6.4) and (6.6)

(ii) $\eta = L$: Suppose this represents the outlet positions. It is assumed that $u \geqslant 0$ here so that, from Eq. (5.3), only one boundary condition is needed. This downstream condition was discussed at some length in both Refs. 1 and 3. The following equation will be used:

$$\rho_t = \frac{\rho}{c}\,u_t . \tag{7.3}$$

The additional characteristic equations are (6.4) and (6.5). We believe that Eq. (7.3) represents the physical situation of an infinite rigid tube of constant diameter beyond the outlet position. Equation (7.3) will be used in the form

$$P_t = \frac{P_0}{c}\,u_t . \tag{7.4}$$

Downstream boundary conditions cause difficulties in almost all areas of fluid dynamics. A common approach is to apply an infinite or steady condition at the boundary. Lieberstein[18] deals with this question carefully and emphasizes the importance of a meaningful boundary condition. He proposes a "Windkessel" condition, which very nicely closes the mathematical system. However, as Lieberstein himself states, the Windkessel condition is of questionable physical significance. As noted above, we think Eq. (7.4) has a well-defined physical interpretation and is at the same time mathematically sound.

(iii) $\zeta = 0$: Because of symmetry the natural condition here is

$$w = 0 . \tag{7.5}$$

From (4.2) this gives $\alpha = 0$ so that, from (5.2), no additional boundary conditions are needed. Instead the characteristic equations (6.1) and (6.2) are used.

(iv) $\zeta = 1$: The natural condition here is that fluid cannot permeate the wall; that is, the normal component of velocity is equal to the normal component of the wall velocity. In terms of our variables, this becomes

$$w - R_z u = R_t . \tag{7.6}$$

As above, Eq. (7.6) gives $\alpha = 0$ so that no further boundary conditions for the flow variables are needed [the characteristic equations (6.1) and (6.2) can be used]. However, an expression for R_t needs to be specified. This is a very important part of the problem. This was discussed in Ref. 1, and we use the same expression given there:

$$\frac{1}{R_0} R_t = \frac{(1 + R_z^2)^{1/2}}{K_1} \left\{ \left(\frac{P - \bar{P}}{P_{eq}} \right) \frac{f_3(z)}{f_4(R)} - \left(\frac{R - R_0}{R_0} \right) \right\} , \tag{7.7}$$

where $\bar{P}$ is the pressure corresponding to the initial state of the tube when it is inflated to the constant radius R_0, and K_1 and P_{eq} will be chosen from the experimental data of Peterson *et al.*[19] f_3 and f_4 are defined as follows:

$$f_3 = \frac{1 - e^{-c_1(z/L)(1 - z/L)}}{1 - e^{-c_1/4}} , \tag{8.1}$$

$$f_4 = 1 + C_2(R/R_0)^2 . \tag{8.2}$$

For $C_1 > 0, f_3$ serves to "tether" each segment at the endpoints; that is, $R_t \equiv 0$ at $\eta = 0$. For $C_2 > 0, f_4$ is such that the distensibility of the tube depends on the expansion that has occurred: The greater the expansion, the more difficult it is to expand further.

Finally, we need to specify conditions at bifurcations. We assume a "sharp" discontinuity, so that the appropriate jump conditions are obtained from consideration of conservation of mass and momentum. This is discussed in the Appendix. Let "0" and "1" denote points on either side of the bifurcation. Let A_0 and A_1 represent respective total areas on either side of the bifurcation. The equations are

$$u_1 A_1 = u_0 A_0 , \tag{9.1}$$

$$P_1 - P_0 = \frac{P_0}{2}\left(\frac{u_0}{c_0}\right)^2 \left[1 - \frac{A_0^2}{A_1^2}\right] = 3.750u_0^2 \left(1 - \frac{A_0^2}{A_1^2}\right) . \tag{9.2}$$

Note that, at position 0 in Fig. 1, Eq. (7.4) is applied to the infinite rigid tube beyond 0 (u is constant in this entire region).

III. QUASISTEADY MODELS

The procedure for deriving the quasisteady equations will be similar to that described in Refs. 2 and 3, and will be based on the following physical assumption [see assumption (2) in Ref. 2]:

> Equation (5.1) is to be solved in a region of radius h and length L, with $h \ll L$. It is assumed that boundary conditions and initial conditions to be imposed on Eq (5.1) are such that the flow varia-

> bles will experience significant variations only over time scales which are large compared to h/c.
>
> (10)

Mathematically, assumption (10) is interpreted as follows. Dependent variables, which can react to perturbation on a time scale of h/c, are in quasisteady equilibrium with the slowly reacting dependent variables. The problem is to find these variables, while simultaneously considering appropriate quasisteady limits of boundary conditions.

To find the quasisteady variables, introduce the following:

$$\tilde{u} = (u - aw)/(1 + a^2)^{1/2}, \qquad (11.1)$$

$$\tilde{w} = (w + au)/(1 + a^2)^{1/2}, \qquad (11.2)$$

where

$$a = -\left(\frac{r}{R}\right)\left(\frac{\partial R}{\partial z}\right). \qquad (11.3)$$

Equation (5.1) becomes

$$W_\tau + A_1 W_\zeta + B_1 W_\eta = -F_1, \qquad (12)$$

where

$$A_1 = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & P_0(1+a^2)^{1/2}\zeta_r \\ 0 & c^2(1+a^2)^{1/2}\zeta_r/P_0 & \alpha \end{pmatrix},$$

$$B_1 = \begin{pmatrix} u & c^2/P_0(1+a^2)^{1/2} & 0 \\ P_0/(1+a^2)^{1/2} & u & aP_0/(1+a^2)^{1/2} \\ 0 & ac^2/P_0(1+a^2)^{1/2} & u \end{pmatrix},$$

$$F_1 = \begin{pmatrix} \left(\dfrac{\tilde{w}}{1+a^2}\right)\dfrac{da}{d\tau} \\ P_0\left(\dfrac{wa_z - ua_r}{1+a^2} + \dfrac{w}{r}\right) \\ -\left(\dfrac{u}{1+a^2}\right)\dfrac{da}{d\tau} \end{pmatrix}, \quad w = \begin{pmatrix} \tilde{u} \\ P \\ \tilde{w} \end{pmatrix},$$

and

$$\frac{d}{d\tau} = \frac{\partial}{\partial\tau} + \alpha\frac{\partial}{\partial\zeta} + u\frac{\partial}{\partial\eta}.$$

