The eigenvalue problem for "arrow" matrices

O. Walter, L. S. Cederbaum, and J. Schirmer
Lehrstuhl für Theoretische Chemie, Institut für Physikalische Chemie, Universität Heidelberg, D-6900 Heidelberg, West Germany

(Received 24 June 1982; accepted for publication 3 June 1983)

The eigenvalue problem for a particular class of "arrow" matrices $\mathbf{Z} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \gamma\mathbf{B}^\dagger & \mathbf{C} \end{pmatrix}$, where $\mathbf{A}$ is a Hermitian $N \times N$ matrix, $\mathbf{C}$ a real diagonal $M \times M$ matrix, $\mathbf{B}$ an arbitrary complex $N \times M$ matrix, and γ a real number, is investigated by means of a partitioning technique. Both Hermitian ($\gamma = 1$) and non-Hermitian ($\gamma \neq 1$) arrow matrices $\mathbf{Z}$ are studied. The one-dimensional case (dimension N of $\mathbf{A}$ equal to 1) is briefly reviewed and a detailed treatment of the multidimensional case ($N > 1$) is presented. For Hermitian arrow matrices, the analysis leads to a new algorithm for computing the eigenvalues and eigenvectors of $\mathbf{Z}$ which is particularly efficient if $M \gg N$.

PACS numbers: 02.10.Sp, 02.70. + d

I. INTRODUCTION

Many physical problems can be described in terms of an eigenvalue problem for "arrow" matrices $\mathbf{Z}$ which have the form

$$\mathbf{Z} = \begin{pmatrix} a_{11} & \cdots & a_{1N} & b_{11} & \cdots & b_{1M} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{1N}^* & \cdots & a_{NN} & b_{N1} & \cdots & b_{NM} \\ \gamma b_{11}^* & \cdots & \gamma b_{N1}^* & c_1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \gamma b_{1M}^* & \cdots & \gamma b_{NM}^* & 0 & \cdots & c_M \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \gamma\mathbf{B}^\dagger & \mathbf{C} \end{pmatrix}. \tag{1.1}$$

A diagonal real matrix $\mathbf{C}$ is coupled to a Hermitian matrix $\mathbf{A}$ by $N \times M$, in general complex, matrix elements b_{ij}. Without loss of generality, one may take the c_m to be ordered such that $c_1 \leqslant c_2 \leqslant \cdots \leqslant c_M$. The real number γ accounts for a special class of non-Hermitian arrow matrices. Setting $\gamma = 1$, one obtains a Hermitian matrix of order $N + M$. Such Hermitian arrow matrices are studied, e.g., in perturbation theory for determining the changes of the unperturbed energies caused by adding additional states to the Hamiltonian matrix.[1,2] For finite Fermi systems like atoms or molecules, Hermitian matrices of the form of Eq. (1.1) are employed for calculating ionization energies, electron affinities, and relative intensities via a Green's function approach.[3–6] Setting $\gamma = -1$, one obtains a non-Hermitian arrow matrix which, e.g., has to be dealt with when studying the spectral distribution of scattered light.[7] Here, the number of complex eigenvalues determines the maximum number of frequency-shifted lines in the spectrum of light scattered from ordinary fluids.

A convenient approach for studying the eigenvalue problem of such arrow matrices is to employ a partitioning technique. To this purpose, one considers as a function of the variable ω the $N \times N$ matrix

$$\mathbf{G}(\omega) = [\omega\mathbf{1} - \mathbf{L}(\omega)]^{-1}, \tag{1.2}$$

where the matrix elements $L_{ij}(\omega)$ of the matrix $\mathbf{L}(\omega)$ are given by

$$L_{ij}(\omega) = a_{ij} + \gamma \sum_{m=1}^{M} \frac{b_{im} b_{jm}^*}{\omega - c_m}. \tag{1.3}$$

$\mathbf{G}(\omega)$ will be referred to as the Green's function associated with $\mathbf{Z}$. According to Eq. (1.1), one immediately identifies $\mathbf{G}(\omega)$ with the principal matrix (i.e., the block corresponding to the submatrix $\mathbf{A}$) of $[\omega\mathbf{1} - \mathbf{Z}]^{-1}$,

$$\mathbf{G}(\omega) = ([\omega\mathbf{1} - \mathbf{Z}]^{-1})_A. \tag{1.4}$$

From Eq. (1.4), it is seen that the poles g_i of $\mathbf{G}(\omega)$ coincide with the eigenvalues of $\mathbf{Z}$. For Hermitian arrow matrices ($\gamma = 1$), the eigenvalue problem is formulated as

$$\mathbf{Z} = \mathbf{X}\mathbf{\Gamma}\mathbf{X}^\dagger, \tag{1.5a}$$

$$\Gamma_{ij} = g_i \delta_{ij}. \tag{1.5b}$$

Here, the matrix $\mathbf{X}$ denotes the unitary transformation that diagonalizes $\mathbf{Z}$; Eq. (1.5a) implies that the residues of a particular pole g_p are linked to the eigenvector $\mathbf{X}_p$ of the corresponding eigenvalue of $\mathbf{Z}$ via the spectral representation of the Green's function

$$G_{ij}(\omega) = \sum_{p=1}^{N+M} \frac{x_{ip} x_{jp}^*}{\omega - g_p}, \tag{1.6}$$

where x_{ip}, $1 \leqslant i \leqslant N$, denote the first N components of $\mathbf{X}_p$. Note that an eigenvalue of $\mathbf{Z}$ does not occur as a pole of $\mathbf{G}(\omega)$ when the first N components of its eigenvector are zero.

Since Eq. (1.4) holds for all values of γ, the connection between the poles of $\mathbf{G}(\omega)$ and the eigenvalues of $\mathbf{Z}$ is not restricted to the Hermitian case. Equation (1.3) can thus be used to show that, for $\gamma > 0$, the eigenvalue problem can be replaced by a Hermitian eigenvalue problem for a matrix $\mathbf{Z}$, where the coupling matrix elements b_{ij} are replaced by $\tilde{b}_{ij} = \sqrt{\gamma} b_{ij}$. Similarly, the case $\gamma < 0$ can be reduced to a special non-Hermitian $\mathbf{Z}$ where $\gamma = -1$. Therefore in the following, we will consider only the two values of γ: $\gamma = 1$ (Hermitian $\mathbf{Z}$) and $\gamma = -1$ (non-Hermitian $\mathbf{Z}$). For the non-Hermitian case, the eigenvalue problem reads

$$\begin{aligned} \mathbf{Z}\mathbf{X}^R &= \mathbf{X}^R\mathbf{\Gamma}, \\ \mathbf{X}^L\mathbf{Z} &= \mathbf{\Gamma}\mathbf{X}^L, \end{aligned} \tag{1.7}$$

where $\mathbf{X}^R$ ($\mathbf{X}^L$) denotes the matrix of the right (left) eigenvectors. The eigenvalues g_p of $\mathbf{Z}$ and, hence, the poles of $\mathbf{G}(\omega)$ may now either be real or complex. Provided that $\mathbf{X}^L\mathbf{X}^R = \mathbf{1}$, the spectral representation of $\mathbf{G}(\omega)$ [Eq. (1.6)] still holds if one replaces $x_{ip}x_{jp}^*$ by $x_{ip}^R x_{pj}^L$.

The partitioning technique has the essential advantage

 0022-2488/84/040729-09$02.50

that the eigenvalues of an $(N+M)$-dimensional matrix can be studied by investigating the $N\times N$ matrix $\mathbf{G}(\omega)$. Here, the special case of $N=1$ is particularly simple. This one-dimensional case has been considered by many authors, see, e.g., Refs. 1, 7, and 8. For Hermitian arrow matrices, all eigenvalues can be evaluated by a straightforward "graphical" algorithm. When $\gamma=-1$, a similar procedure may be applied, however, restricted to real eigenvalues. In the general case where the principal submatrix $\mathbf{A}$ is of dimension $N>1$, no algorithm similar to the one worked out for the one-dimensional case seems to be available. For the Hermitian case, a rather qualitative discussion of an extension of the graphical procedure has been presented,[9] and for non-Hermitian $(\gamma=-1)$ arrow matrices, the maximum number of complex eigenvalues has been determined.[10]

The special structure of the arrow matrices considered here can also be exploited by the conventional technique of triangular factorization, see, e.g., Ref. 11. For a Hermitian arrow matrix $\mathbf{Z}$, one may start from the partitioned factorization of $\omega\mathbf{1}-\mathbf{Z}$:

$$\omega\mathbf{1}-\mathbf{Z}=\mathbf{H}(\omega)\mathbf{D}(\omega)\mathbf{H}^{\dagger}(\omega), \tag{1.8a}$$

where

$$\mathbf{H}(\omega)=\begin{pmatrix}\mathbf{1} & -\mathbf{B}(\omega\mathbf{1}-\mathbf{C})^{-1}\\ \mathbf{0} & \mathbf{1}\end{pmatrix},$$
$$\mathbf{D}(\omega)=\begin{pmatrix}\omega\mathbf{1}-\mathbf{L}(\omega) & \mathbf{0}\\ \mathbf{0} & \omega\mathbf{1}-\mathbf{C}\end{pmatrix}. \tag{1.8b}$$

Thus the complete factorization for a value of $\omega\neq c_m$ reduces to the factorization of $\omega\mathbf{1}-\mathbf{L}(\omega)$. Using Sylvester's inertia theorem,[11] the number of eigenvalues of $\mathbf{Z}$ in a specific interval (a,b), where $a,b\neq c_m$, can be obtained by performing the triangular factorization for $\omega=a$ and $\omega=b$. The eigenvalues of $\mathbf{Z}$ may then be calculated by implementing a standard algorithm such as a bisection or secant iteration procedure,[11] which requires the above factorization at a set of iteration points ω_ν.

The main theoretical drawback of any algorithm based on the factorization of Eq. (1.8), however, is the fact that the matrix $\mathbf{L}(\omega)$ is singular for $\omega=c_m$. Thus for ω close to c_m, the numerical evaluation of the matrix elements $L_{ij}(\omega)$ may suffer from severe rounding errors. In this paper, we present an alternative approach which circumvents such difficulties. By investigating the eigenvalues of $\mathbf{L}(\omega)$ as a function of ω, it is possible to describe properly the singular behavior at $\omega=c_m$. This analysis leads to a simple method for counting the number of eigenvalues of Hermitian arrow matrices in the intervals bounded by two successive matrix elements c_m and c_{m+1} (including $\pm\infty$). Moreover, we will introduce an efficient algorithm for evaluating all eigenvalues and eigenvectors of $\mathbf{Z}$ which can be considered as the multidimensional analog to the "graphical" algorithm for the one-dimensional case.

In Sec. II the one-dimensional case is briefly discussed. The treatment of the multidimensional case is presented in Sec. III. In the last section, the "graphical" algorithm for the Hermitian eigenvalue problem is summarized and its feasibility for numerical applications is briefly discussed.

II. THE ONE-DIMENSIONAL CASE

As an illustrative example for the general procedure, let us first consider the one-dimensional case where the dimension of $\mathbf{A}$ is equal to $N=1$. In that case, the matrix $\mathbf{Z}$ has the form

$$\mathbf{Z}=\begin{pmatrix} a & b_1 & b_2 & b_3 & \cdots\\ \gamma b_1^* & c_1 & 0 & 0 & \cdots\\ \gamma b_2^* & 0 & c_2 & 0 & \cdots\\ \gamma b_3^* & 0 & 0 & c_3 & \cdots\\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}, \tag{2.1}$$

and the corresponding one-dimensional Green's function reads

$$G(\omega)=\frac{1}{\omega-l(\omega)}, \tag{2.2a}$$
$$l(\omega)=a+\gamma\sum_{m=1}^{M}\frac{|b_m|^2}{\omega-c_m}. \tag{2.2b}$$

The analytical properties of this Green's function $G(\omega)$ are well known for both $\gamma=1$[1,8] and $\gamma=-1$.[7] Here, we restrict ourselves to a brief summary of the essential results which will be supplemented by a few remarks concerning non-Hermitian arrow matrices.

The eigenvalues of $\mathbf{Z}$ are obtained as the solutions g_p, $p=1,\ldots,M+1$ of the implicit equation

$$\omega-l(\omega)=0. \tag{2.3}$$

The real zeros of $\omega-l(\omega)$ can easily be found by a graphical analysis. Let us first consider the Hermitian case $\gamma=1$ and assume that all c_m are different. In this case, the eigenvalues g_p are real and distinct; Fig. 1 shows a graph of $l(\omega)$ for a specific set of parameters a, b_m, and c_m. The poles of $l(\omega)$ are indicated by vertical dotted lines, and the intervals $(-\infty, c_1), (c_1, c_2), \ldots, (c_M, \infty)$ are specified by $h=1,2,\ldots,M+1$. As can be seen from this graph, the solutions g_p of Eq. (2.3) are obtained as the values of ω at the intersections of $l(\omega)$ with the dashed line $y=\omega$. Since $l(\omega)$ is monotonically decreasing in each interval h, there is one and only one eigenvalue of $\mathbf{Z}$ in

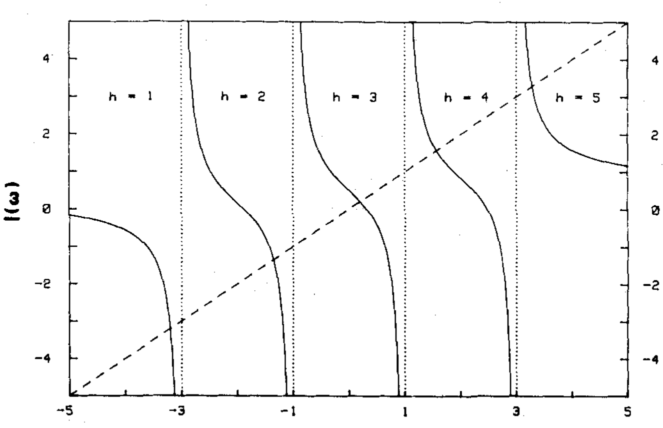


FIG. 1. Graphical determination of the eigenvalues of a Hermitian arrow matrix $\mathbf{Z}$ for the one-dimensional case $N=1$ $(M=4)$. The full lines represent the function $l(\omega)$, the poles of $l(\omega)$ are indicated by vertical dotted lines. The eigenvalues of $\mathbf{Z}$ are obtained as the values of ω at the intersections of $l(\omega)$ with the dashed line $y=\omega$.

each of these intervals. This property is often referred to as the separation theorem.

For $\gamma = -1$, the solutions of Eq. (2.3) are no longer restricted to real values of ω. However, all the real eigenvalues may still be obtained by the graphical method described above. Due to the sign of γ, the function $l(\omega)$ is now monotonically increasing in each interval h. In Fig. 2, the function $l(\omega)$ is plotted for a similar set of parameters as used for Fig. 1. Since $l(\omega)$ has M poles, there are at least $(M-1)$ intersections with the straight line $y=\omega$ in the interval (c_1, c_M), i.e., at least $(M-1)$ roots are real. If the remaining two eigenvalues are complex, there is exactly one eigenvalue of $\mathbf{Z}$ in each of the $(M-1)$ intervals $h = 2, ..., M$. Since the trace of $\mathbf{Z}$ is real and the real eigenvalues are nondegenerate, the complex eigenvalues are conjugate to each other. It should be noted that once the real eigenvalues are known, the complex solutions may directly be calculated by determining the determinant and the trace of $\mathbf{Z}$. If, on the other hand, all eigenvalues are real, then there is either one interval $h = 2, 3, ..., M$ bracketing three roots, which is the case in the example depicted in Fig. 2, or there are two additional solutions in the interval $(-\infty, c_1)$ or (c_M, ∞). In contrast to the Hermitian case, these solutions can be threefold degenerate in the former, and twofold degenerate in the latter case.

Let us now consider the case of a K-fold degeneracy of the diagonal matrix elements of $\mathbf{C}$, say $c_m = c_{m+1} = \cdots = c_{m+K-1}$. From Eq. (2.2b) it follows that this is equivalent to replacing $\mathbf{Z}$ by $\tilde{\mathbf{Z}}$, where the K coupling matrix elements $b_m, ..., b_{m+K-1}$ are replaced by modified quantities $\tilde{b}_m, ..., \tilde{b}_{m+K-1}$:

$$\begin{aligned} \tilde{b}_m &= \sqrt{\textstyle\sum_{i=0}^{K-1} |b_{m+i}|^2}, \\ \tilde{b}_{m+1} &= \tilde{b}_{m+2} = \cdots = \tilde{b}_{m+K-1} = 0. \end{aligned} \tag{2.4}$$

According to Eq. (2.4), $(K-1)$ diagonal elements of $\tilde{\mathbf{Z}}$ decouple, and one immediately sees that c_m is a $(K-1)$-fold degenerate eigenvalue of $\mathbf{Z}$. The residues of the corresponding pole of the Green's function vanish and these $(K-1)$-fold degenerate eigenvalues $g_p = c_m$ do not occur as poles of $G(\omega)$.

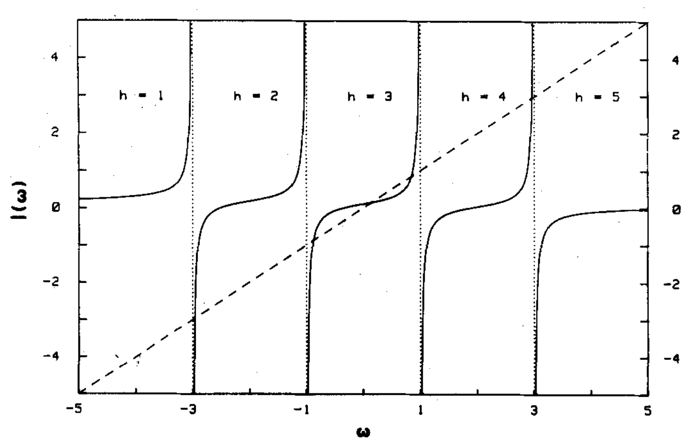


FIG. 2. Graphical determination of the real eigenvalues of a non-Hermitian arrow matrix $\mathbf{Z}$ for the one-dimensional case $N=1$ $(M=4)$. For an explanation of the symbols, see Fig. 1. Since there are three intersection points in the interval $h=3$, all eigenvalues of $\mathbf{Z}$ are real.

Finally, we turn to the problem of determining the eigenvector matrix. For a nondegenerate eigenvalue g_p, the corresponding right (left) eigenvector $\mathbf{X}_p^R$ $(\mathbf{X}_p^L)$ with elements x_{ip}^R (x_{pi}^L) is obtained as

$$\begin{aligned} x_{i+1,p}^R &= \gamma b_i^* x_{1p}^R/(g_p - c_i), \\ x_{p,i+1}^L &= b_i x_{p1}^L/(g_p - c_i), \end{aligned} \qquad 1 \leqslant i \leqslant M. \tag{2.5}$$

The normalizing condition $\mathbf{X}_p^L \mathbf{X}_p^R = 1$ fixes the remaining first eigenvector component to be

$$x_{1p}^R = x_{p1}^L = \sqrt{R_p}, \tag{2.6a}$$

where

$$R_p = \left[1 + \gamma \sum_{m=1}^{M} \frac{|b_m|^2}{(g_p - c_m)^2}\right]^{-1}. \tag{2.6b}$$

For Hermitian $\mathbf{Z}$, Eq. (2.5) reduces to the usual relation $\mathbf{X}_p^R = (\mathbf{X}_p^L)^\dagger$. The quantity R_p can easily be identified with the residue for $G(\omega)$ at the first-order pole $\omega = g_p$:

$$R_p = \left[1 - \frac{\partial}{\partial\omega} l(\omega)\Big|_{\omega=g_p}\right]^{-1}. \tag{2.7}$$

Equation (2.7), on the other hand, shows that the residues of the Green's function are directly related to the slope of $l(\omega)$ at its intersection points with the straight line $y=\omega$.

If g_p is a degenerate eigenvalue of $\mathbf{Z}$, one must distinguish between two cases. The first case is that $g_p = c_m$ is a $(K-1)$-fold degenerate eigenvalue resulting from a K-fold degeneracy of the matrix elements c_m. In this case, there are always $(K-1)$ linearly independent eigenvectors, and the K nonvanishing components $m+1, ..., m+K$ of these eigenvectors are obtained as the orthogonal complement of the vector $(b_m, ..., b_{m+K-1})$. The other possibility is a two- or threefold degeneracy of $g_p \neq c_m$ which only occurs for non-Hermitian arrow matrices $(\gamma = -1)$. In that case, Eq. (2.5) yields only one linearly independent right (left) eigenvector $\mathbf{X}_p^R$ $(\mathbf{X}_p^L)$ and no orthogonal transformation exists which brings $\mathbf{Z}$ to diagonal form. The spectral representation of $\mathbf{G}(\omega)$, Eq. (1.6), must then be modified by introducing a pole of second or third order. From Eq. (2.2), it follows that this additional term can always be written as

$$G(\omega) = \sum_{q \neq p} \frac{R_q}{\omega - g_q} + \frac{\alpha(\omega - g_p) + \beta}{(\omega - g_p)^2} \tag{2.8}$$

if we restrict ourselves to a second-order pole. The constant β is obtained as

$$\beta^{-1} = \sum_{m=1}^{M} \frac{|b_m|^2}{(g_p - c_m)^3} = -\frac{1}{2}\frac{\partial^2}{\partial\omega^2} l(\omega)\Big|_{\omega=g_p}, \tag{2.9a}$$

and for α, one finds, from the asymptotic behavior of $G(\omega)$, that

$$\alpha + \sum_{q \neq p} R_q = 1. \tag{2.9b}$$

We also have a simple means to tell whether such a doubly degenerate eigenvalue will occur. From Eq. (2.7), it follows that if g_p is twofold degenerate, the two relations

$$\left.\frac{\partial}{\partial\omega} l(\omega)\right|_{\omega=g_p} = 1,$$
$$\left.\frac{\partial^2}{\partial\omega^2} l(\omega)\right|_{\omega=g_p} \neq 0 \tag{2.10}$$

must be satisfied. Note that for a third-order pole, similar results hold.

III. THE MULTIDIMENSIONAL CASE

A. The Hermitian case for $N > 1$

For the general case $N > 1$, the Hermitian arrow matrix $\mathbf{Z}$ is of order $N + M$, and the associated Green's function is a matrix of order N. Since the multidimensional analog of Eq. (2.3),

$$\omega\mathbf{1} - \mathbf{L}(\omega) \equiv \mathbf{0}, \tag{3.1}$$

is now a matrix equation, the situation is more involved. In order to determine the poles of $\mathbf{G}(\omega)$, one may consider the set of N eigenfunctions $l_q(\omega)$ which are defined by diagonalizing $\mathbf{L}(\omega)$ at every value of ω. The poles of $\mathbf{G}(\omega)$ are thus obtained by the solutions of the set of equations

$$\omega - l_q(\omega) = 0, \quad 1 \leqslant q \leqslant N. \tag{3.2}$$

In the following, we restrict ourselves to the case that all matrix elements c_m are different; the general case including degeneracies of the diagonal matrix elements of $\mathbf{C}$ will be treated in Sec. IIIB.

By definition, the matrix elements of $\mathbf{L}(\omega)$ are continuous for all values of $\omega \neq c_m$, which implies that the eigenfunctions $l_q(\omega)$ are also continuous for $\omega \neq c_m$. To examine the functions $l_q(\omega)$ at the poles of $\mathbf{L}(\omega)$, let us now consider the analytic properties of the eigenfunctions at a particular pole, say c_k. For $c_{k-1} < \omega < c_{k+1}$, we divide $\mathbf{L}(\omega)$ into two parts:

$$\mathbf{L}(\omega) = \tilde{\mathbf{L}}(\omega) + (1/(\omega - c_k))\mathbf{U}_k,$$
$$\mathbf{U}_k = \mathbf{B}_k \mathbf{B}_k{}^{\dagger}. \tag{3.3}$$

Here, $\mathbf{B}_k$ denotes the vector of the coupling matrix elements b_{ik} for the pole c_k, and $\tilde{\mathbf{L}}(\omega)$ is defined by Eq. (1.3) with the restriction $m \neq k$ in the summation. This decomposition ensures that $\tilde{\mathbf{L}}(\omega)$ is continuous at $\omega = c_k$. The pole part is given as the product of $(\omega - c_k)^{-1}$ and the Hermitian matrix $\mathbf{U}_k$. The eigenvalues $\beta_i^{(k)}$ of $\mathbf{U}_k$ are easily determined as

$$\beta_1^{(k)} = \sum_{n=1}^{N} |b_{nk}|^2,$$
$$\beta_2^{(k)} = \beta_3^{(k)} = \cdots = \beta_N^{(k)} = 0. \tag{3.4}$$

The eigenvector components of the nonvanishing eigenvalue $\beta_1^{(k)}$ are given by

$$S_{i1}^{(k)} = \left[\sum_{j=1}^{N} |b_{jk}|^2\right]^{-1/2} b_{ik}, \tag{3.5}$$

whereas the remaining $(N-1)$ eigenvectors corresponding to the vanishing eigenvalues $\beta_2^{(k)}, \ldots, \beta_N^{(k)}$ span the orthogonal complement of $\mathbf{S}_1^{(k)}$. Transforming $\mathbf{L}(\omega)$ by means of such an eigenvector matrix $\mathbf{S}$, one obtains

$$\mathbf{S}^{\dagger}\mathbf{L}(\omega)\mathbf{S} = \begin{pmatrix} Q_{11}(\omega) + \beta_1^{(k)}/(\omega - c_k) & \mathbf{F}(\omega) \\ \mathbf{F}^{\dagger}(\omega) & \mathbf{D}(\omega) \end{pmatrix}, \tag{3.6a}$$

$$F_{ij}(\omega) = Q_{ij}(\omega), \quad i = 1, \quad j = 2,3,\ldots,N,$$
$$D_{ij}(\omega) = Q_{ij}(\omega), \quad i,j = 2,3,\ldots,N, \tag{3.6b}$$

$$Q_{ij}(\omega) = [\mathbf{S}^{\dagger}\tilde{\mathbf{L}}(\omega)\mathbf{S}]_{ij}, \tag{3.6c}$$

where only the $(1,1)$ element of $\mathbf{S}^{\dagger}\mathbf{L}(\omega)\mathbf{S}$ diverges for $\omega \to c_k$. Now, it is easily seen that only one eigenfunction $l_q(\omega)$ has a (simple) pole and the analytic form of this particular $l_q(\omega)$ in the vicinity of c_k is given as

$$l_q(\omega) \underset{\omega \to c_k}{\simeq} \frac{\beta_1^{(k)}}{\omega - c_k} + Q_{11}(c_k). \tag{3.7}$$

Equation (3.6) also shows how the remaining $(N-1)$ finite eigenvalues of $\mathbf{L}(\omega)$ at $\omega = c_k$ can be determined. Since, for $c_{k-1} < \omega < c_{k+1}$, the matrix elements $Q_{ij}(\omega)$ of the submatrix $\mathbf{D}(\omega)$ are continuous by construction, we can introduce the ω-dependent unitary transformation $\boldsymbol{\Upsilon}(\omega)$ which diagonalizes $\mathbf{D}(\omega)$:

$$\mathbf{D}(\omega) = \boldsymbol{\Upsilon}(\omega)\boldsymbol{\Delta}(\omega)\boldsymbol{\Upsilon}^{\dagger}(\omega),$$
$$\Delta_{ij}(\omega) = d_i(\omega)\delta_{ij}. \tag{3.8}$$

In the limit $\omega \to c_k$, the diverging matrix element $[\mathbf{S}^{\dagger}\mathbf{L}(\omega)\mathbf{S}]_{11}$ decouples from $\mathbf{D}(\omega)$ and the $(N-1)$ finite eigenvalues of $\mathbf{L}(\omega)$ are directly obtained as the eigenvalues $d_i(c_k)$ of $\mathbf{D}(c_k)$. This means that $(N-1)$ eigenfunctions are regular at $\omega = c_k$ and the corresponding values of $l_q(c_k)$ are given as the eigenvalues $d_i(c_k)$. Moreover, the limit on the right and the limit on the left for $\omega \to c_k$ are equal, i.e., the $(N-1)$ regular functions are continuous at c_k. These analytical properties of the functions $l_q(\omega)$ at the poles of $\mathbf{L}(\omega)$ are summarized in the following lemma.

Lemma 1: Let all poles c_m of $\mathbf{L}(\omega)$ be different. It follows that at each pole c_k, one and only one eigenfunction $l_q(\omega)$ has a simple pole. The analytic form of this particular $l_q(\omega)$ in the vicinity of c_k is given by Eq. (3.7). The remaining $(N-1)$ functions are continuous at $\omega = c_k$ and the corresponding values $l_q(c_k)$ are obtained as the eigenvalues of the $(N-1)$-dimensional submatrix $\mathbf{D}(c_k)$ (Eq. 3.6).

Lemma 1 implies that each function $l_q(\omega)$ has I_q poles at the positions $\omega_p = c_{i_1}, \ldots, c_{i_{I_q}}$. In order to study the behavior of the eigenfunctions in the pole-free domains, we consider the explicit form of the function $l_q(\omega)$ for values of $\omega \neq c_m$:

$$l_q(\omega) = [\mathbf{Y}^{\dagger}(\omega)\mathbf{L}(\omega)\mathbf{Y}(\omega)]_{qq},$$
$$\mathbf{Y}(\omega)^{\dagger}\mathbf{Y}(\omega) = \mathbf{1}. \tag{3.9}$$

Making use of the unitarity of $\mathbf{Y}(\omega)$, if follows that the first derivative of $l_q(\omega)$ with respect to ω is given by

$$\frac{\partial}{\partial\omega} l_q(\omega) = \left[\mathbf{Y}^{\dagger}(\omega)\left(\frac{\partial}{\partial\omega}\mathbf{L}(\omega)\right)\mathbf{Y}(\omega)\right]_{qq}. \tag{3.10}$$

Inserting the explicit expression for $\mathbf{L}(\omega)$ [Eq. (1.3)], one obtains the quadratic form

$$\frac{\partial}{\partial\omega} l_q(\omega) = -\gamma \sum_{m=1}^{M} \frac{1}{(\omega - c_m)^2} \left| \sum_{n=1}^{N} y_{nq}(\omega) b_{nm} \right|^2, \quad (3.11)$$

where $y_{nq}(\omega)$ denotes the matrix elements of $\mathbf{Y}(\omega)$. Equation (3.11) shows that for $\omega \neq c_m$, the derivative of $l_q(\omega)$ is negative (semi) definite. Since it holds for all values of $\omega \neq c_m$ and each $l_q(\omega)$ is continuous at all $c_m \neq c_{i_1}, \ldots, c_{i_{I_q}}$, the function $l_q(\omega)$ is monotonically decreasing in each of the intervals $(-\infty, c_{i_1})$, (c_{i_1}, c_{i_2}), ..., $(c_{i_{I_q}}, \infty)$. This property of the eigenfunctions is formulated in Lemma 2:

Lemma 2: Let each eigenfunction $l_q(\omega)$ have I_q poles at $c_{i_1}, c_{i_2}, \ldots, c_{i_{I_q}}$. It follows that each $l_q(\omega)$ is monotonically decreasing in each of the intervals $(-\infty, c_{i_1})$, (c_{i_1}, c_{i_2}), ..., $(c_{i_{I_q}}, \infty)$.

To complete this analysis, let us consider the asymptotic behavior of the functions $l_q(\omega)$. From the fact that $\mathbf{L}(\omega) \to \mathbf{A}$ for $\omega \to \pm\infty$, one immediately obtains the following lemma.

Lemma 3: For $\omega \to \infty$ as well as for $\omega \to -\infty$, each eigenfunction $l_q(\omega)$ asymptotically approaches a particular eigenvalue of the matrix $\mathbf{A}$.

Comparing the above results with those of the one-dimensional case, one finds that the monotonicity and the asymptotic behavior of the function $l(\omega)$ is preserved. However, for $N = 1$ the number of poles of $l(\omega)$ is M, whereas for $N > 1$ the number of poles I_q of a particular function $l_q(\omega)$ is always less than M. Denoting by $l_1(\omega)$ the function which has a pole at $\omega = c_1$, the next poles of $l_1(\omega)$ occur at $c_{1+N}, c_{1+2N}, \ldots, c_{1+(I_1-1)N}$. Similarly, the poles of $l_2(\omega)$ are $c_2, c_{2+N}, \ldots, c_{2+(I_2-1)N}$, etc. It thus follows that I_q is given by

$$I_q = 1 + \left[\frac{M-q}{N} \right], \quad M,N \geqslant q. \quad (3.12)$$

It should be noted that this formula for the number of poles is valid only if all c_m are different.

The behavior of the functions $l_q(\omega)$ according to Lemmas 1–3 is conveniently visualized by a numerical example. Figure 3 shows a graph of the functions $l_q(\omega)$ for the specific choice $N = 3$ and $M = 4$. (The explicit matrix elements of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are listed in the Appendix.) The values of the diagonal matrix elements of $\mathbf{C}$ are marked by vertical dotted lines. Starting from the asymptote $y = -2.5$ the function $l_1(\omega)$ has a simple pole at $\omega = -3$, whereas it is continuous at $\omega = -1$ and $\omega = 1$. The second pole of this function is at $\omega = 3$ and it then approaches the asymptote $y = 0.5$. The remaining two functions $l_2(\omega)$ and $l_3(\omega)$ have only one pole each. The numerical values of the three functions $l_q(\omega)$ at $\omega = c_m$ are also given in the Appendix (Table AI).