The eigenvalues of A_1 are given by Eq. (5.2) and consequently we conclude that P and $\tilde{w}$ are in quasisteady equilibrium with respect to $\tilde{u}$. From Eq. (12) we therefore obtain the following:

$$\frac{d\tilde{u}}{d\tau} = -\frac{c^2}{P_0(1+a^2)^{1/2}}P_\eta - \left(\frac{w}{1+a^2}\right)\frac{da}{d\tau}, \qquad (13.1)$$

$$0 = (\tilde{u}_\eta + a\tilde{w}_\eta) + \zeta_r(1+a^2)\tilde{w}_\zeta + \frac{(1+a^2)^{1/2}}{P_0}(uP_\eta + \alpha P_\zeta) + (1+a^2)^{1/2}\left(\frac{wa_z - ua_r}{1+a^2} + \frac{w}{r}\right), \qquad (13.2)$$

$$0 = aP_\eta + (1+a^2)\zeta_r P_\zeta + \frac{P_0(1+a^2)^{1/2}}{c}(u\tilde{w}_\eta + \alpha\tilde{w}_\zeta) - \left(\frac{P_0\tilde{u}}{c^2(1+a^2)^{1/2}}\right)\frac{da}{d\tau}. \qquad (13.3)$$

In considering boundary conditions, the underlying principles [Ref. 3, Eqs. (15) and (16)] are as follows.

> If a boundary equation involves P_τ or $\tilde{w}_\tau$, and if the equation involves flow conditions internal to the region of computation, then the equation will be put in quasisteady equilibrium.
>
> (14)
>
> The basic quasisteady assumption, Eq. (10), requires that (13.2) and (13.3) be applied at all interior points of the lateral boundaries.
>
> (15)

Remark: As discussed in Ref. 3, if (15) is not imposed, the solution may develop steep gradients which are then inconsistent with assumption (10). We now apply (14) and (15) to the various boundaries.

(i) $\eta = 0$: Eq. (7.2) is deleted by (15) and Eq. (6.6) is deleted by (14).

(ii) $\eta = L$: Eqs. (6.4) and (6.5) are deleted by (14).

(iii) $\zeta = 0$: Eqs. (6.1) and (6.2) are deleted by (14).

(iv) $\zeta = 1$: Eqs. (6.1) and (6.2) are deleted by (14).

Summarizing, the boundary conditions to be used in conjuction with Eqs. (13) are as follows.

> Equation (7.1) at $\eta = 0$, Eq. (7.4) at $\eta = L$, Eq. (7.5) at $\zeta = 0$, Eqs. (7.6) and (7.7) at $\zeta = 1$, and Eqs. (9) at bifurcation points.
>
> (16)

Equation (13) and (16) would seem to be a reasonable quasisteady model and would be particularly applicable to the case of a "long" tube for which $\Delta t \sim L/c$, where Δt is the time scale of interest in terms of variation of boundary conditions. In the problems to be considered herein, $L \sim 0.3$ m, $\Delta t \sim 1$ sec., and $c = 1469.14$ m/sec. Consequently, Eq. (13.1) can also be put into quasisteady equilibrium. Application of (14) and (15), with $\tilde{u}_\tau$ replacing $\tilde{w}_\tau$, shows that Eqs. (16) still are suitable.

A second model therefore consists of the following.

(i) $A_1 W_\zeta + B_1 W_\eta + F_1 = 0$ [from Eq. (12)],

(ii) boundary conditions (16). (17)

Preliminary calculations were made with Eqs. (17); actually, Eqs. (13) were iterated to a steady state. The results were reasonable, but for the problems to be considered herein the equations were much more complicated than necessary, particularly because of the small variation in pressure.

Expected magnitudes for various quantities are as follows:

$$|u| < 1, \quad |R_t| \sim |w| < 0.01, \quad |a| \sim |R_z| < 0.01.$$

One expects to obtain P_ζ from Eq. (13.3). With the quantities given above, $|P_\zeta| < 10^{-4}|P_\eta|$ and the dominant terms in Eq. (13.1) are $P_\eta \sim -(P_0/c^2)uu_\eta$. Thus, we can assume

$$P_\zeta = 0, \qquad (18.1)$$

$$P_\eta = -\frac{P_0}{c^2}uu_\eta. \qquad (18.2)$$

Also, one finds in Eq. (13.2) that the dominant terms should be

$$\tilde{u}_\eta + \frac{1}{R}\left[\tilde{w}_\xi + \frac{\tilde{w}}{\xi} + 2R_z\tilde{u}\right] = 0 . \tag{18.3}$$

Equation (18) and the appropriate boundary conditions need to be solved simultaneously, but the general dependence is as follows.

(i) $\tilde{w}(t,0,\eta) = 0$ and Eq. (18.3) give the $\tilde{w}$ field,
(ii) Eq. (7.6) gives $R\,(t,\eta)$,
(iii) Eqs. (7.7) and (18.1) give the P field,
(iv) Eqs. (7.1) and (18.2) give the $\tilde{u}$ field.
The solution is correct when Eq. (7.4) is satisfied.

One further simplification is obtained by assuming that both $|u_\xi|$ and $|P_\xi|$ are negligible. Let starred quantities represent wall values, $f^*(t,\eta) = f(t,R\,(\eta),\eta)$. Specifically, we assume that w is linear in ξ, or

$$\tilde{w} = \xi w^* .$$

Then, Eqs. (18.2) and (18.3) take the form

$$\frac{\partial u^*}{\partial\eta} = -\frac{2}{R}\,[w^* + R_z u^*] , \tag{19.1}$$

$$\frac{\partial P^*}{\partial\eta} = -\frac{P_0}{c^2}\,u^*\left(\frac{\partial u^*}{\partial\eta}\right) . \tag{19.2}$$

It turns out that, for the velocity distributions to be considered, the right side of Eq. (19.2) is negligible (this will be discussed further in Sec. IV). With this further assumption, we obtain

$$\frac{\partial u^*}{\partial\eta} = -\frac{2}{r}\,[w^* + R_z u^*] , \tag{20.1}$$

$$\frac{\partial P^*}{\partial\eta} = 0, \quad \text{or } P^* = P^*(\tau) . \tag{20.2}$$

Equation (20.2) is perhaps surprising since it assumes that the spatial variation of pressure is negligible throughout the system. This would seem to be consistent with the results reported by Gams *et al.*[20]

From Eqs. (7.6) and (7.7),

$$\frac{\partial R}{\partial\tau} = w^*(1 + a^2)^{1/2} \sim w^* , \tag{20.3}$$

$$w^* = \frac{R_0}{K_1}\left[\left(\frac{P^* - \bar{P}}{P_{eq}}\right)\frac{f_3}{f_4} - \left(\frac{R - R_0}{R_0}\right)\right] , \tag{20.4}$$

where f_3 and f_4 are defined by Eqs. (8.1) and (8.2). The boundary conditions to be used with Eqs. (20) are as follows.

$$u^*(\tau,0) = f_1(\tau) , \tag{21.1}$$

$$\frac{\partial P^*}{\partial\tau} = \frac{P_0}{c}\frac{\partial u^*}{\partial\tau}, \quad \text{at the outlet point.} \tag{21.2}$$

Summarizing, we have presented the following quasisteady systems of equations.