As in the one-dimensional case, the $N + M$ eigenvalues of $\mathbf{Z}$ are obtained as the values of ω at the intersections of the straight line $y = \omega$ with the curves $y = l_q(\omega)$, $1 \leqslant q \leqslant N$. Now let us consider a specific interval $[c_k, c_{k+1}]$ bounded by two successive poles of $\mathbf{L}(\omega)$. By evaluating the values of the functions $l_q(\omega)$ numerically at the border points c_k and c_{k+1} (Lemma 1), and by keeping in mind the monotonicity (Lemma 2), one can directly determine the number of intersections in the particular interval. The same is achieved for

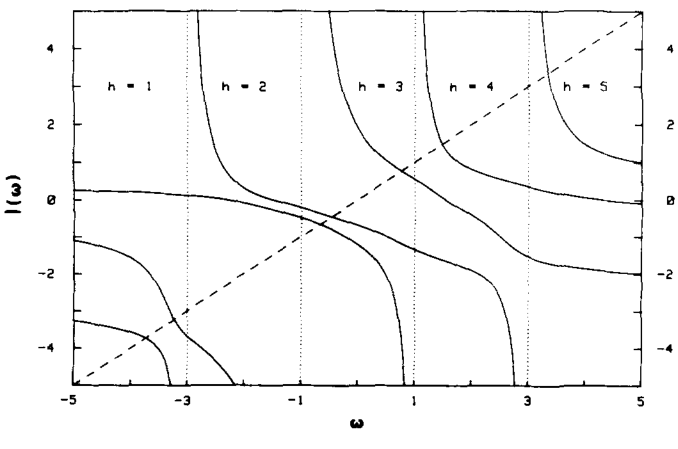


FIG. 3. Graphical determination of the eigenvalues of a Hermitian arrow matrix $\mathbf{Z}$ for the multidimensional case ($N = 3$, $M = 4$). The full lines represent the three functions $l_1(\omega)$, $l_2(\omega)$, and $l_3(\omega)$; the poles of $\mathbf{L}(\omega)$ are indicated by vertical dotted lines. At each pole of $\mathbf{L}(\omega)$, exactly one $l_q(\omega)$ has a simple pole. There are three intersection points of $l_q(\omega)$ with the dashed line $y = \omega$ in the interval $h = 3$, two in the interval $h = 1$, but none in the interval $h = 2$. The slope of $l_q(\omega)$ at an intersection determines the residues of the Green's function and, thus, the eigenvectors of $\mathbf{Z}$.

the two intervals $(-\infty, c_1]$ and $[c_M, \infty)$ if one takes into account the asymptotic behavior of the eigenfunctions (Lemma 3). This procedure allows the extension of the eigenvalue count to values of ω including the points $\omega = c_m$; it also implies that the set of intervals $(-\infty, c_1]$, $[c_1, c_2]$, ..., $[c_m, \infty)$ provides a particular choice for slicing the whole ω-range.

As an illustration let us return to our numerical example. From Fig. 3 or, equivalently, from Table AI in the Appendix, one sees that there are three eigenvalues of $\mathbf{Z}$ in the interval $h = 3$, two in the interval $h = 1$, but none in the interval $h = 2$. Since there are N eigenfunctions $l_q(\omega)$, the upper limit for the number of intersection points in each interval h is N, i.e., the total number of eigenvalues may vary between 0 and N. Note that this statement, as well as additional conclusions on the eigenvalues of $\mathbf{Z}$, follows directly from the interlace theorem of Cauchy.[11]

B. Degenerate poles of $\mathbf{L}(\omega)$

We now drop the condition that all matrix elements c_m are different. From the previous discussion, it is evident that the occurrence of degenerate poles of $\mathbf{L}(\omega)$ does not influence the monotonicity nor the asymptotic behavior of the eigenfunctions $l_q(\omega)$. However, as will now be shown, Lemma 1 must be modified. To be more explicit, suppose K matrix elements c_m are equal, say $c_k = c_{k+1} = \cdots = c_{k+K-1}$. The corresponding coupling matrix elements form K vectors $\mathbf{B}_k$, ..., $\mathbf{B}_{k+K-1}$. As in the last section, $\mathbf{L}(\omega)$ is decomposed according to Eq. (3.3). This leads to an expression analogous to Eq. (3.3), where $\mathbf{U}_k$ is given by

$$\mathbf{U}_k = \sum_{i=0}^{K-1} \mathbf{B}_{k+i} \mathbf{B}_{k+i}^\dagger. \quad (3.13)$$

Now let J be the number of linearly independent vectors in the set of vectors $\mathbf{B}_k, \ldots, \mathbf{B}_{k+K-1}$. Then $\mathbf{U}_k$ is of rank J, and one finds that it has J nonvanishing eigenvalues $\beta_1^{(k)}, ..., \beta_J^{(k)}$ (note that $J \leqslant N$). Transforming $\mathbf{L}(\omega)$ by means of a unitary transformation $\mathbf{S}$ which diagonalizes $\mathbf{U}_k$, one arrives at a matrix similar to that of Eq. (3.6):

$$\mathbf{S}^\dagger \mathbf{L}(\omega)\mathbf{S} = \begin{pmatrix} \mathbf{Q}(\omega) + \dfrac{\mathbf{P}^{(k)}}{\omega - c_k} & \mathbf{F}(\omega) \\ \mathbf{F}^\dagger(\omega) & \mathbf{D}(\omega) \end{pmatrix}, \tag{3.14a}$$

$$P_{ij}^{(k)} = \beta_i^{(k)}\delta_{ij}, \quad i,j = 1,2,...,J, \tag{3.14b}$$

$$D_{ij}(\omega) = Q_{ij}(\omega), \quad i,j = J+1,...,N, \tag{3.14c}$$

$$F_{ij}(\omega) = Q_{ij}(\omega), \quad i = 1,...,J, \quad j = J+1,...,N, \tag{3.14d}$$

where the matrix elements $Q_{ij}(\omega)$ have already been defined in Eq. (3.6c). Equation (3.14) shows that for $\omega \to c_k$, the first J diagonal elements of $\mathbf{S}^\dagger\mathbf{L}(\omega)\mathbf{S}$ decouple from the $(N-J)$-dimensional submatrix $\mathbf{D}(\omega)$. By a consideration similar to the one given in the previous section, we can thus generalize the contents of Lemma 1 to the case of degenerate poles of $\mathbf{L}(\omega)$:

Lemma 4: Let c_k be a K-fold degenerate pole of $\mathbf{L}(\omega)$, $c_k = c_{k+1} = \cdots = c_{k+K-1}$, and J be the number of linearly independent vectors in the set $\mathbf{B}_k, ..., \mathbf{B}_{k+K-1}$ formed by the coupling matrix elements $b_{ik}, ..., b_{ik+K-1}$. It follows that J functions $l_q(\omega)$ have a (simple) pole at $\omega = c_k$. Provided that the nonvanishing eigenvalues $\beta_i^{(k)}$ of the matrix $\mathbf{U}_k$ [Eq. (3.13)] are nondegenerate,[12] the analytic form of these functions in the vicinity of c_k is given by

$$l_q(\omega) \underset{\omega \to c_k}{\simeq} \frac{\beta_i^{(k)}}{\omega - c_k} + Q_{ii}(c_k), \quad i = 1,2,...,J. \tag{3.15}$$

The remaining $(N-J)$ functions are continuous at $\omega = c_k$ and the corresponding values at c_k are obtained as the eigenvalues of the $(N-J)$-dimensional submatrix $\mathbf{D}(c_k)$ specified in Eq. (3.14).

To illustrate the significance of this last lemma, we have plotted the three functions $l_1(\omega)$, $l_2(\omega)$, and $l_3(\omega)$ for the same set of parameters used in Fig. 3 with one additional pole at $\omega = -1$ which is now twofold degenerate. The coupling matrix elements are chosen such that the two vectors $\mathbf{B}_2$ and $\mathbf{B}_3$ are linearly independent. Figure 4 shows that now, at $\omega = -1$, both the functions $l_2(\omega)$ and $l_3(\omega)$ have a pole. Since the additional pole of $\mathbf{L}(\omega)$ increases the dimension of $\mathbf{Z}$ by 1, there must be one additional eigenvalue of $\mathbf{Z}$. From Fig. 4 it follows that this solution lies in the interval $h = 2$. Lemma 4 also shows that the case of degenerate poles of $\mathbf{L}(\omega)$ can easily be incorporated into the procedure for the eigenvalue count described in the last section. The only modification to be made is that one must take into account that, in the general case, up to N eigenfunctions $l_q(\omega)$ may have a simple pole at the left or right boundary of h.

To conclude this analysis of the eigenvalues g_p of a Hermitian arrow matrix $\mathbf{Z}$, we consider the possibility of degenerate eigenvalues. If we exclude the case of degeneracies due to linear dependencies of the first N columns of $\mathbf{Z} - g_p\mathbf{1}$ [which implies that two or more functions $l_q(\omega)$ are tangent of one another at $\omega = g_p$], such degeneracies (as in the one-dimensional case) can only occur for $g_p = c_k$. If c_k is an eigenvalue of $\mathbf{Z}$, it must satisfy the relation

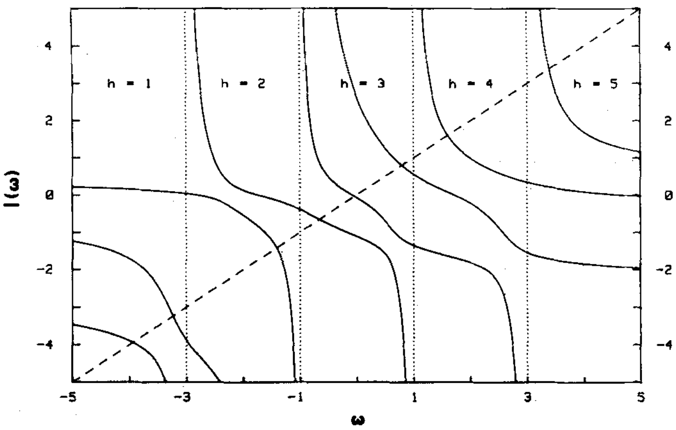


FIG. 4. Graphical determination of the eigenvalues of a Hermitian arrow matrix $\mathbf{Z}$ for the multidimensional case ($N = 3$, $M = 5$). For an explanation of the symbols see Fig. 3. The matrix $\mathbf{L}(\omega)$ is the same as in Fig. 3 except that an additional pole at $\omega = -1$ has been added. Therefore, two functions $l_q(\omega)$ have simple poles at this point. The additional eigenvalue of $\mathbf{Z}$ lies in the interval $h = 2$.

$$|Z - c_k\mathbf{1}| = 0. \tag{3.16}$$

Decomposing $\mathbf{Z}$ according to Eq. (3.3) and (3.13), one obtains

$$\begin{vmatrix} \mathbf{A} - c_k\mathbf{1} & \tilde{\mathbf{B}} & \tilde{\tilde{\mathbf{B}}}_k \\ \gamma\tilde{\mathbf{B}} & \tilde{\mathbf{C}} - c_k\mathbf{1} & \mathbf{0} \\ \gamma\tilde{\tilde{\mathbf{B}}}_k^\dagger & \mathbf{0} & \mathbf{0} \end{vmatrix} = 0, \tag{3.17}$$

where $\tilde{\tilde{\mathbf{B}}}_k$ is a $N \times K$ matrix of rank J containing all coupling matrix elements which correspond to the K-fold degenerate pole c_k. Since $J < K$, the columns of $\mathbf{Z} - c_k\mathbf{1}$ are linearly dependent. It follows that $|\mathbf{Z} - c_k\mathbf{1}|$ is zero and c_k is a $(K-J)$-fold degenerate eigenvalue of $\mathbf{Z}$. As in the one-dimensional case, this eigenvalue of $\mathbf{Z}$ has vanishing residues, i.e., it does not occur as a pole of $\mathbf{G}(\omega)$. To demonstrate this, it is sufficient to show that the first eigenvector components of the matching eigenvector are zero. Decomposing $\mathbf{X}_k$ according to Eq. (3.17) as $\mathbf{X}_k = (\mathbf{X}_A^{(k)}, \mathbf{X}_{\tilde{B}}^{(k)}, \mathbf{X}_{\tilde{\tilde{B}}}^{(k)})$, it follows from Eq. (3.17) that one can choose $\mathbf{X}_A^{(k)} = \mathbf{X}_{\tilde{B}}^{(k)} = 0$ because there are $(K-J)$ linearly independent solutions of the equation

$$\tilde{\tilde{\mathbf{B}}}_k\mathbf{X}_{\tilde{\tilde{B}}}^{(k)} = \mathbf{0}. \tag{3.18}$$

Let us summarize this result in the following lemma.

Lemma 5: Let c_k be a K-fold degenerate pole of $\mathbf{L}(\omega)$, $c_k = c_{k+1} = \cdots = c_{k+K-1}$, J be the number of linearly independent vectors in the set of vectors $\mathbf{B}_k, ..., \mathbf{B}_{k+K-1}$ formed by the coupling matrix elements $b_{ik}, ..., b_{ik+K-1}$, and $J < K$. Let further the first N columns of $\mathbf{Z} - c_k\mathbf{1}$ be linearly independent. It follows that $g_p = c_k$ is a $(K-J)$-fold degenerate eigenvalue of $\mathbf{Z}$. Such a degenerate eigenvalue cannot occur as a pole of the associated Green's function because its corresponding residues are zero. Apart from this case, a particular eigenvalue g_p is degenerate if and only if the rank of the first N columns of $\mathbf{Z} - g_p\mathbf{1}$ is smaller than $N - 1$.

C. The non-Hermitian case $\gamma = -1$

In this subsection, we briefly discuss the special class of non-Hermitian arrow matrices $\mathbf{Z}$ corresponding to $\gamma = -1$. In that case, the matrix $\mathbf{L}(\omega)$ reads

$$\mathbf{L}(\omega) = \mathbf{A} - \mathbf{B}(\omega\mathbf{1} - \mathbf{C})^{-1}\mathbf{B}^\dagger. \tag{3.19}$$

As has been discussed for the one-dimensional case, the sign of γ changes the features of the Green's function drastically, and complex solutions of Eq. (3.2) may occur.[13] With respect to the real eigenvalues of $\mathbf{Z}$, the same graphical analysis can be applied as for Hermitian arrow matrices. One then finds that the above results concerning the pole structure of the eigenfunctions $l_q(\omega)$ (Lemmas 1 and 4) and the asymptotic behavior (Lemma 3) also hold for $\gamma = -1$. The derivative of each function $l_q(\omega)$, however, is, according to Eq. (3.11), always positive (semi) definite, implying that each $l_q(\omega)$ is monotonically increasing in each interval $h = (c_{i_1}, c_{i_2}), ..., (c_{i_{l_q-1}}, c_{i_{l_q}})$ where the $c_{i_1}, ..., c_{i_{l_q}}$ denote the pole positions of $l_q(\omega)$. As a direct consequence, it follows that in each interval h there is *at least* one intersection point of $y = \omega$ with the particular function $l_q(\omega)$. The number of eigenvalues in a specific interval $[c_k, c_{k+K}]$ between two successive (nodegenerate) matrix elements c_m can no longer be determined by the values of the functions $l_q(\omega)$ at $\omega = c_k$ and $\omega = c_{k+K}$ because each $l_q(\omega)$ may intersect the straight line $y = \omega$ more than once in the considered interval. This special feature of the eigenfunctions may also introduce multiple zeros of the function $\omega - l_q(\omega)$ leading to higher-order poles of $\mathbf{G}(\omega)$ which destroys the simple spectral representation given in Eq. (1.6).

D. Determination of eigenvectors

This paragraph contains a summary of how the eigenvectors of an eigenvalue g_p can be obtained by calculating the residues of the corresponding poles of $\mathbf{G}(\omega)$. Here, we restrict ourselves to eigenvalues of Hermitian arrow matrices and to nondegenerate real eigenvalues of non-Hermitian arrow matrices ($\gamma = -1$).

In both cases the spectral representation of the Green's function, Eq. (1.6), relates the first N components of an eigenvector $\mathbf{X}_p$ to the residues of the Green's function R^p_{ij} which, for first-order poles, are given by

$$R^p_{ij} = y_{iq}(g_p)\left[1 - \frac{\partial}{\partial\omega} l_q(\omega)\Big|_{\omega = g_p}\right]^{-1} y^*_{jq}(g_p), \tag{3.20}$$

where $\mathbf{Y}_q(\omega)$ is the eigenvector of the particular function $l_q(\omega)$ which intersects the line $y = \omega$ at $\omega = g_p$. The derivative of $l_q(\omega)$ with respect to ω is explicitly given by Eq. (3.11). Comparing Eq. (3.20) with the spectral representation of $\mathbf{G}(\omega)$, one obtains

$$x^R_{ip} = y_{iq}(g_p)\left[1 - \frac{\partial}{\partial\omega} l_q(\omega)\Big|_{\omega = g_p}\right]^{-1/2},$$
$$x^L_{pi} = (x^R_{ip})^*; \quad 1 \leqslant i \leqslant N. \tag{3.21a}$$

The remaining components $x_{N+i,p}$, $1 \leqslant i \leqslant M$, are then given by ($g_p \neq c_i$).

$$x^R_{N+i,p} = \gamma \frac{\sum_{n=1}^N b^*_{ni} x^R_{np}}{g_p - c_i}, \qquad x^L_{p,N+i} = \frac{\sum_{n=1}^N x^L_{pn} b_{ni}}{g_p - c_i}, \qquad 1 \leqslant i \leqslant M. \tag{3.21b}$$

If $g_p = c_i$, either Eq. (3.18) must be used to obtain the K components of $\mathbf{X}^R_p$ ($\mathbf{X}^L_p$) different from zero or $x^R_{N+i,p} = x^L_{p,N+i} = 0$, $1 \leqslant i \leqslant M$. Note that the eigenvectors are orthonormal and that for Hermitian $\mathbf{Z}$ the usual relation $\mathbf{X}^L_p = (\mathbf{X}^R_p)^\dagger$ is recovered.

IV. A POLE SEARCH ALGORITHM FOR SOLVING THE EIGENVALUE PROBLEM OF HERMITIAN ARROW MATRICES

The preceding analysis has shown that the eigenvalue problem of Hermitian arrow matrices can be treated by implementing a pole search algorithm for $\mathbf{G}(\omega)$. Such a pole search algorithm should be viewed in connection with well-established procedures based on the triangular factorization mentioned in Sec. I. The latter methods all suffer from the fact that they cannot be applied to eigenvalues of $\mathbf{Z}$ which are very close or equal to the diagonal elements c_m of $\mathbf{C}$ because, for $\omega = c_m$, the matrix elements of $\mathbf{L}(\omega)$ are singular. However, considering the eigenfunctions $l_q(\omega)$ rather than $\mathbf{L}(\omega)$ itself shows that (in general) a part of the eigenvalue spectrum of $\mathbf{L}(c_m)$ is regular. The analysis given here constitutes a means for obtaining the regular eigenvalues numerically owing to the fact that the diverging eigenfunctions are separable by a unitary transformation. Another advantage of the new algorithm results from the fact that the derivative of each function $l_q(\omega)$ is directly obtainable, thus permitting the use of iterative methods of higher order for the pole search. In the following, the essential features of this pole search algorithm are now summarized.

(1) As has been discussed in Sec. IIIA, the diagonal matrix elements c_m of $\mathbf{C}$, $1 \leqslant m \leqslant M$, provide a distinct set of intervals $(-\infty, c_1], ..., [c_k, c_{k+K}], ..., [c_M, \infty)$. For a particular interval $h = [c_k, c_{k+K}]$, where K denotes the degeneracy of c_k, the eigenvalue count is performed in the following way. Since the eigenfunctions $l_q(\omega)$ of $\mathbf{L}(\omega)$ are monotonically decreasing in h, there is at most one intersection of a particular $l_q(\omega)$ with the straight line $y = \omega$. Consequently, the values of the functions $l_q(\omega)$ at $\omega = c_k$ and $\omega = c_{k+K}$ directly allow one to evaluate the number of eigenvalues occuring in h. To see how many functions $l_q(\omega)$ have a pole at $\omega = c_k$, one determines the number J_k of nonvanishing eigenvalues of the matrix $\mathbf{U}_k$ of Eq. (3.3) or (3.13). The values of the remaining $N - J_k$ functions, being continuous at $\omega = c_k$, are then calculated by diagonalizing the $(N - J_k)$-dimensional submatrix $\mathbf{D}(c_k)$ defined in Eqs. (3.6) and (3.14). In the same way, the behavior of the functions $l_q(\omega)$ at $\omega = c_{k+K}$ is investigated. For the two intervals $h = (-\infty, c_1]$ and $h = [c_M, \infty)$, the above procedure can also be applied by taking into account the asymptotes of the eigenfunctions which are given as the eigenvalues of the matrix $\mathbf{A}$.

(2) The eigenvalues g_p in the interval h are determined by solving the equations $\omega - l_q(\omega) = 0$. Due to the fact that the derivatives of the functions $l_q(\omega)$ can directly be obtained via Eq. (3.11), these zeros can be evaluated by employing iterative methods of higher order such as a Newton–Raphson iteration. The iteration involves the calculation of $l_q(\omega)$ at iteration points ω_ν, where $c_k < \omega_\nu < c_{k+K}$. To obtain $l_q(\omega_\nu)$, the matrix elements $L_{ij}(\omega_\nu)$ are constructed via Eq.

(1.3) and the resulting matrix $\mathbf{L}(\omega_v)$ is diagonalized. The value $l_q(\omega_v)$ is then given as the qth (ordered) eigenvalue of $\mathbf{L}(\omega_v)$. By determining the corresponding eigenvector $\mathbf{Y}_q(\omega_v)$, the derivative of the eigenfunction at $\omega = \omega_v$ is obtained via Eq. (3.11). If a particular ω_v is so close to the poles c_k or c_{k+K} that the matrix elements $L_{ij}(\omega_v)$ become numerically singular, the regular eigenvalues of $\mathbf{L}(\omega_v)$ are obtained by diagonalizing the matrix $\mathbf{D}(\omega_v)$ [Eq. (3.6) or (3.14)], whereas the diverging functions (if needed) may be approximated by their analytic form (see Lemmas 1 and 4). In the iterative cycle, only $N \times N$ matrices have to be dealt with and, consequently, the computing time for a particular eigenvalue is (N being fixed) proportional to M.

(3) The eigenvector $\mathbf{X}_p$ of an eigenvalue g_p is directly obtained from the residues of corresponding poles g_p of $\mathbf{G}(\omega)$. The evaluation of the residue involves the derivatives of the intersected function $l_q(\omega)$ at $\omega = g_p$. Having calculated this quantity via Eq. (3.11) all eigenvector components of $\mathbf{X}_p$ follow from Eq. (3.22). For fixed N, the computing time for a particular eigenvector is linear in M, i.e., the time for calculating the whole eigenvector matrix is proportional to M^2.

(4) The procedure described above applies to all eigenvalues of $\mathbf{Z}$ that correspond to poles of the Green's function. To complete the eigenvalue analysis, one must determine the remaining eigenvalues which are characterized by the fact that the first N components of their eigenvectors are zero. Such eigenvalues can only occur if a diagonal matrix element c_m of $\mathbf{C}$ is degenerate. If a particular c_k is K-fold degenerate, and if the K vectors $\mathbf{B}_k, ..., \mathbf{B}_{k+K-1}$ formed by the coupling matrix elements $b_{ik}, ..., b_{ik+K-1}$ are linearly dependent, then c_k is a $(K - J_k)$-fold degenerate eigenvalue of $\mathbf{Z}$. Here, J_k denotes the number of linearly independent vectors in the set $\mathbf{B}_k, ..., \mathbf{B}_{k+K-1}$. The $(K - J_k)$ nonvanishing components of the corresponding eigenvectors are obtained as the solutions of Eq. (3.18).

To illustrate the savings that result from using higher-order iterative methods we return to the numerical example depicted in Fig. 3. For the three eigenvalues in the interval $[-1, 1]$, the progress of a Newton–Raphson iteration is listed in Table I, starting from the same initial value $g_i = 0$ for all three eigenvalues. After four cycles, the three eigenvalues are converged to all seven digits shown in Table I. (The actual accuracy after the fourth cycle is even higher.) If, on the other hand, one uses a bisection procedure, more than 20 cycles are needed to obtain the same precision. Note that the initial value $g_i = 0$ has been chosen for convenience only. For actual computations it is possible to obtain better initial guesses by taking into account the analytic properties of the eigenfunctions $l_q(\omega)$ in the interval $[c_k, c_{k+K}]$, e.g., by approximating a regular function $l_q(\omega)$ by a straight line or by using the explicit form for the diverging functions.

Comparing the above algorithm to standard diagonalization procedures that do not exploit the particular structure of the arrow matrix $\mathbf{Z}$, the features of the pole search algorithm suggest that it will be particularly efficient when only a certain range of the eigenvalue spectrum is needed. The situation is even more promising if, in addition, the matching eigenvectors have to be calculated. In both cases, the larger the "tail" M of the arrow $\mathbf{C}$ compared to the dimension N of the submatrix $\mathbf{A}$, the better the efficiency. If all eigenvalues and eigenvectors are to be calculated, even for medium sized matrices ($M = 300$), we found that for $N = 3$, a factor of 7 in cpu time is gained. For $N = 6$, this saving is reduced to a factor of 3. The drastic reduction of computing time is, of course, due to the fact that the time for evaluating all eigenvectors is proportional to M^2, whereas standard routines exhibit an M^3 dependence. A further saving is accomplished if not all eigenvalues are needed. We also would like to emphasize that very close or degenerate eigenvalues of $\mathbf{Z}$ do not pose any additional difficulties within the framework of the above algorithm because this problem is reduced to the problem of coping with degeneracies of the very small $N \times N$ matrix $\mathbf{L}(\omega)$.

For large matrices ($M \geqslant 1000$), the Lanczos algorithm is often used. If the Lanczos algorithm (without reorthogonalization) is applied to the arrow matrix $\mathbf{Z}$, a comparison shows that the cpu times for the two methods are roughly the same as long as only eigenvalues are computed. However, the great numerical stability of the pole search algorithm is a quality which cannot be matched by the Lanczos algorithm. With the latter, considerable numerical effort is required to eliminate spurious eigenvalues or to calculate eigenvectors. It should also be mentioned that, instead of applying the Lanczos algorithm to $\mathbf{Z}$, it might be more cost effective to apply it to the inverted matrix $(\sigma\mathbf{1} - \mathbf{Z})^{-1}$ for a few values of σ.[14] Once the eigenvalues have been determined via the Lanczos algorithm, the eigenvectors can directly be computed by using the explicit formulas given in Sec. IIID.

TABLE I. Newton iteration for the three eigenvalues g_p in the interval $[-1, 1]$ of the example depicted in Fig. 3. For simplicity, the initial value ($k = 0$) for all three eigenvalues was chosen to be $g_i = 0$.

k \ ω	g_1	g_2	g_3
1	− 0.590 849 3	− 0.458 696 5	0.529 051 7
2	− 0.664 259 4	− 0.468 363 3	0.789 365 9
3	− 0.665 360 7	− 0.468 362 1	0.803 717 0
4	− 0.665 360 9	− 0.468 362 1	0.803 724 4

ACKNOWLEDGMENTS

The authors are indebted to H. Köppel for useful discussions. Financial support by the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

APPENDIX

The matrices $\mathbf{A}$ and $\mathbf{B}$ for the numerical example shown in Fig. 3 were chosen as

$$\mathbf{A} = \begin{pmatrix} -2.5 & 0 & 0 \\ 0 & -0.5 & 0 \\ 0 & 0 & 0.5 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} 0.2 & 1.5 & 0.5 & 0.6 \\ 1.0 & 0.1 & 0.6 & 0.7 \\ 0.1 & 1.0 & 0.35 & 0.5 \end{pmatrix},$$

TABLE AI. Values of the three functions $l_q(\omega)$ depicted in Fig. 3 at the poles c_m. If there is only one entry, the corresponding $l_q(\omega)$ is continuous at $\omega = c_m$. If there are two entries, the function $l_q(\omega)$ has a simple pole at this particular point. The first number is the residue $\beta_1^{(m)}$ of $l_q(\omega)$ at the pole c_m, the second entry refers to the quantity $Q_{11}(c_m)$ [see Eq. (3.7)].

$l_q(c_m)$ \ c_m	-3	-1	1	3
$l_1(c_m)$	$1.05,\ -0.9345$	-0.2258	-1.3257	$1.1,\ -0.0426$
$l_2(c_m)$	-3.6840	$3.26,\ -1.9074$	0.6274	-1.5386
$l_3(c_m)$	0.1221	-0.4831	$0.7325,\ -0.4593$	0.3522

and the diagonal elements c_m of $\mathbf{C}$ are given by

$$c_m = \{-3, -1, 1, 3\}.$$

The values of the eigenfunctions $l_1(\omega)$, $l_2(\omega)$, and $l_3(\omega)$ at the poles c_m of $\mathbf{L}(\omega)$ are calculated via the procedure described in Sec. IIIA and are also listed in Table AI.

[1] E. A. Hylleraas and B. Undheim, Z. Phys. **65**, 759 (1930); J. K. L. MacDonald, Phys. Rev. A **43**, 830 (1933).
[2] P. Löwdin, J. Mol. Spectrosc. **10**, 12 (1963).
[3] J. D. Doll and W. P. Reinhardt, J. Chem. Phys. **57**, 1169 (1972).
[4] J. Simons and W. D. Smith, J. Chem. Phys. **58**, 4899 (1973).
[5] G. D. Purvis and Y. Öhrn, J. Chem. Phys. **60**, 4063 (1974).
[6] L. S. Cederbaum, W. Domcke, J. Schirmer, W. von Niessen, G. H. F. Diercksen, and W. P. Kraemer, J. Chem. Phys. **69**, 1591 (1978).
[7] H. N. W. Lekkerkerker and W. G. Laidlaw, Phys. Rev. A **5**, 1604 (1972); E. van Oost and H. N. W. Lekkerkerker, J. Math. Phys. **19**, 2223 (1978).
[8] E. U. Condon and G.H. Shortley, *The Theory of Atomic Spectra* (Cambridge U.P., Cambridge, England, 1967), Sec. II 11.
[9] P. Löwdin, in *Perturbation Theory and its Application in Quantum Mechanics*, edited by C. H. Wilcox (Wiley, New York, 1966).
[10] H. N. W. Lekkerkerker and W. G. Laidlaw, Phys. Rev. A **9**, 431 (1974); J. A. McLennan, Phys. Rev. A **10**, 1272 (1974).
[11] B. N. Parlett, *The symmetric eigenvalue problem* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
[12] If there are degenerate eigenvalues $\beta_i^{(k)}$, it is straightforward to evaluate the additive constant. The expression, however, is somewhat more complicated.
[13] The maximum number of complex eigenvalues is given as the smaller of the two quantities $2N$ and $2M$, see Ref. 10.
[14] T. Ericson and A. Ruhe, Math. Comp. **35**, 1251 (1980).

The relationship between finite groups and Clifford algebras

Nikos Salingaros
Division of Mathematics, Computer Science, and Systems Design, The University of Texas at San Antonio, San Antonio, Texas 78285

(Received 17 May 1983; accepted for publication 23 June 1983)

Clifford algebras are traditionally realized in terms of a specific set of representation matrices. This paper provides a more effective alternative by giving the finite group associated with each Clifford algebra. All the representation-independent algebraic results, which are really direct consequences of the underlying group structure, can thus be derived in an easier and more general manner. There are five related but distinct classes of finite groups associated with the Clifford algebras. These groups are constructed from the complex, cyclic, quaternion, and dihedral groups in a way which is discussed here in detail. Of particular utility is a table which lists the order structure of each group: this permits the immediate identification of any Clifford algebra in any dimension.

PACS numbers: 02.10.Tq, 02.20. + b

I. INTRODUCTION

The Clifford algebras are of considerable interest in discussions of space-time symmetries and their extensions to unified descriptions of gauge fields. In previous work, we showed how the study of Clifford algebras can be facilitated considerably, if one studies a finite multiplicative group which is uniquely defined by each Clifford algebra.[1,2] Then the apparatus of the theory of finite groups can be utilized to obtain results for the Clifford algebra of interest. This procedure stands in sharp contrast to the more traditional method of using a set of representation matrices to derive structural results in a particular Clifford algebra. Moreover, in this alternate framework, there is no risk of deriving results which may turn out to be dependent upon a specific matrix representation.

In this note, we amplify and extend the results of Refs. 1 and 2. In particular, we identify those finite groups which arise via Clifford algebras in the setting of the finite groups of dimension 4, 8, and 16 (Sec. II). Even though all these groups are well known and classified, this information has been usually unavailable in the form which is most useful for our present purposes: namely, a listing of all these groups in a format which allows a particular group to be directly identified by knowing its center, its order structure, and its subgroup structure. (Just knowing the center and order structure is sufficient to identify all but one pair of groups up to order sixteen.) We also relate the different notations used for these groups either in group theoretical discussions or crystallographic usage.

In Sec. III of this paper, we construct the five types of groups which are defined by Clifford algebras in any dimension. Following this, we review the central product of groups in Sec. IV, and utilize the Frobenius–I. Schur theorem in a general construction of the groups associated with Clifford algebras. A particularly useful feature is Table IV, which lists those groups associated with Clifford algebras up to order 1024 and gives their order structure. It is possible to identify any group which arises from a Clifford algebra directly, so that this table is useful in identifying all the Clifford algebras.

I would like to note that related but distinct discussions of the subject matter of this paper appear in Refs. 3 and 4.

II. REVIEW OF FINITE GROUPS OF ORDER 4, 8, AND 16

In this section we review the finite groups of order 2^n, where $n = 2$, 3, and 4, and identify which of these groups arise as the groups corresponding to a Clifford algebra. We discuss the group structure in some detail for comparison purposes. In particular, we list all the groups in a way which relates the various distinct notations used in the literature.[5–7]

To begin with, there are two groups of order four: The cyclic group Z_4 is isomorphic to the complex group with elements $\{1, -1, i, -i\}$. The Gauss–Klein four-group $Z_2 \otimes Z_2$ is the direct group product of two cyclic groups $Z_2 = \{1, -1\}$ and is isomorphic to the dihedral group D_2. Both of these groups are abelian.

There are five groups of order eight. One is cyclic and two are direct group products of cyclic groups, hence these three are abelian. The remaining two groups are the quaternion group Q_4 with elements $\{\pm 1, \pm \mathbf{i}, \pm \mathbf{j}, \pm \mathbf{k}\}$ where $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$, and the dihedral group D_4. Both of these are nonabelian. The groups of order four and eight are displayed in Table I in the ordering of Ref. 6.