(i) Eqs. (13) and (16): From a numerical point of view, the time step restriction is related to L/c rather than R/c.

(ii) Eqs. (17): This in fact was the model that we attempted to solve in Ref. 1.

(iii) Eqs. (18) and (16): This is a simplified form of Eqs. (17).

(iv) Eqs. (20) and (21): This is the simplest model since it involves only one space dimension. Solutions to this will be presented in the next section.

IV. DISCUSSION OF NUMERICAL CALCULATIONS

The calculations to be described have been made with the model consisting of Eqs. (20) and (21). The intent is to show that the model is qualitatively reasonable. Although the results are significantly affected quantitatively by the values of the various parameters that need to be specified, a complete parameter study will not be undertaken. (This is expected to be part of a later study in which an attempt will be made to model the arterial system of an animal).

Several preliminary comments need perhaps to be made.

(1) Equation (20.4) is a "passive" equation in that the wall moves in response to the flow field. Thus, the heart does not function as a "peripheral heart": This was one of the conclusions of the study in which Eq. (20.4) was derived.[19]

(2) This model resembles in some aspects the usual incompressible inviscid equations. However, there are important distinctions. For example, the quasisteady nature of the horizontal velocity is not usually associated with the incompressible model. Also, boundary conditions (21.2) involve the sound speed and would not usually be part of an incompressible model.

The following values were chosen and kept fixed for all layers for all calculations:

$$P_0 = 1.6190\times 10^7\text{mm Hg} , \tag{22.1}$$

$$K_1 = 0.0355 \text{ sec} , \tag{22.2}$$

$$P_{eq} = 1159.9 \text{ mm Hg} . \tag{22.3}$$

The above values correspond to the experimental conditions of Peterson *et al.*[19]; in particular, K_1 and P_{eq} are the reported results for the thoracic aorta of one of the experimental animals.

The function f_3, Eq. (8.1), is shown in Fig. 3 for various values of C_1. It turns out that the results are not greatly dependent on the particular value chosen (although the calculations become sensitive for extremely large values). For all calculations we chose the value

$$C_1 = 4 . \tag{22.4}$$

The input functions f_1 and f_2, Eqs. (7.1) and (7.2), are specified as follows.

$$w = 0 , \tag{23.1}$$

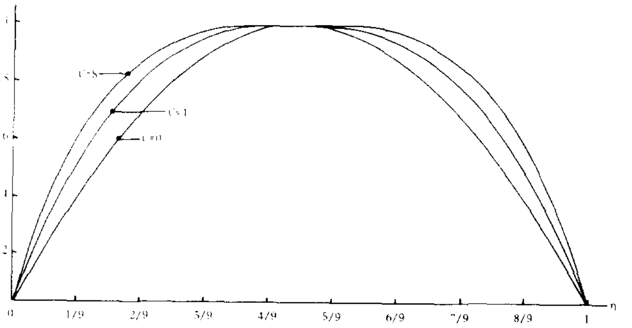


FIG. 3. Tethering function, Eq. (8.1).

$$u = f_1(t) = f_1(t + P_2) = \begin{cases} 16U_{\max}(t/P_1)^2(1 - t/P_1)^2, & 0 \leqslant t \leqslant P_1 \\ 0, & P_1 \leqslant t \leqslant P_2 \end{cases}. \qquad (23.2)$$

The above function is intended to mimic the periodic heart cycle. The flow has period P_2, maximum velocity $U_{\max}$, with inflow occurring during the time duration P_1.

The first problem consists of a single tube of length $L = 0.3048$ m and initial radius $R_0 = 0.006096$ m. (The 1 ft length is 25 times the diameter). At time zero we assumed stationary flow ($P \equiv \bar{P}, u \equiv 0, w \equiv 0, R \equiv R_0$). The inflow function was specified with $U_{\max} = 0.3048$ and $P_1 = 0.4$. For the first case we took $C_2 = 0$ in Eq. (8.2) and $P_2 = 1$ in Eq. (23.2). This produces an inflow distribution f_1 as shown in Fig. 4. The calculation was run for 40 sec (40 cycles), with $\Delta t = 0.02$ and $\Delta x = L/9$. Pressure, $\tilde{P} = P - \bar{P}$, is shown in Figs. 5 and 6. Figure 5 gives the continuous pressure distribution, while Fig. 6 plots only the maximum pressures (but a continuous curve has been drawn through the discrete points). Several points of interest are the following.

(1) Many cycles are required to achieve a steady state (Fig. 6). At 10 cycles, the peak pressure differs by approximately 15% from its final value, while at 20 cycles the difference is about 3%. A similar situation was reported by Ling and Atabek.[21]

(2) The steady-state distribution, shown in Fig. 5 as the curve between $t = 39$ and $t = 40$, has a minimum pressure well above the "rest" pressure $\bar{P}$. This is consistent with the thought that the arteries are constantly in a state of tension; i.e., pressure never returns to the $\bar{P}$ value.[22]

(3) The pressures corresponding to the systolic and diastolic pressures are, respectively, 752.9 and 675.9 mm Hg. These values are physiologically large although the difference, 77 mm Hg, is not unreasonable.

The outlet "steady-state" velocity distribution $u^*(t,L)$ is shown in Fig. 4 and with an expanded scale in Fig. 7. Several points of interest are as follows.

(1) For this case $\int_0^1 u(t,0)dt = \int_0^1 f_1(t)dt = 0.06501$. Since we are in equilibrium, and since the area at both the inlet and outlet points are constant and equal, this value should also be the average outlet velocity. For this mesh a numerical integration gave the average outlet velocity as 0.06480.

(2) Figure 4 shows clearly how the pulsatile inlet flow is transformed into a much more continuous outlet flow.

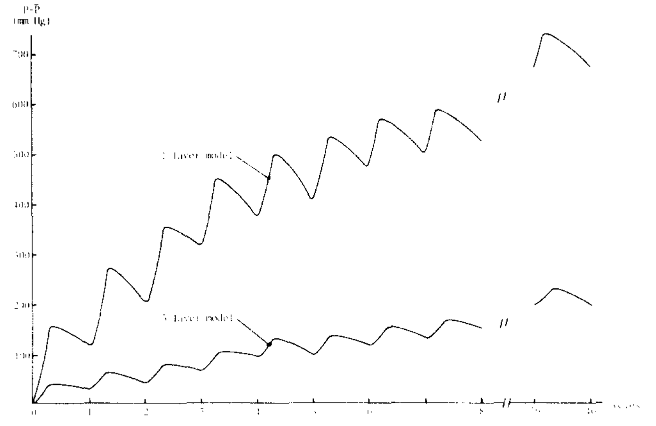


FIG. 5. Pressure distribution.