An important property of each finite group is its order structure. The order of a particular element α in the group is the smallest integer p for which $\alpha^p = 1$. The table lists the number of distinct elements in each group which have order 2, 4, or 8. (The identity 1 is the only element of order 1).

TABLE I. Groups of order 4 and 8.

		Order structure		
Order 4	Type	2	4	8
1. $4\Gamma_1(1^2) = Z_2 \otimes Z_2 = D_2$	abelian	3		
2. $4\Gamma_1(2) = Z_4$		1	2	
Order 8				
1. $8\Gamma_1(1^3) = Z_2 \otimes Z_2 \otimes Z_2$	abelian	7		
2. $8\Gamma_1(2, 1) = Z_4 \otimes Z_2$		3	4	
3. $8\Gamma_1(3) = Z_8$		1	2	4
4. $8\Gamma_2 a_1 = D_4 = N_1$	Dihedral	5	2	
5. $8\Gamma_2 a_2 = \langle 2, 2, 2\rangle = Q_4 = N_2$	Quaternion	1	6	

There are fourteen groups of order sixteen: five of them are abelian, and nine are nonabelian. In addition to the order structure, we give the center of each nonabelian group, which is the subgroup of commuting elements. In the abelian case, the center is the entire group. Although this information is standard, it is not always given in the most practical form. There are several common notations in use, but one cannot easily find the correspondence between them. A more serious matter is the fact that some lists of groups of order 16 are incomplete. We list all the groups in Table II in the most common notations.

We have used the ordering of Ref. 6 which introduced the $n\Gamma_k d_r$ notation. The other notation involving brackets derives from the crystallographic groups and is much older; it is extensively employed in Ref. 5. The notation N_k,S_k, and Ω_k identifies those groups which arise in connection with Clifford algebras, and is discussed in detail in the next section. For completeness, we review the most common notation and structure relations for some of the groups.

(i) The cyclic group Z_n is of order n and is defined by the generating relation (note that Z_n is frequently denoted by C_n)

$$Z_n:\{\alpha^n = 1\}. \tag{1}$$

(ii) The dihedral group D_n is of order $2n$, and is defined by the generating relations

$$D_n:\{\alpha_1^n = 1, \alpha_2^2 = 1, (\alpha_1\alpha_2)^2 = 1\}. \tag{2}$$

(iii) The quaternion or dicyclic group $Q_{2m} = \langle 2,2,m\rangle$ is of order $4m$, and is defined by the generating relations (Q_{2m} is written Q_m in Ref. 7)

$$Q_{2m}:\{\alpha_1^{2m} = 1,\ \alpha_2^2 = \alpha_1^m, (\alpha_1\alpha_2)^2 = \alpha_1^m\}. \tag{3}$$

Many of the groups in Tables I and II are obtained as direct group products of the above groups. The structure and generating relations of the remaining groups are given in Refs. 5–7. The Pauli group $16\Gamma_2 b$ is the group of order sixteen without a name in the earlier editions of Ref. 5. A particularly useful reference which has recently appeared is Ref. 8. Note, however, that the center of the group $16\Gamma_2 d$ is given there incorrectly.

III. STRUCTURE OF THE FINITE GROUPS CORRESPONDING TO CLIFFORD ALGEBRAS

There are five distinct types of finite groups that arise from Clifford algebras. In Refs. 1 and 2 they were called "vee groups" and were labelled as

$$N_{\text{odd}}, N_{\text{even}}, \Omega_{\text{odd}}, \Omega_{\text{even}}, S_k. \tag{4}$$

The S-groups are the "spinor" groups: S_1 is the group of the complex Pauli matrices, and S_2 is the group of the Dirac matrices. The odd N-groups correspond to real spinors; for example, N_1 is related to real 2-spinors, and N_3 is the group of the real Majorana matrices.[1,2] The even N-groups define the quaternionic spinors.

Each Clifford algebra $A^{p,q}$ in a space of dimension $p+q$ and metric signature $p-q$ corresponds to one of the vee groups (4). There are isomorphisms between Clifford algebras of the same dimension which limit the distinct types of Clifford algebras to five, corresponding to the finite groups (4). In particular, there are only two N-groups corresponding to even-dimensional Clifford algebras; there are two Ω-groups and one S-group which correspond to odd-dimensional Clifford algebras. The precise identification is given in Refs. 1 and 2. In Table III, we list the vee groups in terms of the dimension of the base space of the corresponding Clifford algebra, and also give the order of the finite group.

In Sec. II we identified the first few vee groups as follows:

$$N_1 = 8\Gamma_2 a_1, \quad N_2 = 8\Gamma_2 a_2,$$
$$\Omega_1 = 16\Gamma_2 a_1, \quad \Omega_2 = 16\Gamma_2 a_2, \quad S_1 = 16\Gamma_2 b. \tag{5}$$

These groups correspond to the Clifford algebras in dimensions two and three, respectively. All the groups belong to the same family[6] Γ_2. Furthermore, the Ω-groups are double copies of the N-groups and can be written as a direct group product of the N-groups with the group of two elements[1] Z_2:

$$\Omega_k = N_k \otimes Z_2. \tag{6}$$

Because of (6), it is easy to see why N_1,Ω_1 and N_2,Ω_2,

TABLE II. Groups of order 16.

Group	Type		Order structure 2	4	8	16
1. $16\Gamma_1(1^4) = Z_2 \otimes Z_2 \otimes Z_2 \otimes Z_2$	abelian		15			
2. $16\Gamma_1(2, 1^2) = Z_4 \otimes Z_2 \otimes Z_2$			7	8		
3. $16\Gamma_1(2^2) = Z_4 \otimes Z_4$			3	12		
4. $16\Gamma_1(3, 1) = Z_8 \otimes Z_2$			3	4	8	
5. $16\Gamma_1(4) = Z_{16}$			1	2	4	8
		Center				
6. $16\Gamma_2 a_1 = D_4 \otimes Z_2 = \Omega_1$	Generalized dihedral	D_2	11	4		
7. $16\Gamma_2 a_2 = Q_4 \otimes Z_2 = \Omega_2$	Hamiltonian	D_2	3	12		
8. $16\Gamma_2 b = \langle 2, 2, 2\rangle_2 = S_1$	Pauli	Z_4	7	8		
9. $16\Gamma_2 c_1 = (4, 4 \vert 2, 2)$		D_2	7	8		
10. $16\Gamma_2 c_2 = \langle 2, 2 \vert 4; 2\rangle$		D_2	3	12		
11. $16\Gamma_2 d = \langle 2, 2 \vert 2\rangle$		Z_4	3	4	8	
12. $16\Gamma_3 a_1 = D_8$	dihedral	Z_2	9	2	4	
13. $16\Gamma_3 a_2 = \langle -2, 4 \vert 2\rangle$		Z_2	5	6	4	
14. $16\Gamma_3 a_3 = \langle 2, 2, 4\rangle = Q_8$	dicyclic	Z_2	1	10	4	

TABLE III. Finite groups associated with Clifford algebras.

Group	Center	Order of group	Dimension of algebra
N_{2k-1}	Z_2	2^{2k+1}	$2k$
N_{2k}	Z_2		
Ω_{2k-1}	D_2	2^{2k+2}	$2k+1$
Ω_{2k}	D_2		
S_k	Z_4		

respectively, belong to the same genus a_1 and a_2. The spinor group S_1 belongs to the genus b in the classification of Ref. 6. This establishes the relationship between the vee groups in terms of family and genus.

A key result on the group structure was given in Ref. 1, where we determined the center $C(G)$ of each vee group G, and also the factor group $G/C(G)$, which is abelian. The different cases are given by the following theorem.

Theorem 1:

$$\frac{G}{C(G)}:\frac{N_{2k-1}}{Z_2}\approx\frac{N_{2k}}{Z_2}\approx\frac{\Omega_{2k-1}}{D_2}\approx\frac{\Omega_{2k}}{D_2}\approx\frac{S_k}{Z_4}\approx(Z_2)^{\otimes 2k}. \tag{7}$$

All the factor groups are equal to the direct product of $2k$ copies of the group Z_2. This property (7) identifies the vee groups as "extra-special-2-groups," and was utilized in Ref. 2 to investigate the algebraic structure. The identity (7), combined with the order structure of the groups, is sufficient to identify the larger vee groups. For example, the vee groups of order 32 and 64 can be uniquely identified in the classification of Ref. 6 as follows:

$$N_3 = 32\Gamma_5 a_1, \quad N_4 = 32\Gamma_5 a_2,$$
$$\Omega_3 = 64\Gamma_5 a_1, \quad \Omega_4 = 64\Gamma_5 a_2, \quad S_2 = 64\Gamma_5 b. \tag{8}$$

The algebras corresponding to N_3 and S_2 are the algebras of the Majorana and Dirac matrices, respectively.

The N-groups, which are the stem groups of the classification, can be described as finite groups in the following manner. [It is interesting to see that the N-groups can be obtained by changing only signs in the multiplication table for $(Z_2)^{\otimes 2k}$, as is suggested by Theorem 1.] We can use a specific identification for the N-groups in terms of Clifford algebras $A^{p,q}$, in order to obtain a set of generating relations for the finite groups. We stress that this identification is not unique, but is for the present purpose the most convenient one: that is, the following, obtained from Table I of Ref. 2,

$$N_{2k-1}\approx A^{k,k}, \quad N_{2k}\approx A^{k-1,k+1}. \tag{9}$$

From Table III, the dimension of the base space is obviously $2k$ in both cases. Therefore the Clifford algebra has $2k$ generators $\alpha_1,...,\alpha_{2k}$ which mutually anticommute. There is also the volume element $\omega = \alpha_1\alpha_2\cdots\alpha_{2k}$ which anticommutes with all of the α_j. The squares of the α_j are determined by the metric. For instance, in $A^{k,k}$, k of the α_j have square equal to 1, while the other k have square equal to -1. (It is important to note that 1 and -1 are distinct elements of the finite group). The square of ω can be determined from the well-known identity[1]

$$\omega^2 = (-1)^{n(n-1)/2}\det g. \tag{10}$$

Here n is the dimension of the base space, and $\det g$ is the determinant of the diagonal metric which equals the product of all the squares of the generators. In the specific case (9), we therefore have

$$\omega^2 = \begin{cases} 1, & \text{for } A^{k,k} \\ -1, & \text{for } A^{k-1,k+1}. \end{cases} \tag{11}$$

By combining all the above information, we obtain a realization of the N-groups. Even though we used a specific Clifford algebra in each case, the Clifford algebra isomorphisms guarantee that the following are, in fact, general definitions. The generating relations for the N- groups can therefore be given as

$N_{2k-1}:\{\alpha_1,...,\alpha_{2k},\omega\}$ mutually anticommute.

$$\begin{cases} \alpha_1{}^2 = \cdots = \alpha_k{}^2 = \omega^2 = 1, \\ \alpha_{k+1}{}^2 = \cdots = \alpha_{2k}{}^2 = -1, \\ \omega = \alpha_1\alpha_2\cdots\alpha_{2k}. \end{cases} \tag{12a}$$

$N_{2k}:\{\alpha_1,...,\alpha_{2k},\omega\}$ mutually anticommute.

$$\begin{cases} \alpha_1{}^2 = \cdots = \alpha_{k-1}{}^2 = 1, \\ \alpha_k{}^2 = \cdots = \alpha_{2k}{}^2 = \omega^2 = -1, \\ \omega = \alpha_1\alpha_2\cdots\alpha_{2k}. \end{cases} \tag{12b}$$

The simplest cases have already been discussed: $N_1 = D_4$ is the dihedral group, and $N_2 = Q_4$ is the quaternion group. One can see how definitions (12) generalize the dihedral and quaternion groups in a way that is radically different from D_n and Q_{2m} defined in Sec. II.

The S-groups are obtained by complexifying the Clifford algebra. What this process does to the group is to add a commuting element β to the group generators; one additionally identifies β^2 with the element -1 in the group. One possible definition using (12a) is the following:

$S_k:\{\alpha_1,...,\alpha_{2k},\omega\}$ mutually anticommute; β commutes.

$$\begin{cases} \alpha_1{}^2 = \cdots = \alpha_k{}^2 = \omega^2 = 1, \\ \alpha_{k+1}{}^2 = \cdots = \alpha_{2k}{}^2 = \beta^2 = -1, \\ \omega = \alpha_1\alpha_2\cdots\alpha_{2k}. \end{cases} \tag{13}$$

The smallest S-group is the Pauli group S_1. For example, the realization $(\alpha_1,\alpha_2,\omega,\beta) = (\tau_1,i\tau_2,-\tau_3,i)$ can be given in terms of the Pauli matrices. The next S-group is the group of the Dirac matrices S_2, which may be realized in the standard (Bjorken and Drell) representation as $(\alpha_1,\alpha_2,\alpha_3,\alpha_4,\omega,\beta) = (i\gamma^1,\gamma^4,\gamma^3,\gamma^2,\gamma^5,i)$.

We stress that the above definition of the group (13) is *not* equal to $N\otimes Z_4$, because of the identification of $\beta^2\in Z_4$ with $-1\in N$. Comparing the orders, we see that S_k is of order 2^{2k+2} (Table III), while $N_{2k-1}\otimes Z_4$ is of order $2^{2k+1}\times 4 = 2^{2k+3}$, i.e., twice the correct order. By factoring out a Z_2 from $N\otimes Z_4$, then one can easily obtain an abelian group of the correct order because there are several ways to factor Z_2 from $N\otimes Z_4$; this is not, however, the group S_k. (This error was made in Refs. 1 and 2, so that a formula given for the S-groups there is incorrect). A discussion of the proper construction of the S-groups from the N-groups is the subject of the following section.

Finally, the Ω-groups are obtained from the N-groups via (6). One adds a commuting element γ, where $\gamma^2 = 1$ to each N-group (12a) or (12b) to obtain the two distinct groups $\Omega_{\rm odd}$ and $\Omega_{\rm even}$.

IV. THE GROUPS AS CENTRAL PRODUCTS OF THE QUATERNION AND DIHEDRAL GROUPS

The discussion in the preceding section (Theorem 1 and Table III) identifies the N-groups to be "extra-special-2-groups" which are defined as follows.[9]

Definition 1: A finite group G is called an extra-special-2-group if G is of order 2^n, $C(G)$ is of order 2, and $G/C(G)$ is the abelian group $Z_2 \otimes \cdots \otimes Z_2 (n-1 \text{ times})$.

Once this fact is established, the structure of all the vee groups can be determined from the body of results on extra-special-2-groups. First, we recall the *central product* of two finite groups and show how it differs from the direct product. As an illustrative example, consider the direct product between a quaternion and a dihedral group $Q_4 \otimes D_4$. The resulting group has order $8 \times 8 = 64$. Furthermore, its center is the direct product of the two individual centers and is equal to $Z_2 \otimes Z_2$.

The central product of Q_4 with D_4, denoted as $Q_4 \odot D_4$ here, amalgamates the Z_2 center of Q_4 with the Z_2 center of D_4 to give the center of $Q_4 \odot D_4$ as Z_2. Hence the central product of Q_4 with D_4 is of order 32. In the case where one center is a subgroup of the other center, they both amalgamate into the larger center.

We now recall the corollary to the Frobenius–Schur theorem from Ref. 9, p. 193, Ref. 10, p. 355, and Ref. 11, p. 277.

Theorem 2: If G is an extra-special-2-group of order 2^{2k+1}, then

(i) $G \approx D_4 \odot \cdots \odot D_4$ (k times), or

(ii) $G \approx Q_4 \odot D_4 \odot \cdots \odot D_4$ ($k-1$ times);

(iii) type (i) contains exactly $2^{2k} + 2^k - 1$ involutions, and type (ii) contains exactly $2^{2k} - 2^k - 1$ involutions.

An involution is an element of order two. We can apply the above theorem to give the structure of the vee groups as follows (the powers mean repeated central products, and $N_1 \approx D_4$, $N_2 \approx Q_4$):

Theorem 3:

$$\begin{aligned}
&N_{2k-1} \approx (N_1)^{\odot k},\\
&N_{2k} \approx N_2 \odot (N_1)^{\odot (k-1)},\\
&\Omega_{2k-1} \approx N_{2k-1} \odot D_2,\\
&\Omega_{2k} \approx N_{2k} \odot D_2,\\
&S_k \approx N_{2k-1} \odot Z_4 \approx N_{2k} \odot Z_4.
\end{aligned} \tag{14}$$

The elementary identity $N \odot (Z_2 \otimes Z_2) \approx N \otimes Z_2$ relates (14) to the definition of the Ω-groups (6). Also, $N_1 \odot N_1 \approx N_2 \odot N_2$.[1,2,9,10]

Note that in taking direct products of *algebras*, one always amalgamates their centers, so that the usual $\otimes$ product between algebras in fact corresponds to the $\odot$ product between the underlying groups, and not to the $\otimes$ product of those groups.

Because the structure of the vee groups is determined by their connection to Clifford algebras, a knowledge of the order structure is sufficient to identify them uniquely. The order structure of the N-groups is determined as follows: Theorems 2 and 3 give the number of elements of order two as the involutions. Since the order of the N-groups is 2^{2k+1} (Table III), the number of elements of order four is just the difference (remembering that there is also the unit of order one to subtract).

$$\begin{aligned}
&\text{type (i):} \quad 2^{2k} - 2^k,\\
&\text{type (ii):} \quad 2^{2k} + 2^k.
\end{aligned} \tag{15}$$

It is convenient to introduce the notation $(1,a,b)$, or more simply (a,b) to denote the order structure of a group: one element of order one, a elements of order two, and b elements of order four. Since the vee groups arise from Clifford algebras, they contain no elements of any higher order. The order structure of the N-groups can therefore be given in general from Theorem 2 and Eq. (15) as

$$\begin{aligned}
&N_{2k-1}: (2^{2k} + 2^k - 1,\ 2^{2k} - 2^k),\\
&N_{2k}: (2^{2k} - 2^k - 1,\ 2^{2k} + 2^k).
\end{aligned} \tag{16}$$

The order structure of the Ω and S groups is obtained by using the central product and (14). In general, it is easy to see that the direct product of two groups with orders (a,b) and (c,d), respectively, will have order

$$(a,b) \otimes (c,d) = (a + c + ac + bd, b + d + bc + ad). \tag{17}$$

In the case of a direct product with the group Z_2, we have, from Table I and (17),

$$(a,b) \otimes (1,0) = (1 + 2a, 2b). \tag{18}$$

This identity (18) allows the calculation of the order structure of the Ω-groups from (6) and (16). One obtains

$$\begin{aligned}
&\Omega_{2k-1}: (2^{2k+1} + 2^{k+1} - 1,\ 2^{2k+1} - 2^{k+1}),\\
&\Omega_{2k}: (2^{2k+1} - 2^{k+1} - 1,\ 2^{2k+1} + 2^{k+1}).
\end{aligned} \tag{19}$$

The central product of an N-group with any other vee group will amalgamate the Z_2 center of the N-group in the process. We can use (17) and (18) to obtain an expression for the order of the central product as follows:

$$N \otimes G \approx Z_2 \otimes (N \odot G), \quad C(N) = Z_2, \quad C(G) \supset Z_2. \tag{20a}$$

$$(a,b) \otimes (c,d) = (1,0) \otimes \left(\frac{a + c + ac + bd - 1}{2}, \frac{b + d + bc + ad}{2}\right). \tag{20b}$$

The expression on the right-hand side of (20b) gives the order structure of $N \odot G$. In particular, the central product of an N-group with the complex group Z_4 gives the order structure of the S-groups as

$$N \otimes Z_4 \approx Z_2 \otimes (N \odot Z_4) = Z_2 \otimes S, \tag{21a}$$

$$(a,b) \otimes (1,2) = (1,0) \otimes (a + b, a + b + 1). \tag{21b}$$

It is easy to check from (16) and (21) that both $N_{\rm odd}$ and $N_{\rm even}$ give the same S-group via (14). The order structure of the S-groups is, therefore,

$$S_k: (2^{2k+1} - 1,\ 2^{2k+1}). \tag{22}$$

We include an explicit listing of the order structure of the vee groups for the first few cases, which serves as a useful

TABLE IV. Order structure of the finite groups associated with Clifford algebras.

Vee group	Corresponding Clifford algebra	Order of group	Dimension of algebra	Order 2	Structure 4
N_1	real 2-spinors, generalized quaternions	8	2	5	2
N_2	quaternions	8	2	1	6
Ω_1		16	3	11	4
Ω_2	Clifford biquaternions	16	3	3	12
S_1	Pauli algebra, complex quaternions	16	3	7	8
N_3	Eddington–Majorana	32	4	19	12
N_4	Minkowski space-time	32	4	11	20
Ω_3		64	5	39	24
Ω_4		64	5	23	40
S_2	Dirac algebra	64	5	31	32
N_5		128	6	71	56
N_6		128	6	55	72
Ω_5		256	7	143	112
Ω_6		256	7	111	144
S_3		256	7	127	128
N_7		512	8	271	240
N_8		512	8	239	272
Ω_7		1024	9	543	480
Ω_8		1024	9	479	544
S_4	Eddington double E-frame	1024	9	511	512

reference (Table IV). In Ref. 1 we gave a similar table which was calculated from the Clifford algebras directly. (Note that Ref. 1 reverses N_5 and N_6, and does not employ the Ω notation). Table IV can be used to uniquely identify any finite group which arises from a Clifford algebra. This concludes our discussion of the vee groups. The order structure of the vee groups in general has also been computed in Ref. 4.

Comment: For completeness, we note that the construction of Clifford algebras given in Ref. 2 and utilized in this paper is a direct extension of classic results of R. Brauer, E. Cartan, H. Hasse, H. Weyl, and E. Witt. The construction of n-dimensional spinors from tensor products of 2-spinor representations is discussed in Ref. 12. The construction of Clifford algebras as tensor products of quaternion algebras is discussed in Refs. 13 and 14.

V. CONCLUSION

In this paper, the properties of Clifford algebras were studied in terms of the finite groups that they define. This procedure generalizes the usual method of deriving properties of Clifford algebras from manipulations of the representation matrices. That has been the standard practice followed in physics, but it is a method with severe limitations, especially when the dimension of the Clifford algebra is large. Since physical processes are independent of a specific matrix representation, an entirely representation-free formulation should appear as a more elegant and useful alternative to the more traditional methods.

The practical content of this paper is summarized in Table IV. From this table, one can identify any Clifford algebra which arises in a physical application simply by computing its order structure. This calculation may be done using any particular matrix (or other) representation of the algebra. Once the algebra has been identified in Table IV, then one has a set of structural results that the algebra possesses by virtue of belonging to a specific family. All the algebraic properties then follow from the group-theoretical construction of this paper.

ACKNOWLEDGMENT

I am greatly indebted to Dr. R. W. Hartung for stimulating my interest with valuable suggestions, and for substantial assistance in preparing the tables of groups.

[1] N. Salingaros, "Realization, extension, and classification of certain physically important groups and algebras," J. Math.Phys. **22**, 226–232 (1981).

[2] N. Salingaros, "On the classification of Clifford algebras and their relation to spinors in n dimensions," J. Math. Phys. **23**, 1–7; 1231 (1982).

[3] K. C. Hannabuss, "Clifford Algebras and Discrete Groups," preprint, Oxford University.

[4] H. W. Braden, "N-dimensional spinors: Their properties in terms of finite groups," preprint, Cambridge University.

[5] H. S. M. Coxeter and W. O. J. Moser, *Generators and Relations for Discrete Groups* (Springer, New York, 1980),4th ed.

[6] M. Hall, Jr. and J. K. Senior, *The Groups of Order 2^n* (MacMillan, New York, 1964).

[7] L. C. Biedenharn, W. Brouwer, and W. T. Sharp, "The Algebra of Representations of Some Finite Groups," Rice University Studies **54**, No. 2 (Houston, Texas, 1968).

[8] A. D. Thomas and G. V. Wood, *Group Tables* (Shiva, Orpington, Kent, England, 1980).

[9] L. Dornhoff, *Group Representation Theory, Part A* (Marcel Dekker, New York, 1971).

[10] B. Huppert, *Endliche Gruppen I* (Springer, Berlin, 1967).

[11] B. Huppert and N. Blackburn, *Finite Groups II* (Springer, Berlin, 1982). (Note that this reference includes the unit in the involution count.)

[12] R. Brauer and H. Weyl, "Spinors in n-dimensions," Am. J. Math. **57**, 425–449 (1935).

[13] E. Witt, "Theorie der quadratischen Formen in beliebigen Körpern," J. Reine Angew. Math. **176**, 31–44 (1936).

[14] O. T. O'Meara, *Introduction to Quadratic Forms* (Springer, Berlin, 1973).

Infinitesimal operators and structure of the most degenerate representations of the groups Sp($p+q$) and Sp(p,q) in an Sp(p) $\times$ Sp(q) basis

A. U. Klimyk
Institute for Theoretical Physics, Academy of Sciences of the Ukrainian SSR, Kiev-130, USSR

Bruno Gruber
Physics Department, Southern Illinois University, Carbondale, Illinois 62901

(Received 24 January 1983; accepted for publication 25 March 1983)

Explicit expressions are derived for the infinitesimal operators of the representations π_λ of the most degenerate series of the group Sp(p,q) and for the irreducible unitary representations of the group Sp($p+q$) with highest weight ($m, m, 0, ..., 0$) with respect to an Sp(p) $\times$ Sp(q) basis. The Sp(p) $\times$ Sp(q) basis can be chosen arbitrarily. The irreducibility conditions for the representations π_λ and the structure of the reducible representations π_λ are discussed. All the unitary irreducible representations of the most degenerate series of Sp(p,q) are obtained. It is shown that there are four series of unitary representations (principal, supplementary, discrete, and ladder).

PACS numbers: 02.20. + b

I. INTRODUCTION

In this article, explicit expressions are obtained for the infinitesimal operators of the representations of the most degenerate series of Sp(p,q) and of the unitary irreducible representations of Sp($p+q$) which have highest weights (m, m, 0, ..., 0) in an Sp(p) $\times$ Sp(q) basis. Using these infinitesimal operators, we can then investigate the most degenerate series representations π_λ of Sp(p,q). In particular, we have found the irreducibility conditions for the representations π_λ of Sp(p,q). We have defined all the irreducible representations among the set of the representations π_λ. Some of the representations π_λ are reducible. We have found the irreducible representations of Sp(p,q) which are contained in reducible representations of π_λ. Among the set of irreducible representations of Sp(p,q), which are contained in reducible representations π_λ of Sp(p,q), there are finite-dimensional representations. They lead to finite-dimensional irreducible representations of the compact group Sp($p+q$). We have determined all irreducible representations of Sp(p) $\times$ Sp(q) which are contained in the finite-dimensional representations of Sp(p,q) and Sp($p+q$) which have highest weights (m, m, 0, ..., 0). The multiplicities do not exceed 1.

Analyzing the unitarity condition for the irreducible representations π_λ and for the irreducible components of the reducible representations π_λ, we have obtained four series of irreducible unitary representations of Sp(p,q). They represent all the unitary irreducible representations of Sp(p,q) which can be extracted from the irreducible representations of Sp(p,q) which are considered in this article.

In this paper we use the method which was developed by us in Refs. 1–4. The Lie group Sp(p,q) and its algebra sp(p,q) are discussed in Chap. 9 in Ref. 5. These groups can be realized by complex $2(p+q)\times 2(p+q)$ matrices. Let us note that physicists frequently use the notation Sp($2p$, $2q$) and Sp($2p$) for the groups Sp(p,q) and Sp(p). This notation refers to the dimensionality of the matrices. Here we use the notation Sp(p,q) and Sp(p), following Helgason.[5]

For the notation and definitions which we use in the following we refer to Refs. 5 and 6.

II. THE REPRESENTATIONS OF THE MOST DEGENERATE SERIES OF Sp(p,q)

The maximal compact subgroup K of Sp(p,q) is isomorphic to Sp(p) $\times$ Sp(q). It consists of the matrices

diag[Sp(p), Sp(q)].

The Lie algebra $\mathfrak{k} = \mathrm{sp}(p)\times \mathrm{sp}(q)$ of K consists of the matrices[5]

$$\begin{bmatrix} X_{11} & X_{13} & 0 & 0 \\ -\overline{X_{13}} & \overline{X_{11}} & 0 & 0 \\ 0 & 0 & X_{22} & X_{24} \\ 0 & 0 & -\overline{X_{24}} & \overline{X_{22}} \end{bmatrix}, \tag{1}$$

where $X_{11}\in u(p)$, $X_{22}\in u(q)$, and u(p) is the Lie algebra of U(p). Here X_{13} denotes a symmetric $p\times p$ matrix, while X_{24} denote a symmetric $q\times q$ matrix. Further below we shall consider the case for which $p\geqslant q$.

We have the Cartan decomposition $\mathrm{sp}(p,q) = \mathfrak{k} + \mathfrak{p}$, where $\mathfrak{p}$ consists of the matrices[5]

$$\begin{bmatrix} 0 & 0 & Y_{12} & Y_{14} \\ 0 & 0 & \overline{Y_{14}} & -\overline{Y_{12}} \\ \overline{Y}^T_{12} & Y^T_{14} & 0 & 0 \\ \overline{Y}^T_{14} & -Y^T_{12} & 0 & 0 \end{bmatrix}, \tag{2}$$

with Y_{12} and Y_{14} complex $p\times q$ matrices.

We shall now consider the representations of the most degenerate series of Sp(p,q). The representations are induced by one-dimensional representations of the subgroup[4,6]

$$P = ANM(K) = A'N'M'. \tag{3}$$

The subgroups A, N, $M(K)$, A', N', M' are defined in the following manner. The subgroups A and N are defined by the Iwasawa decomposition[6,7] $G = ANK$ of the group $G =$ Sp(p,q). Here A is a commutative subgroup, and N a nilpotent subgroup. The subgroup A' is one-dimensional and $A'\subset A$. We have $A' = \exp \mathfrak{a}'$, where $\mathfrak{a}'$ is the one-dimensional subalgebra of sp(p,q) which is generated by the matrix (2) for

 0022-2488/84/040743-08\$02.50

which $Y_{14}=0$, and

$$Y_{12}=\begin{bmatrix}0 & \cdots & 0 & 0\\ \cdot & \cdots & \cdot & \cdot\\ 0 & \cdots & 0 & 0\\ 0 & \cdots & 0 & 1\end{bmatrix}. \tag{4}$$

We denote this matrix by h_1. Therefore, $A' = \{\exp th_1, t\in R\}$. The subgroup $M(K)$ is a maximal connected subgroup of K, such that $mh = hm$ for every $m\in M(K)$ and every $h\in A'$. It is easy to verify that $M(K)$ consists of the matrices

$$\mathrm{diag}(\mathrm{Sp}(p-1), u, \mathrm{Sp}(q-1), u) \quad u\in \mathrm{Sp}(1).$$

It is clear that the group $M(K)$ is isomorphic to the group $\mathrm{Sp}(p-1)\times\mathrm{Sp}(q-1)\times\mathrm{Sp}(1)$. The subgroup M' is a maximal connected subgroup in $\mathrm{Sp}(p,q)$, such that $mh=hm$ for every $m\in M'$ and every $h\in A'$. We do not need the explicit form of M'. The nilpotent subgroup N' is generated by the root subspaces which correspond to the positive restricted roots of the pair $(\mathrm{sp}(p,q), \mathfrak{a}')$. We have $N'\subset N$.

Let us note that the subgroup A can be represented as $A = A'A''$, where $A'' = \exp \mathfrak{a}''$ and $\mathfrak{a}''$ consists of those matrices of the Lie algebra $\mathfrak{a}$ of A, which are orthogonal to $\mathfrak{a}'$ with respect to the Killing–Cartan bilinear form.[6] Every element $h\in A$ is uniquely decomposed into the product $h'h''$, $h'\in A'$, $h''\in A''$.

Now we consider the one-dimensional representation

$$h'n'm' \to \exp[\lambda(\log h')], \quad h'\in A',\ n'\in N',\ m'\in M', \tag{5}$$

of the subgroup (3), where λ is a complex linear form on $\mathfrak{a}'$. It is clear that

$$\exp[\lambda(th_1)] = \exp \sigma t, \quad \sigma\in C, \tag{6}$$

where $h' = \exp th_1 \in A'$, and $h_1\in\mathfrak{a}'$ is defined above.

The representation (5) of P induces the representation π_λ of $\mathrm{Sp}(p,q)$. This representation acts upon the space $L_0^2(K)$, $K=\mathrm{Sp}(p)\times\mathrm{Sp}(q)$, consisting of functions $f\in L^2(K)$ for which

$$f(mk)=f(k), \quad m\in M(K). \tag{7}$$

The representation π_λ is given by the formula

$$\pi_\lambda(g)f(k) = \exp[\lambda(\log h')]f(k_g), \tag{8}$$

where $h'\in A'$ and $k_g\in K$ are defined by the Iwasawa decomposition of the element kg:

$$kg = hnk_g = h'h''nk_g, \quad h\in A,\ h''\in A'',\ n\in N.$$

It is clear from (6) that π_λ is given by a single complex number σ. In what follows we consider the case that $p\geqslant q>1$. The infinitesimal operators of the representations of $\mathrm{Sp}(p,1)$ and $\mathrm{Sp}(p+1)$ in $\mathrm{Sp}(p)\times\mathrm{Sp}(1)$ basis are considered in Ref. 8.

III. INFINITESIMAL OPERATORS OF THE REPRESENTATIONS π_λ OF $\mathrm{Sp}(p,q)$

In this section we will determine the explicit form for the infinitesimal operators of the representations π_λ of $\mathrm{Sp}(p,q)$. For this construction we shall make use of Lemma 5.2 of Ref. 4. Let $B(.,.)$ be the Killing–Cartan form on $\mathrm{sp}(p,q)$, and let θ denote the Cartan involution on $\mathrm{sp}(p,q)$. Then

$$\langle x,y\rangle = -cB(x,\theta y), \quad c>0, \tag{9}$$

is a scalar product on $\mathrm{sp}(p,q)$. The adjoint representation of $\mathrm{Sp}(p,q)$ on the carrier space $\mathrm{sp}(p,q)$ will be denoted by Ad. For the representations π_λ of $\mathrm{Sp}(p,q)$ Lemma 5.2 of Ref. 4 is formulated in the following manner.