(3) Through boundary condition (21.2) one sees that the peak pressure (Fig. 6) is essentially determined by the peak outlet velocity, while the large pressure variation through the cycle correlates with the relatively small variation in the outlet velocity. Figure 7 shows the outlet velocity more clearly. The difference between the maximum and minimum velocity is $\Delta u = 0.006992$; from Eq. (21.2), this produces $\Delta p = 77.05$.

Figures 8, 9, and 10 display, respectively, the horizontal velocity, vertical velocity, and the wall position at several times during an equilibrium cycle. Figure 11 is intended to be identical to Fig. 10 except that in Fig. 11 the length and radius are in scale. Vertical velocity is of course much less in magnitude than the horizontal velocity. For this problem, even though the radial distension is more than 50% (at the center) of the initial tube radius, the actual motion of the wall is rather small. Nevertheless, this small motion accounts for the relatively constant outlet velocity.

We next investigated the effect of the period. In Eq. (23), $U_{\max}$ and P_1 were kept, respectively, at the values of 0.3048 and 0.4. The first three rows of Table I illustrate some of the steady-state results obtained by varying P_2. In all three cases the same amount of inflow occurred per period. In effect, the time between "beats" varied: for $P_2 = 1$ the rest period was 0.6, for $P_2 = 0.8$ the rest period was 0.4, and for $P_2 = 0.5$ the rest period was 0.1. The average outlet velocity should there-

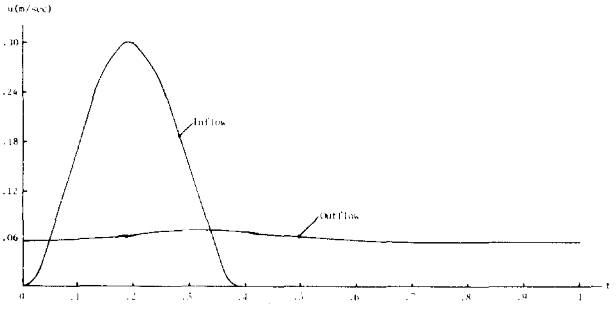


FIG. 4. Horizontal velocity (inflow and outflow).

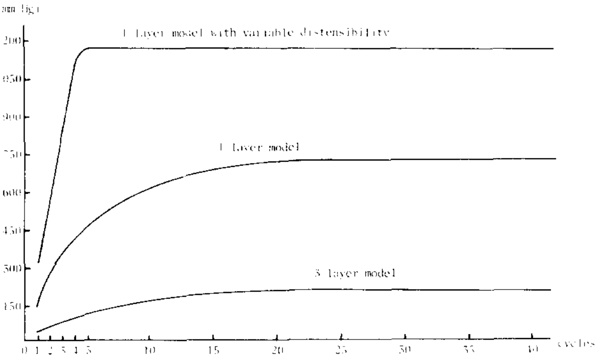


FIG. 6. Maximum pressure/cycle.

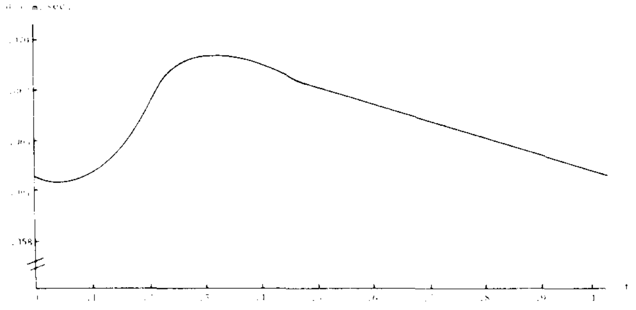


FIG. 7. Outlet velocity for the 1 layer model.

fore vary inversely as the period, and this will produce a corresponding rise in pressure. The decrease in ΔP was not anticipated: At the shorter period, the outflow is higher but more continuous. Also, at the shorter period the amount of wall distension increases significantly, while the actual wall motion decreases. The required number of cycles to achieve a steady state also increases significantly as the period decreases; this would seem to be related to the amount of wall distension that occurs.

The effect of variable distensibility was also investigated with the one-layer model. This is achieved by setting $C_2 \neq 0$ in Eq. (8.4). For our case, we chose $C_2 = 100$. Some results of this calculation are displayed in the fourth row of Table I and should be compared with the first row. Although the average outlet velocity is unchanged, the outlet velocity is not nearly as continuous for the variable distensibility case: u varies from 0.03728 to 0.1066 (as compared to 0.06130 and 0.06831). This fact is also reflected in the pressure distribution. As expected, wall distension decreases significantly and this would seem to correlate with the fact that only a few cycles are required to achieve a steady state (see Fig. 6). In summary, variable distensibility allows the flow to more rapidly adjust to the inflow conditions, but the resulting flow has a much more severe pressure requirement.

We next ran the bifurcating model. We chose the case of three layers. Referring to Fig. 1, Secs. I, II, and III are distensible tubes, while Secs. IV are assumed to be infinite rigid tubes; bifurcation conditions are applied at locations B and the outlet condition is specified at 0. Each of the first three sections was assumed to be of length 0.3048 m. Also, it was assumed that the total area increases by a factor of 50% at each bifurcation. Letting $R_0(i)$ be the initial radius of segment i, we specified $R_0(1) = 0.006096$ m, so that $R_0(2) = 0.005279$ m, $R_0(3) = 0.004572$ m, and $R_0(4) = 0.003959$ m. The input function f_1 is as before, with period 1 sec.; the constants K_1 and P_{eq} were kept at those values specified by Eqs. (22) for all layers; C_1 in Eq. (8.1) was taken as 4 in all layers, while for the first run C_2 in Eq. (8.2) was zero in all layers. The mesh was $\Delta x = 0.3048/9$ in all layers and $\Delta t = 0.02$.

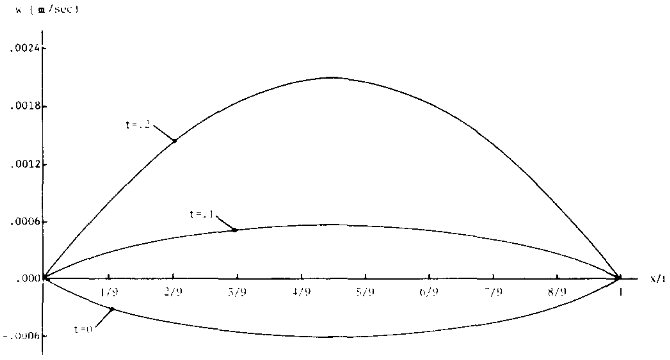


FIG. 9. Vertical velocity.