Lemma 1: The infinitesimal operators $\pi_\lambda(Y)$, $Y\in\mathfrak{p}_c$ ($\mathfrak{p}_c$ is a complexification of $\mathfrak{p}$), of the representations π_λ of the group $\mathrm{Sp}(p,q)$ act upon the infinitely differentiable functions of $L_0^2(K)$ in the manner

$$\begin{aligned}\pi_\lambda(Y)f(k) = {} & \langle(\mathrm{Ad}\,k)Y, H\rangle\lambda(H)f(k)\\ & - \langle(\mathrm{Ad}\,k)Y,\rho\rangle f(k)\\ & + \tfrac{1}{2}[Q, \langle(\mathrm{Ad}\,k)Y, h\rangle]f(k),\end{aligned} \tag{10}$$

where H is a normalized element of $\mathfrak{a}'$ with respect to the scalar product (9), h is an element of $\mathfrak{a}'$ such that $\alpha(h)=1$ [α is a simple restricted root of the pair $(\mathrm{sp}(p,q),\mathfrak{a}')$], Q is identical to the operator Q_1 of formula (5) of Ref. 2, ρ denotes half the sum of the positive restricted roots of the pair $(\mathrm{sp}(p,q), \mathfrak{a}')$, including multiple roots, and [.,.] denotes the commutator of the operator Q with the multiplication operator.

In the following we will need an orthonormal basis for $L_0^2(K)$. The space $L^2(K)$ has a basis which consists of all matrix elements of all nonequivalent irreducible representations of $K=\mathrm{Sp}(p)\times\mathrm{Sp}(q)$. The functions of $L_0^2(K)$ satisfy the condition (7). Therefore, all matrix elements of the irreducible representations of K, which satisfy the condition (7), can be taken as a basis of $L_0^2(K)$. From (7) it is clear that the condition (7) is satisfied by the matrix elements of those irreducible representations of K which contain the one-dimensional indentity representation of $M(K)=\mathrm{Sp}(p-1)\times\mathrm{Sp}(q-1)\times\mathrm{Sp}(1)$. We shall find these representations of K. An irreducible representation of $\mathrm{Sp}(p)$ is characterized by p integers $m_1, m_2, ..., m_p$, $m_1\geqslant m_2\geqslant\cdots\geqslant m_p\geqslant 0$. The representations of $\mathrm{Sp}(p)$ with highest weights $(m_1, m_2, 0, ..., 0)$ (and only these) contain the one-dimensional identity representation of $\mathrm{Sp}(p)$ (see §130 in Ref. 9). The restriction of a representation of $\mathrm{Sp}(p)$ with highest weight $(m_1, m_2, 0, ..., 0)$ onto $\mathrm{Sp}(p-1)\times\mathrm{Sp}(1)$ contains the irreducible representation of $\mathrm{Sp}(p-1)\times\mathrm{Sp}(1)$ with highest weight[8] $(0,...,0)_{\mathrm{Sp}(p-1)}(m_1-m_2)_{\mathrm{Sp}(1)}$ with unit multiplicity. Moreover, this restriction does not contain other irreducible representations of $\mathrm{Sp}(p-1)\times\mathrm{Sp}(1)$ which have an identical part for[8] $\mathrm{Sp}(p-1)$ [i.e., the same highest weight $(0,...,0)_{\mathrm{Sp}(p-1)}$].

Therefore, the restriction of the irreducible representation of $K=\mathrm{Sp}(p)\times\mathrm{Sp}(q)$ with highest weight $(m_1, m_2, 0,...,0)$ $(m_1', m_2', 0,...,0)$ onto $K'=[\mathrm{Sp}(p-1)\times\mathrm{Sp}(1)]\times[\mathrm{Sp}(q-1)\times\mathrm{Sp}(1)]$ contains the representation of K' with highest weight $(0,...,0)_{\mathrm{Sp}(p-1)}(m_1-m_2)_{\mathrm{Sp}(1)}(0,...,0)_{\mathrm{Sp}(q-1)}(m_1'-m_2')_{\mathrm{Sp}(1)}$ and does not contain other representations of K' which have identical parts for $\mathrm{Sp}(p-1)$ and $\mathrm{Sp}(q-1)$. The subgroup $M(K)=\mathrm{Sp}(p-1)\times\mathrm{Sp}(q-1)\times\mathrm{Sp}(1)$ is embedded into K' in such a way that $\mathrm{Sp}(1)$ is embedded diagonally into $\mathrm{Sp}(1)\times\mathrm{Sp}(1)$:

$$g\ni\mathrm{Sp}(1)\to g\times g\in\mathrm{Sp}(1)\times\mathrm{Sp}(1).$$

Therefore, the restriction of the representation of K' with highest weight $(0,...,0)(m_1-m_2)(0,...,0)(m_1'-m_2')$ onto $M(K)$ contains the one-dimensional identity representation of $M(K)$ if and only if the tensor product of the representations $(m_1 - m_2)$ and $(m_1' - m_2')$ of Sp(1) contains the one-dimensional identity representations of Sp(1). Since $\mathrm{Sp}(1)\sim \mathrm{SO}(3)$ this implies that this assertion is valid if and only if

$$m_1 - m_2 = m_1' - m_2'.$$

Hence, the one-dimensional identity representation of $M(K)$ is contained only in the representations of $\mathrm{Sp}(p)\times\mathrm{Sp}(q)$ with highest weights $(m_1, m_2, 0,...,0)$ $(m_1', m_2', 0,...,0)$ for which

$$m_1 - m_2 = m_1' - m_2'. \tag{11}$$

Moreover, the multiplicity of this representation of $M(K)$ is 1. We shall denote these representations of $K = \mathrm{Sp}(p)\times\mathrm{Sp}(q)$ by $[m_1, m_2]$ $[m_1', m_2']$.

Let Ω denote an $M(K)$-invariant normalized vector of the carrier space of the representation $[m_1, m_2]$ $[m_1', m_2']$ of K. This vector is unique (up to the multiplier $e^{i\varphi}$). Let $|r\rangle$ denote a set of orthonormal basis elements for this space, $r = 1,2,...,\dim[m_1,m_2]$ $[m_1', m_2']$. We will not restrict the choice of this orthonormal basis. Let $D(k)$ denote the operator which represents k in the representation $[m_1m_2][m_1'm_2']$ of K. Then the functions

$$\begin{aligned}(\dim[m_1, m_2][m_1', m_2'])^{1/2}\langle\Omega|D(k)|r\rangle \\ \equiv(\dim[m_1, m_2][m_1', m_2'])^{1/2}D_{\Omega r}(k) \\ \equiv|m_1, m_2, m_1', m_2', r\rangle\end{aligned} \tag{12}$$

for all r and for all m_1, m_2, m_1', m_2', for which the condition (11) is satisfied, constitute an orthonormal basis for the space $L_0^2(K)$. Namely, the matrix elements (12) (and only these) satisfy the condition (7).

The restriction of π_λ onto K acts upon $L_0^2(K)$ by the formula

$$\pi_\lambda(k_0)f(k) = f(kk_0)$$

Therefore, all functions (12), with fixed m_1, m_2, m_1', m_2', constitute a basis for the subspace on which π_λ realizes the irreducible representation $[m_1, m_2]$ $[m_1', m_2']$ of K. It follows from this that the irreducible representations of K are contained in π_λ not more than once (with unit multiplicity).

We shall now derive the infinitesimal operators $\pi_\lambda(Y)$ in the basis (12). We use Lemma 1 which was formulated above.

The scalar product (9) can be given by the formula

$$\langle X, Y\rangle = \tfrac{1}{4}\,\mathrm{Tr}\,X\overline{Y}^T. \tag{13}$$

The element H of Lemma 1 coincides with the matrix (2) for which $Y_{14} = 0$ and Y_{12} is given by (4). A simple restricted root α of the pair $(\mathrm{sp}(p, q), \mathfrak{a}')$ is given by $\alpha(H) = 1$. Then the element h of Lemma 1 coincides with H. Due to the formula

$$\alpha(h') = \langle h_\alpha, h'\rangle, \quad h'\in\mathfrak{a}',$$

the element $h_\alpha\in\mathfrak{a}'$ corresponds to the root α. It is easy to see that $h_\alpha = H$.

Now we have

$$\langle(\mathrm{Ad}\,k)Y, H\rangle\lambda(H) = \langle(\mathrm{Ad}\,k)Y, h\rangle\lambda(h_\alpha), \tag{14}$$

$$\langle(\mathrm{Ad}\,k)Y, \rho\rangle = \tfrac{1}{2}(r+2s)\langle h_\alpha, h_\alpha\rangle\langle(\mathrm{Ad}\,k)Y, h\rangle, \tag{15}$$

where r denotes the multiplicity of the root α and s the multiplicity[6] of the root 2α. The algebra $\mathrm{sp}(p, q)$ is an algebra of the type CII (see pp. 30–32 in Ref. 6 or Table 3 in Ref. 4). Using the root system of this algebra, we find that

$$\tfrac{1}{2}(r+2s)\langle h_\alpha, h_\alpha\rangle = 2p + 2q - 1. \tag{16}$$

Since the representations π_λ are of the degenerate series, the chain (2) of subgroups of Ref. 2 reduces to

$$\mathrm{Sp}(p)\times\mathrm{Sp}(q) = K \equiv K_1 \supset K_2 = M(K).$$

Moreover, between the subgroups K_1 and K_2 there is one subgroup $K_{\frac{1}{2}}$ [see the chain (3) in Ref. 2)], which is different from K_1 and K_2. The subgroup $K_{\frac{1}{2}}$ can be found in the same way as in the case of the group $\mathrm{U}(p, q)$ in Ref. 2. We find that

$$K_{\frac{1}{2}} = \mathrm{diag}(\mathrm{Sp}(p-1), \mathrm{Sp}(1), \mathrm{Sp}(q-1), \mathrm{Sp}(1)).$$

The subgroups $K\supset K_{\frac{1}{2}}\supset M(K)$ are used to evaluate the eigenvalues of the operator[2] Q. The operator Q acts upon the states (12) as

$$\begin{aligned}&Q\,|m_1, m_2, m_1', m_2', r\rangle \\ &\quad = q(m_1, m_2, m_1', m_2')|m_1, m_2, m_1', m_2', r\rangle,\end{aligned} \tag{17}$$

where $q(\cdots)$ is a number.

According to (10), (14)–(17), it follows that

$$\begin{aligned}&\pi_\lambda(Y)|m_1, m_2, m_1', m_2', r\rangle \\ &\quad = [\lambda(h_\alpha) - 2p - 2q + 1 + \tfrac{1}{2}Q \\ &\qquad - \tfrac{1}{2}q(m_1, m_2, m_1', m_2')] \\ &\qquad \times\langle(\mathrm{Ad}\,k)Y, h\rangle|m_1, m_2, m_1', m_2', r\rangle.\end{aligned} \tag{18}$$

For the elements $Y\in\mathfrak{p}$ we take the basis elements of $\mathfrak{p}$. Since $[\mathfrak{k}, \mathfrak{p}]\subset\mathfrak{p}$, it follows that $\mathfrak{p}$ is a carrier space for the representation of $\mathfrak{k}$ with respect to the action ad b, $b\in\mathfrak{k}$. This representation of $\mathfrak{k} = \mathrm{sp}(p) + \mathrm{sp}(q)$ has the highest weight $(1,0,...,0)$ $(1,0,...,0)$. [The first part is a highest weight for $\mathrm{sp}(p)$, the second part for $\mathrm{sp}(q)$.] We choose an orthonormal basis of $\mathfrak{p}$ with respect to the scalar product (13). Let $Y_1, Y_2, ..., Y_n$, $n = \dim\mathfrak{p}$, denote the elements of this basis.

The expression $\langle(\mathrm{Ad}\,k)Y_j, h\rangle$ of (18) is a matrix element of the representation of $\mathrm{Sp}(p)\times\mathrm{Sp}(q)$ with highest weight $(1,0,...,0)$ $(1,0,...,0)$. Since $|m_1, m_2, m_1', m_2', r\rangle$ is also a matrix element of the representation of K [see (12)], we have

$$\begin{aligned}&\langle(\mathrm{Ad}\,k)Y, h\rangle|m_1, m_2, m_1', m_2', r\rangle \\ &\quad = \sum_{\substack{n_1,n_2\\ n_1'n_2'}}\sum_{r'}\left(\frac{\dim[m_1,m_2][m_1',m_2']}{\dim[n_1,n_2][n_1', n_2']}\right)^{1/2} \\ &\qquad\times\langle m_1m_2m_1'm_2', \Omega; (1,0,...,0)(1,0,...,0), h\,|n_1n_2n_1'n_2', \Omega\rangle \\ &\qquad\times\langle n_1n_2n_1'n_2', r'|m_1m_2m_1'm_2', r; (1,0,...,0)(1,0,...,0), Y_j\rangle \\ &\qquad\times|n_1n_2, n_1', n_2', r'\rangle,\end{aligned} \tag{19}$$

where $\langle\cdots|\cdots\rangle$ denotes the Clebsch–Gordan coefficients for the tensor product of the representations $\mathrm{Sp}(p)\times\mathrm{Sp}(q)$ with highest weights $(m_1,m_2,0,...,0)$ $(m_1',m_2',0,...,0)$ and $(1,0,...,0)$ $(1,0,...,0)$. Since the element $h\in\mathfrak{p}$ is invariant with respect to $M(K)$, it follows that the same Ω occurs twice in the first Clebsch–Gordan coefficient on the right-hand side of (19).

Let us find the values of n_1, n_2, n_1', n_2' over which the summation in (19) is to be carried out. The tensor product of the representations of $\mathrm{Sp}(p)$ with highest weights $(m_1,m_2,0,\ldots,0)$ and $(1,0,\ldots,0)$ is decomposed into irreducible representations by making use of the same formula as is done for the representations of $\mathrm{SO}(2p)$ with the highest weights $(m_1,m_2,0,\ldots,0)$ and $(1,0,\ldots,0)$ [see, for example, formula (4.5) in Ref. 4]. We have also to take into account that due to the first Clebsch–Gordan coefficient of (19) the integers n_1, n_2, n_1', n_2' obey the condition

$$n_1 - n_2 = n_1' - n_2'.$$

Therefore, the summation in (19) is over the following values of (n_1, n_2, n_1', n_2'):

$$\begin{gathered}(m_1+1, m_2, m_1'+1, m_2), (m_1, m_2+1, m_1', m_2'+1), (m_1-1, m_2, m_1'-1, m_2'),\\ (m_1, m_2-1, m_1', m_2'-1), (m_1+1, m_2, m_1', m_2'-1), (m_1-1, m_2, m_1', m_2'+1),\\ (m_1, m_2+1, m_1'-1, m_2'), (m_1, m_2-1, m_1'+1, m_2').\end{gathered} \tag{20}$$

Now we have to substitute (19) into (18) and to evaluate the numbers

$$\tfrac{1}{2}q(n_1, n_2, n_1', n_2') - \tfrac{1}{2}q(m_1, m_2, m_1', m_2') \tag{21}$$

for all values of (n_1, n_2, n_1', n_2') which are enumerated in (20). The numbers (21) are to be evaluated by means of the formula (50) of Ref. 2. The reader also finds in this paper the definitions and notation which is used in formula (50) of Ref. 2. For our present case it holds $\Lambda^1 = (m_1, m_2, m_1', m_2')$, $\Lambda^2 = (m_1 - m_2, m_1' - m_2')$. The values which one obtains in evaluating formula (50) of Ref. 2 for all (n_1, n_2, n_1', n_2') enumerated in (20) are given in Table I.

Using Table I, formulas (18) and (20), we find that

$$\begin{aligned}\pi_\lambda(Y_j)|m_1, m_2, m_1', m_2', r\rangle = {} & \sum_{r'} (\sigma + M + M')K^{m_1+1,m_2}_{m_1'+1,m_2'}(r',j)|m_1+1, m_2, m_1'+1, m_2', r'\rangle\\ & + \sum_{r'} (\sigma + M + M')K^{m_1,m_2+1}_{m_1',m_2'+1}(r',j)|m_1, m_2+1, m_1', m_2'+1, r'\rangle\\ & + \sum_{r'} (\sigma - M - M' - 4p - 4q + 4)K^{m_1-1,m_2}_{m_1'-1,m_2}(r',j)|m_1-1, m_2, m_1'-1, m_2', r'\rangle\\ & + \sum_{r'} (\sigma - M - M' - 4p - 4q + 4)K^{m_1,m_2-1}_{m_1',m_2'-1}(r',j)|m_1, m_2-1, m_1', m_2'-1, r'\rangle\\ & + \sum_{r'} (\sigma + M - M' - 4q + 2)K^{m_1+1,m_2}_{m_1',m_2'-1}(r',j)|m_1+1, m_2, m_1', m_2'-1, r'\rangle\\ & + \sum_{r'} (\sigma - M + M' - 4p + 2)K^{m_1-1,m_2}_{m_1',m_2'+1}(r',j)|m_1-1, m_2, m_1', m_2'+1, r'\rangle\\ & + \sum_{r'} (\sigma + M - M' - 4q + 2)K^{m_1,m_2,+1}_{m_1'-1,m_2'}(r',j)|m_1, m_2+1, m_1'-1, m_2', r'\rangle\\ & + \sum_{r'} (\sigma - M + M' - 4p + 2)K^{m_1,m_2-1}_{m_1'+1,m_2'}(r',j)|m_1, m_2-1, m_1'+1, m_2', r'\rangle,\end{aligned} \tag{22}$$

TABLE I.