The pressure jump, predicted by Eq. (9.2), is negligible for our cases, and so we again assumed pressure to be constant throughout. The horizontal velocity distribution for this problem is shown in Fig. 12 for several time steps. The outgoing velocity, shown in Fig. 12 as a constant value of 0.0183, is plotted more precisely in Fig. 13. The jumps in velocity are a result of Eq. (9.1).

The pressure distribution as a function of time is shown in Figs. 5 and 6. Both the absolute pressure rise and the variation per cycle are significantly reduced, with both parameters now lying within reasonable physiological ranges. Table I displays some of the data for this case, as well as for the three-layer model with variable distensibility ($C_2 = 100$). The effect of variable distensibility is similar to that for the one-layer model. It is perhaps of interest to note that the ratio $\Delta P/P$ is approximately 10% for both the one-layer and three-layer models (for the case of constant distensibility).

Several other cases, which will not be described in detail, have also been run.

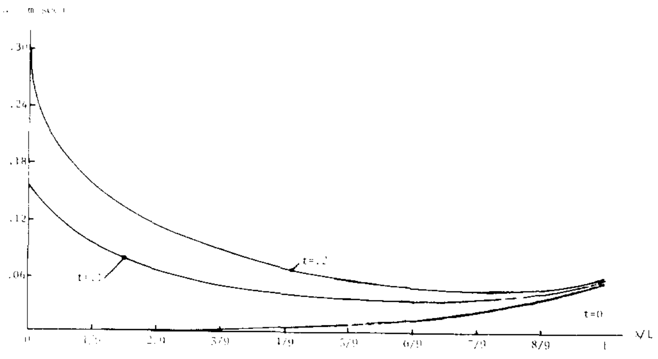


FIG. 8. Horizontal velocity.

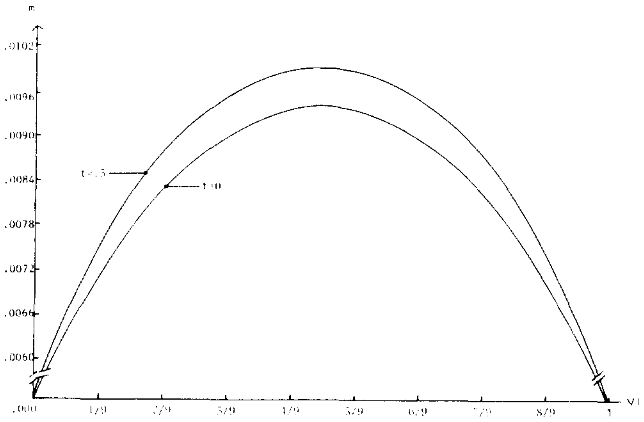


FIG. 10. Wall position.

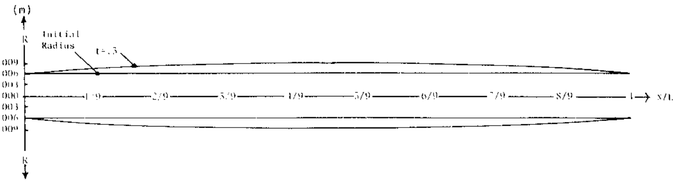


FIG. 11. Radius.

(1) The three-layer model with pressure varying with distance [Eq. (9.2) was applied at the bifurcations and Eq. (18.2) at interior points]. The maximum total variation in horizontal pressure at any time was 0.25 mm Hg, while the difference from those values shown in Table I was at most 0.5 mm Hg.

(2) Cases with backflow at the inlet point (during the "rest" part of the cycle) were calculated. The results were reasonable and indicated that the boundary conditions were satisfactory.

Finally, we note that the accuracy of the computation was tested in two ways.

(1) As discussed in previous work[2,3] a sequence of three calculations with decreasing mesh size was made; the mesh size, both Δx and Δt, is cut by a factor of 2. Since the numerics is essentially first order, the differences in calculated values should be cut by a factor of 2. That is, if $\theta(t,x,h)$ represents any of the flow variables calculated at time t and point x with mesh size $(\Delta x, \Delta t)$, then

$$|\theta(t,x,h/2) - \theta(t,x,h/4)| \leqslant \tfrac{1}{2}|\theta(t,x,h) - \theta(t,x,h/2)| . \quad (24)$$

Equation (24) is considered a good test for the numerics, and it was satisfied.

(2) A mass balance calculation was performed at each time step. Essentially, we integrate the inflow and outflow and compare their difference with the mass stored in the expanding tubes. In all cases the error was less than 0.1%. Since this error did not decrease significantly with mesh size, it is attributed to the quasisteady assumptions.

V. SUMMARY AND CONCLUSIONS

We have attempted to apply the time-dependent inviscid compressible Euler equations to flow through elastic tubes with parameters comparable to arterial flow in mammals. The mathematical procedure was as follows.

(i) Using characteristics (principle 3), boundary conditions were specified for Eqs. (2). It was assumed that the result is a well-posed mathematical problem.

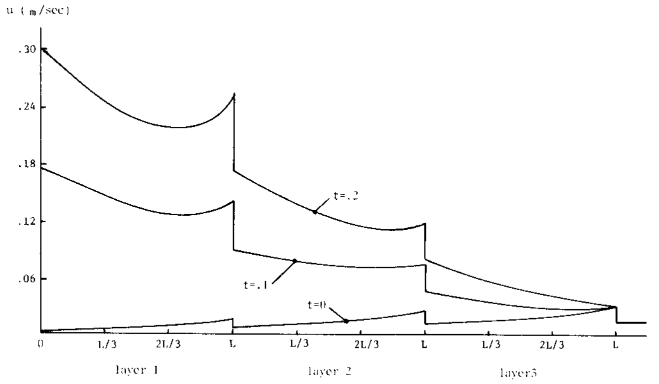


FIG. 12. Horizontal velocity for 3 layer model.

(ii) Using quasisteady analysis (principles 10, 14, and 15), various sets of quasisteady models were obtained. These models are similar in some respects to the usual incompressible equations, but at the same time show important differences.

The numerical calculations indicated the following.

(1) The models are well posed (this was seen through a perturbation analysis via a sequence of runs with decreasing mesh).

(2) The results are physically sensible. In particular, one saw displayed the mechanism by which the elastic wall converts the pulsatile inflow to a much more continuous outflow.

(3) The fluid sound speed, rather than a wave speed, was the pertinent parameter in the outflow boundary conditions. This results in a direct relationship between the pressure variation and the variation in the outflow velocity.

(4) It was seen that bifurcations had the expected effect in terms of decreasing pressure requirements.

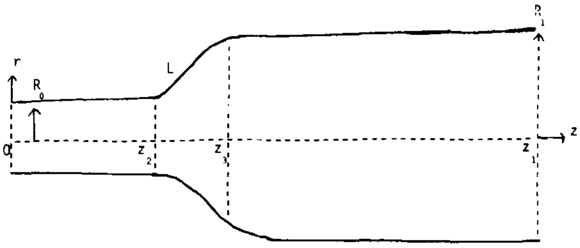
APPENDIX: BIFURCATION BOUNDARY CONDITIONS

The thought is to treat the bifurcation as a discontinuity over which "jump" conditions are to be applied. It is further assumed that the essential consideration is that of the increasing cross-sectional area. We will therefore derive conditions for a tube with an increasing cross-sectional area and apply these conditions to the bifurcation.

Let the cross section of the tube be

TABLE I. Selected results of the computations.

No. of layers	Variable distensibility	Period	$P_{\max}$	$P_{\min}$	$\Delta P = P_{\max} - P_{\min}$	$\Delta R_{\max}$	$\Delta R_{\min}$	$\Delta(\Delta R) = \Delta R_{\max} - \Delta R_{\min}$	max U_{out}	min U_{out}	av U_{out}	No. of cycles for steady state
1	no	1	752.9	675.9	77.0	0.003 911	0.003 542	0.000 369	0.0683	0.0613	0.0648	30
1	no	0.8	921.9	860.9	61.0	0.004 795	0.004 505	0.000 290	0.0836	0.0781	0.0809	45
1	no	0.5	1434	1407	26.6	0.007 471	0.007 352	0.000 119	0.1302	0.1277	0.1300	100
1	yes	1	1176	411.1	765.1	0.001 216	0.000 799	0.000 417	0.1066	0.0373	0.0648	6
3	no	1	221.5	199.2	22.3	0.001 149	0.001 042	0.000 107	0.0201	0.0180	0.0192	30
3	yes	1	251.5	178.2	73.3	0.000 631	0.000 530	0.000 101	0.0228	0.0162	0.0192	14



On L, we let $r = R(z)$ or $z = Z_1(r)$. Also let A denote the indicated area.

The following lemmas are easily established.

Lemma 1: Let $f(r,z)$ be continuously differentiable, and suppose $f(0,r) \equiv f_0$ and $f(z_1,r) \equiv f_1$. Then,

$$\int\int_A r\frac{\partial f}{\partial z}dA = \frac{R_1^2 f_1}{2} - \frac{R_1^2 f_0}{2} - \int_L rf dr.$$

Lemma 2: Let $f(r,z)$ be continuously differentiable, and suppose $f = 0$ for $(0 \leqslant z \leqslant z_2, r = R_0)$, $(z_3 \leqslant z \leqslant z_1, r = R_1)$, and $(0 \leqslant z \leqslant z_1, r = 0)$. Then, $\int\int_A f_r dA = \int_L f dz$.

We consider the following steady-state flow: $P \equiv P_0, u \equiv U_0, w \equiv 0$ for $0 \leqslant z \leqslant z_2$, and $P \equiv P_1, u \equiv U_1, w \equiv 0$ for $z_3 \leqslant z \leqslant z_1$. Note that on L, from Eq. (7.6)

$$wdz = udR \,. \tag{A1}$$

With $\rho_t \equiv 0$, the conservation of mass equation, Eq. (1.2), can be written as

$$(\rho wr)_r + r(\rho u)_z = 0\,.$$

Applying Lemma 1 and Lemma 2, we obtain

$$\rho_1 U_1 \frac{R_1^2}{2} - \frac{\rho_0 U_0 R_0^2}{2} + \int_L \rho wr dz - \int_L \rho ur dr = 0\,. \tag{A2}$$

From Eq. (A1), the integrals over L cancel. This leaves the expected result, $R_1^2 \rho_1 U_1 = R_0^2 \rho_0 U_0$. If we let A_0 and A_1 denote the inlet and outlet cross-sectional areas, we obtain $A_1 \rho_1 U_1 = A_0 \rho_0 U_0$. Finally, if we assume $\rho_0 = \rho_1$, then a further approximation is given by

$$A_1 U_1 = A_0 U_0 \,. \tag{A3}$$

From Eqs. (1.2) and (1.4), we obtain

$$r(\rho u)_t = -(\rho uwr)_r - r(\rho u^2 + P)_z \,. \tag{A4}$$

Assuming steady state, we obtain from the lemmas, and Eq. (A3),

$$\Delta P = P_1 - P_0 = \frac{2}{R_1^2}\int_L rPdr + \frac{A_1 - A_0}{A_1}\left(\frac{A_0 \rho}{A_1} U_0^2 - P_0\right). \tag{A5}$$

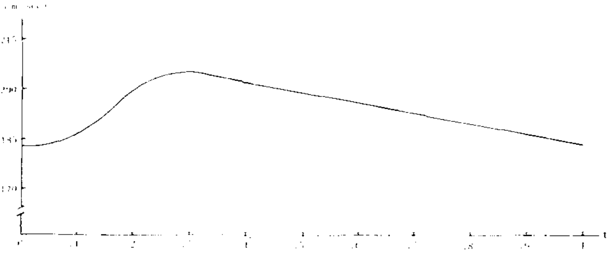


FIG. 13. Outlet velocity for 3 layer model.

The integral in Eq. (A5) cannot be evaluated without precise knowledge of P. However, extreme values can be found as follows.

(i) $P \equiv P_0 \Rightarrow \int_L rPdr = P_0 \dfrac{R_1^2 - R_0^2}{2}$ or

$$(\Delta P)_{\min} = \left(\frac{A_1 - A_0}{A_1}\right)(\rho U_0^2)\left(\frac{A_0}{A_1}\right), \tag{A6}$$

(ii) $P \equiv P_1 \Rightarrow \int_L rPdr = P_1\left(\dfrac{R_1^2 - R_0^2}{2}\right)$ or

$$(\Delta P)_{\max} = \left(\frac{A_1 - A_0}{A_1}\right)\rho U_0^2 \,. \tag{A7}$$

An average therefore might be

$$(\Delta P)_{\rm av} = \frac{(\Delta P)_{\min} + (\Delta P)_{\max}}{2} = \left(\frac{A_1^2 - A_0^2}{2A_1^2}\right)\rho U_0^2 \,. \tag{A8}$$

Remark: Equation (A8) is also the result obtained by integrating Eq. (1.4) on the centerline $r = 0$. Taking ρ to be constant consistent with Eq. (22.1), we have

$$(\Delta P)_{\rm av} = 3.750 U_0^2 \left(\frac{A_1^2 - A_0^2}{A_1^2}\right) \text{mm Hg}. \tag{A9}$$

[1] P. Gordon and S. M. Scala, "Nonlinear Theory of Pulsatile Blood Flow Through Viscoelastic Blood Vessels," in *Proceedings of the AGARD Specialists' Meeting on Fluid Dynamics of Blood Circulation and Respiratory Flow*, May 4–6, 1970, Naples, Italy (Technical Editing and Reproduction Ltd., London, 1970).

[2] P. Gordon, "Quasisteady Primitive Equations with Associated Upper Boundary Conditions," J. Math. Phys. **20**, 634–658 (1979).

[3] P. Gordon, "Lateral boundary conditions for quasisteady atmosphere flows," J. Math. Phys. **21**, 2612–2627 (1980) (Progress Report, Contract No. N00014-78-C-0303, NR 061-254).

[4] J. R. Womersley, "Oscillatory Motion of a Viscous Liquid in a Thin-Walled Elastic Tube I: The Linear Approximation for Long Waves," Philos. Mag. **46**, 199–221 (1955).

[5] D. A. McDonald, *Blood Flow in Arteries* (Williams and Wilkins, Baltimore, 1974), 2nd ed.

[6] M. H. Friedman, V. O'Brien, and L. W. Ehrlich, "Calculations of Pulsatile Flow Through a Branch—Implications for the Hemodynamics of Atherogenesis," Circ. Res. **36**, 277–285 (1975).

[7] S. C. Ling, H. B. Atabek, W. G. Letzing, and D. J. Patel, "Nonlinear Analysis of Aortic Flow in Living Dogs," Circ. Res. **33**, 198–212 (1973).

[8] R. Skalak, "Wave Propagation in Blood Flow," in *Biomechanics* (American Society of Mechanical Engineers, New York, 1966).

[9] James A. Maxwell and Max Anliker, "The Dissipation of Small Waves in Arteries and Veins with Viscoelastic Wall Properties," Biophys. J. **8**, 920–950 (1968).

[10] M. F. O'Rourke and A. P. Avolis, "Pulsatile Flow and Pressure in Human Systemic Arteries," Circ. Res. **46**, 363–372 (1980).

[11] J. C. Hunsaker and B. G. Rightmire, *Engineering Applications of Fluid Mechanics* (McGraw-Hill, New York, 1947).

[12] N. R. Kuchar and S. Ostrach, "Flows in Entrance Regions of Circular Elastic Tubes," in *Biomedical Fluid Mechanics Symposium* (ASME, New York, 1966), pp. 45–69.

[13] D. A. McDonald, "The Velocity Profiles of Pulsatile Flow" in *Flow Properties of Blood and Other Biological Systems* (Pergamon, London, 1960), pp. 84–96.

[14] E. W. Merrill and R. E. Wells, "Flow Properties of Biological Fluids," Appl. Mech. Rev. **14**, 663–673 (1961).

[15] S. M. Scala and P. Gordon, "Solution of the Time-Dependent Navier-Stokes Equations for the Flow Around a Circular Cylinder," AIAA Paper, 67-221. January 1967; AIAA J. **6**, 815–822 (1968); Comp. Fluid Dynamics **4**, 114-121 (1968).

[16] S. M. Scala and P. Gordon, "Solution of the Navier-Stokes Equations for Viscous Supersonic Flows Adjacent to Isothermal and Adiabatic Sur-

faces," G. E. Co., Space Sciences Laboratory, Document TIS 69SD1001, April 28, 1969; *Proceedings of the Symposium on Viscous Interaction Phenomena in Supersonic and Hypersonic Flow*, Wright-Patterson Air Force Base, Ohio, May 7–8, 1969 (University of Dayton, Dayton, Ohio, 1970).

[17] P. Gordon, "Stability of Solutions of the Compressible Navier-Stokes Equations," J. Math. Phys. **18**, 1543–1552 (1977).

[18] H. M. Lieberstein, *Mathematical Physiology, Blood Flow and Electrically Active Cells*" in *Modern Analytic and Computational Methods in Science and Mathematics*, No. 40, edited by R. Bellman (Elsevier, New York, 1973).

[19] L. Peterson, R. Jensen, and J. Parnell, "Mechanical Properties of Arteries in Vivo," Circ. Res. **8**, 622 (1960).

[20] E. Gams, L. L. Hunstman, and J. E. Chimoskey, "Correlation of Maximum Aortic and Carotid Flow Acceleration in Chronically Instrumented Dogs," Circ. Res. **34**, 302–308 (1974).

[21] S. S. Ling and H. B. Atabek, "Nonlinear Analysis of Pulsatile Flow in Arteries," J. Fluid Mech. **55**, 493–511 (1972).

[22] Robert F. Rushmer, *Cardiovascular Dynamics* (Saunders, Philadelphia, 1961).

Nonlinear periodic waves in a self-gravitating fluid

Dipankar Ray
Physics Department, Queen Mary College, Mile End Road, London E1 4NS, England

(Received 7 March 1980; accepted for publication 2 May 1981)

Liang has obtained equations for nonlinear plane waves in a Jeans universe. While Liang has studied the equation for a specific equation of state, this paper makes a general study. Among other things it is shown that only a subsonic propagation is possible.

PACS numbers: 98.80. − k, 47.75. + f

Liang[1] has studied one-dimensional perturbations to a homogeneous isotropic expanding universe in the following way. Let u_0 and ρ_0 be unperturbed velocity and density, and u and ρ be perturbed values of the same quantities. Taking the x direction as the direction of perturbation, v and D are defined by

$$\vec{u} = \vec{u}_0 + v(x,t)/R, \quad \rho = \rho_0 e^{D(x,t)}, \tag{1}$$

where t is time and R is the expansion coefficient. By specializing in Jeans universe where R, the expansion coefficient, can be set equal to 1, Liang has obtained the following equations:

$$[C_s^2 - U^2(1-w)^2]\frac{d^2w}{d\xi^2} + \frac{[U^2(1-w)^2 + C_s^2 + dC_s^2/dD]}{1-w}\left(\frac{dw}{d\xi}\right)^2 + 4\pi G\rho_0 w = 0 \tag{2a}$$

and

$$e^D = 1/(1-w), \tag{2b}$$

where

$$D = D(\xi), \quad v = Uw(\xi), \quad \xi = u - U\tau, \quad \tau = \int \frac{dt}{R^2}. \tag{3}$$

U is the constant phase velocity.

Equations (2) have been studied by Liang[1] for a specific functional dependence of C_s on D. In this paper we plan to make a general study.

It is obvious that, for physical solutions, w and $dw/d\xi$ have to be bounded. Further, if $C_s^2 - U^2(1-w)^2 = 0$ over a finite range of values of ξ then, using (1) and (2), we get $w = 0$ over that finite range of ξ, which corresponds to the no-wave case. If we leave that out,

$$C_s^2 - U^2(1-w)^2 \neq 0, \tag{4}$$

except possibly at isolated points. Multiplying (2a) by $C_s^2 - U^2(1-w)^2$ and integrating, Eqs. (2) can be replaced by

$$F^2\left(\frac{dw}{d\xi}\right)^2 + V = 0 \tag{5}$$

provided (4) holds,

$$F(w) \equiv F \equiv \frac{C_s^2}{1-w} - U^2(1-w), \tag{6a}$$

and $V = V(w)$ is defined by

$$\frac{dV(w)}{dw} = \frac{8\pi G\rho_0 w}{1-w}F(w). \tag{6b}$$

Since w is bounded, let w_1 and w_2 be, respectively, the lower and upper bounds of w; then

$$\frac{dw}{d\xi} = 0 \quad \text{at} \quad w = w_1 \quad \text{and} \quad w_2. \tag{7}$$

Also, from (1) and (2b), the condition that ρ must be finite and positive requires

$$w_2 < 1. \tag{8}$$

From (6a) and (8), F is finite and so from (7)

$$V(w_1) = 0, \quad V(w_2) = 0. \tag{9}$$

Also from (5)

$$V(w) \leqslant 0 \quad \text{for} \quad w_1, w, w_2. \tag{10}$$

From (7), by Roll's theorem, there exists w_3 such that

$$w_1 < w_3 < w_2 \tag{11}$$

and $dV/dw = 0$ for $w = w_3$. Therefore, from (6b) either

$$w_3 = 0,$$

or

$$F(w_3) = 0. \tag{12}$$

However, we shall see that

$$F(w) \neq 0 \quad \text{for} \quad w_1 < w < w_2. \tag{13}$$

This can be seen as follows: From (4) and (5) and the finiteness of $dw/d\xi$,

$$v(w) \neq 0 \tag{14}$$

in (w_1, w_2) except possibly at isolated points. Now if possible let $F(w_0) = 0$ where w_0 is a particular point in (w_1, w_2). Then from (5) and the finiteness of $dw/d\xi$

$$V(w_0) = 0.$$

Then, in view of (9), (10), and (13), there exists at least one minimum for $V(w)$ between w_1 and w_0 and at least one minimum between w_0 and w_2.

Therefore, there exists w_4 and w_5 such that $dV/dw = 0$ and $V < 0$ for $w = w_4$, w_5 and $w_1 < w_4 < w_0 < w_5 < w_2$. Then, from (6), $w_4 F(w_4) = 0 = w_5 F(w_5)$. Since both w_4 and w_5 cannot be zero, either $F(w_4) = 0$ or $F(w_5) = 0$. Therefore, in view of (5) and the finiteness of $dw/d\xi$, either $V(w_4) = 0$ or $V(w_5) = 0$, which violates previous assumptions.

Therefore, (13) must hold and from (12) and (13) we get $w_3 = 0$.

Therefore, from (8) and (12),

$$w_1 < 0 < w_2 < 1. \tag{15}$$

In view of (6), (13), and (15), V has one, and only one, extremum in (w_1,w_2) and that is at $w = 0$. Owing to (10), this can only be a minimum, which, owing to (16), requires that

$$C_s^2 - U^2(1 - w)^2 > 0 \tag{16}$$

for all w such that $w_1 < w < w_2$.

At this stage we have seen that (9), (15), and (16) are necessary conditions for physically acceptable solutions of (2) or equivalently that of (5) subject to (4). To show that they are also sufficient, one has to show that (9), (15), and (16) together, automatically satisfy (10). For this we note from (6), (15), and (16) that V is monotonically decreasing between w and 0 and monotonically increasing between 0 and w_2. Equation (10) then follows from Eq. (9).

Therefore, summarily with F and V defined through (6), necessary and sufficient conditions that solutions of (2) give bounded and positive ρ and bounded w and $dw/d\xi$ are given by (9), (15), and (16). Also (15) and (16) lead to

$$C_0^2 - U^2 > 0,$$

where $C_0 = C_s$ for $\rho = \rho_0$, i.e., $w = 0$.

In other words, only a subsonic wave is possible. This result was proved by Liang for a specific equation of state, but is proved in general here.

---

[1] E. P. T. Liang, Astrophys. J. **230**, 325 (1979).

Erratum: Reduction of inner-product representations of unitary groups [J. Math. Phys. 24, 233 (1983)]

R. S. Nikam, K. V. Dinesha, and C. R. Sarma
Department of Physics, Indian Institute of Technology, Bombay-400 076 India

(Received 24 February 1983; accepted for publication 25 February 1983)

PACS numbers: 02.20.Qs, 99.10. + g

The names of the three authors were inadvertently omitted from beneath the title of the first page of their article.

Erratum: Geometrical perturbation theory: action-principle surface terms in homogeneous cosmology [J. Math. Phys. 23, 2151 (1982)]

Robert H. Gowdy
Department of Physics, Virginia Commonwealth University, Richmond, Virginia 23284

(Received 8 December 1982; accepted for publication 17 December 1982)

PACS numbers: 04.20.Cv, 02.40. – k, 98.80.Dr, 99.10. + g

Equation (3.4) omits an important term which invalidates the proposed correction to the action. The elimination of the surface terms from the action principle by imposing a gauge condition remains valid but is complicated by the additional terms.

Erratum: Regge trajectories in confining potentials [J. Math. Phys. 23, 665 (1982)]

D. P. Datta and S. Mukherjee
Department of Physics, North Bengal University, Raja-Rammohunpur, Darjeeling-734430, India

(Received 7 June 1982; accepted for publication 1 October 1982)

PACS numbers: 11.60. + c, 14.80.Dq, 12.35.Ht, 99.10. + g

Although the proof of the analyticity of $E_n(l)$ given in Sec. II is invalid since the conditions (11) do not hold simultaneously for $\phi = \pi/3$, the analyticity follows for potentials (4) even for a greater range of β, $0 < \beta < \alpha$. Note that the asymptotic behavior of the wavefunction $u_n(z)$ is

$$u_n/u_0 \rightarrow 1,$$

$$u_0 = z^{-\alpha/4}\exp\left\{-\frac{z^{\alpha/2+1}}{\alpha/2+1} - \frac{k}{2}\frac{z^{\beta-\alpha/2+1}}{\beta-\alpha/2+1} + \text{lower order terms}\right\}$$

as $z \rightarrow \infty$. Thus $u_n(z)$ is square-integrable in the sector $|\arg z| \leqslant \pi/(2+\alpha)$ of the complex z plane. An equation similar to Eq. (12), therefore, can be used to prove the constancy of the number of zeros of u_n in $|\phi| \leqslant \pi/(2+\alpha)$ for any complex λ, and hence the result follows.

In the proof of Lemma 3.2 infimum should be replaced by supremum.

We are thankful to A. Martin for his comments in this connection.

0022-2488/83/041013-01$02.50