	$\Lambda^1+\tau^1$	$\langle\Lambda^1+R^1,\tau^1\rangle$	$\frac{1}{2}\langle\tau^1,\tau^1\rangle$	$\Lambda^2+\tau^2$	$\frac{1}{2}\langle\Lambda^2+R^2,\tau^2\rangle$	$\frac{1}{4}\langle\tau^2,\tau^2\rangle$
1	$m_1+1, m_2,$ $m_1'+1, m_2'$	$2(m_1+m_1'+p+q)$	2	$m_1-m_2+1,$ $m_1'-m_2'+1$	$(m_1-m_2+m_1'-m_2'+2)$	1
2	m_1, m_2+1 $m_1', m_2'+1$	$2(m_2+m_2'+p+q-2)$	2	$m_1-m_2-1,$ $m_1'-m_2'-1$	$(-m_1+m_2-m_1'+m_2'-2)$	1
3	$m_1-1, m_2,$ $m_1'-1, m_2'$	$2(-m_1-m_1'-p-q)$	2	$m_1-m_2-1,$ $m_1'-m_2'-1$	$(-m_1+m_2-m_1'+m_2'-2)$	1
4	m_1, m_2-1 $m_1', m_2'-1$	$2(-m_2-m_2'-p-q+2)$	2	$m_1-m_2+1,$ $m_1'-m_2'+1$	$(m_1-m_2+m_1'-m_2'+2)$	1
5	$m_1+1, m_2,$ $m_1', m_2'-1$	$2(m_1-m_2'+p-q+1)$	2	m_1-m_2+1 $m_1'-m_2'+1$	$(m_1-m_2+m_1'-m_2'+2)$	1
6	$m_1-1, m_2,$ $m_1', m_2'+1$	$2(-m_1+m_2'-p+q-1)$	2	m_1-m_2-1 $m_1'-m_2'-1$	$(-m_1+m_2-m_1'+m_2'-2)$	1
7	$m_1, m_2+1,$ $m_1'-1, m_2'$	$2(m_2-m_1'+p-q-1)$	2	m_1-m_2-1 $m_1'-m_2'-1$	$(-m_1+m_2-m_1'+m_2'-2)$	1
8	m_1, m_2-1 $m_1'+1, m_2'$	$2(-m_2+m_1'-p+q+1)$	2	m_1-m_2+1 $m_1'-m_2'+1$	$(m_1-m_2+m_1'-m_2'+2)$	1

where

$$\sigma = \lambda(h_\alpha), \quad M = m_1 + m_2, \quad M' = m_1' + m_2',$$

and $K^{n_1 n_2}_{n_1' n_2'}(r', j)$ is a coefficient of $|n_1, n_2, n_1', n_2', r\rangle$ on the right-hand side of (19). In (22) the summation over r' is defined by the nonzero Clebsch–Gordan coefficients in (19); these Clebsch–Gordan coefficients are contained in $K^{n_1 n_2}_{n_1' n_2'}(r'j)$.

The integers m_1, m_2, m_1', m_2' are dependent. Namely, they satisfy the relation (11). Therefore, instead of m_1, m_2, m_1', m_2', we can introduce the integers

$$M = m_1 + m_2, \quad M' = m_1' + m_2', \quad J = m_1 - m_2. \tag{23}$$

Then we obtain for the functions (12) the notation

$$|m_1, m_2, m_1', m_2, r\rangle \equiv |J, M, M', r\rangle. \tag{24}$$

For the representations $[m_1, m_2][m_1', m_2']$ of $\mathrm{Sp}(p) \times \mathrm{Sp}(q)$, which are contained in the representation π_λ of $\mathrm{Sp}(p, q)$, the values of J, M, M' run over all nonnegative integers such that either all of them are even or all of them are odd. This follows from the equalities

$$J = M - 2m_2, \quad J = M' - 2m_2'.$$

Therefore, the sum $M + M'$ (and $M - M'$) is even.

Thus we have obtained the infinitesimal operators $\pi_\lambda(Y_j)$ which correspond to the elements of $\mathfrak{p}$. In order to obtain all infinitesimal operators of the representation π_λ, we must also have the infinitesimal operators for irreducible representations of $K = \mathrm{Sp}(p) \times \mathrm{Sp}(q)$. These were found in Ref. 8.

Let us note that the expressions for the infinitesimal operators (22) are valid for any $\mathrm{Sp}(p) \times \mathrm{Sp}(q)$ basis. The basis $|m_1, m_2, m_1', m_2', r\rangle$ in (22) coincides with the basis $|m_1, m_2, m_1', m_2', r\rangle$ used for the second Clebsch–Gordan coefficient on the right-hand side of (19). The Clebsch–Gordan coefficients for the most interesting $\mathrm{Sp}(p) \times \mathrm{Sp}(q)$ bases will be investigated in a forthcoming paper.

IV. STRUCTURE OF THE REPRESENTATIONS π_λ OF $\mathrm{Sp}(p, q)$

In this section we determine the irreducible representations π_λ which are contained among the set of all representations π_λ of $\mathrm{Sp}(p, q)$, and, moreover, we investigate the structure of the reducible representations π_λ.

Lemma 2: A representation π_λ of $\mathrm{Sp}(p, q)$ is irreducible if for every representation $[m_1, m_2][m_1', m_2']$ of $\mathrm{Sp}(p) \times \mathrm{Sp}(q)$, which is contained in π_λ, no one of the numbers

$$\sigma + M + M', \quad \sigma - M - M' - 4p - 4q + 4,$$
$$\sigma - M + M' - 4p + 2, \quad \sigma + M - M' - 4q + 2,$$

[the coefficients on the right-hand side of (22)] is equal to zero.

This lemma is proven in exactly the same manner as statement 7.1 in Ref. 4. We omit the proof.

Theorem 1: A representation π_λ of $\mathrm{Sp}(p, q)$ is irreducible if and only if σ is not an even integer.

The irreducibility of a representation π_λ for which $\sigma = \lambda(h_\alpha)$ is not an even integer, follows from Lemma 2. The reducibility of a representations π_λ, for which σ is an even integer, will be shown below.

Let us consider the representations π_λ for which σ is an even integer. Below the representations π_λ will be denoted by π^σ, where $\sigma = \lambda(h_\alpha)$. It is known[4] that for real σ the representations π^σ and $\pi^{-\sigma + 4p + 4q - 2}$ are either both reducible or both are irreducible. Moreover, the irreducible representations π^σ and $\pi^{-\sigma + 4p + 4q - 2}$ are equivalent, and the reducible representations π^σ and $\pi^{-\sigma + 4p + 4q - 2}$ contain the same irreducible representations of $\mathrm{Sp}(p, q)$ [see Chapter 5 in Ref. 4]. Therefore, it is sufficient to consider the representations π^σ with $\sigma < 2p + 2q - 1$.

Let us investigate the structure of the representtions π^σ with even σ, $\sigma < 2p + 2q - 1$. In order to do this, we set the coefficients in the parenthesis of the right-hand side of (22) equal to zero:

$$\sigma + M + M' = 0, \tag{25}$$

$$\sigma - M - M' - 4p - 4q + 4 = 0, \tag{26}$$

$$\sigma + M - M' - 4q + 2 = 0, \tag{27}$$

$$\sigma - M + M' - 4p + 2 = 0. \tag{28}$$

Since $\sigma < 2p + 2q - 1$, the relation (26) cannot be satisfied. We shall consider the relations (25), (27), (28). We distinguish three cases.

Case 1: The integer σ is even and $\sigma \leqslant 0$. In Fig. 1 we plot the points $(M = m_1 + m_2, M' = m_1' + m_2')$ which correspond to the representations $[m_1, m_2][m_1', m_2']$ of $\mathrm{Sp}(p) \times \mathrm{Sp}(q)$ which are contained in π^σ. If we plot Eqs. (25), (27), and (28) in Fig. 1, then it is seen that the points (M, M') are divided into four regions, namely D^F, D^0, D^+, D^-. The points which are located on a straight line belong to that region from which the arrow points to the line.

Let us now consider the points (M, M') of D^F. It is seen from (22) that the operators $\pi^\sigma(Y_j)$ can increase the sum $M + M'$ in a vector $|J, M, M', r\rangle$ only in the first and in the second summands of the right-hand side of (22). These summands vanish for $M + M' = -\sigma$. Therefore, the operators $\pi^\sigma(Y_j)$ cannot transform vectors $|J, M, M', r\rangle$ with $(M, M') \in D^F$ into vectors $|J, M, M', r\rangle$ with $(M, M') \in D^0$ (see Fig. 1). Hence, all the vectors $|J, M, M', r\rangle$ with $(M, M') \in D^F$

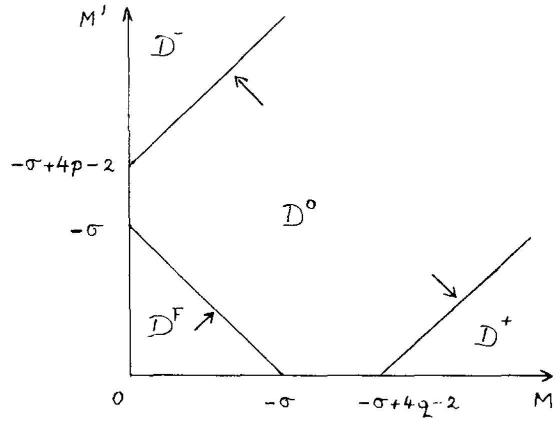


FIG. 1. Figure for Case 1.

constitute the basis for an invariant subspace of $L_0^2(K)$ with respect to sp(p,q). The representation π^σ realizes a subrepresentation on this subspace. We denote this subrepresentation by D_σ^F. Let us remark that the operators $\pi^\sigma(Y_j)$ can transform vectors $|J, M, M', r\rangle$ with $(M, M')\in D^0$ into vectors $|J, M, M', r\rangle$ with $(M, M')\in D^F$. This follows from (22). This means that D_σ^F is not contained in π^σ as a direct summand.

Let us now consider the points (M, M') from $D^F + D^0$. One can see from (22) that the operators $\pi^\sigma(Y_j)$ increase $M - M'$ in the vector $|J,M,M',r\rangle$ only in the fifth and in the seventh summand of the right-hand side of (22), and $M' - M$ only in the sixth and in the eighth summand. The fifth and the seventh summands vanish for $M - M' = -\sigma + 4q - 2$. The sixth and the eighth summands vanish for $M' - M = -\sigma + 4p - 2$. Therefore (see Fig. 1), the operators $\pi^\sigma(Y_j)$ cannot transform the vectors $|J, M, M', r\rangle$ with $(M, M')\in D^F + D^0$ into vectors $|J, M, M', r\rangle$ with $(M, M')\in D^+ + D^-$. Hence, the representation π^σ of Sp(p,q) realizes on the vectors $|J, M, M', r\rangle$ with $(M, M')\in D^F + D^0$ a subrepresentation, which we denote by D_σ^{0F}. The quotient representation D_σ^{0F}/D_σ^F will be denoted by D_σ^0. The quotient representation π^σ/D_σ^{0F} thus decomposes into a direct sum of representations of Sp(p,q), which correspond to the vectors $|J,M,M'r\rangle$ with $(M,M')\in D^+$ and to the vectors $|J,M,M'r\rangle$ with $(M, M')\in D^-$. We denote these representations by D_σ^+ and D_σ^-.

Thus for case 1 the representation π^σ leads to four representations $D_\sigma^F, D_\sigma^0, D_\sigma^+, D_\sigma^-$ of Sp(p,q). Irreducible representations of Sp$(p)\times$Sp(q) are contained in these representations of Sp(p,q) at most once (i.e., with unit multiplicity).

The representations $D_\sigma^F, D_\sigma^0, D_\sigma^+, D_\sigma^-$ of Sp(p,q) are irreducible. The proof of this irreducibility is analogous to the proof of irreducibility of the representations of U$(n, 1)$ and SO$_0(n, 1)$ which are contained in the principal nonunitary series of these groups (see §2 in Chap. 7 and §7 in Chap. 8 in Ref. 4). We omit the proof.

The representations π^σ for Case 1 have thus the following structure:

$$\begin{bmatrix} D_\sigma^F & * & 0 & 0 \\ 0 & D_\sigma^0 & * & * \\ 0 & 0 & D_\sigma^+ & 0 \\ 0 & 0 & 0 & D_\sigma^- \end{bmatrix},$$

where $*$ denotes a nonzero matrix.

The representation D_σ^F is finite-dimensional. Its highest weight is $(-\sigma/2, -\sigma/2, 0, ..., 0)$. This follows from Theorem 5.13a of Ref. 4. The considerations given above permit us to find all highest weights of the irreducible representations of $K = $ Sp$(p)\times$Sp(q), which are contained in the restriction of the representation D_σ^F of Sp(p,q) [and, therefore, of Sp$(p+q)$] onto Sp$(p)\times$Sp(q). These irreducible representations of Sp$(p)\times$Sp(q) have highest weights $(m_1, m_2, 0,...,0)$ $(m_1', m_2', 0,...,0)$ for which $m_1 - m_2 = m_1' - m_2'$ (therefore, $m_1 + m_2 + m_1' + m_2'$ is even) and

$$m_1 + m_2 + m_1' + m_2' \leqslant -\sigma.$$

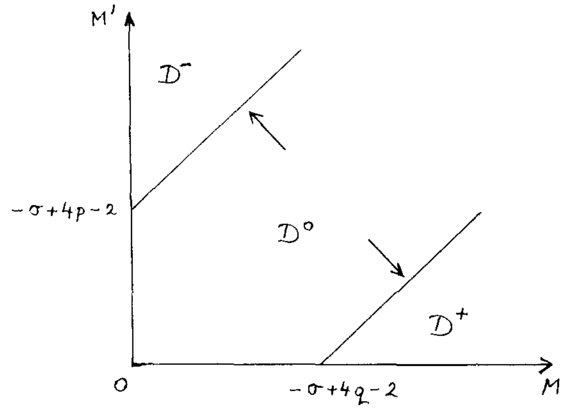


FIG. 2. Figure for Case 2.

These representations of Sp$(p)\times$Sp(q) are contained in D_σ^F with multiplicity one.

Case 2: The integer σ is even and $0<\sigma<2p+2q-2$. The considerations for this case are the same as for Case 1. For this case the relation (25) cannot be satisfied. Equations (27) and (28) are shown in Fig. 2. The points (M, M') are separated into three parts D^0, D^+, D^-. On the subspace for which the vectors $|J, M, M', r\rangle$, $(M, M')\in D^0$ form a basis a subrepresentation of π^σ is realized. We denote it by D_σ^0. The quotient representation π^σ/D_σ^0 decomposes into a direct sum of two representations of Sp(p,q). One of these corresponds to the points $(M, M')\in D^+$ and will be denoted by D_σ^+; the other corresponds to the points $(M, M')\in D^-$ and will be denoted by D_σ^-. It is easily observed that the representation $D_\sigma^0, D_\sigma^+, D_\sigma^-$ are irreducible. The representation π^σ for Case 2 has thus the structure

$$\begin{bmatrix} D_\sigma^0 & * & * \\ 0 & D_\sigma^+ & 0 \\ 0 & 0 & D_\sigma^- \end{bmatrix}. \tag{29}$$

Case 3: $\sigma = 2p + 2q - 2$: For this case the relation (25) cannot be satisfied. The relations (27) and (28) become identical. This case is shown in Fig. 3. In fact this case coincides with Case 2; the difference is that in this case the points of D^0 are located on straight line only. As in Case 2, we obtain three irreducible representations $D_\sigma^0, D_\sigma^+, D_\sigma^-$ of Sp(p,q) and the representation π^σ has the structure (29).

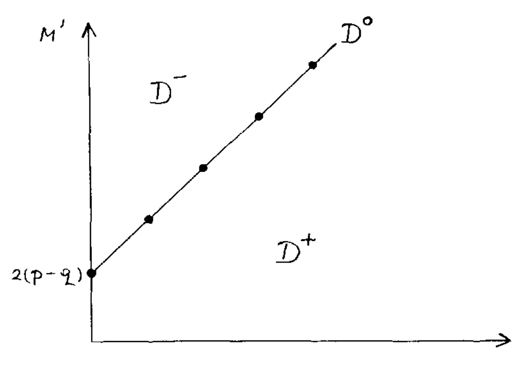


FIG. 3. Figure for Case 3.

V. MOST DEGENERATE UNITARY REPRESENTATIONS OF $\mathrm{Sp}(p,q)$

We have constructed the irreducible representations π^σ (σ is not an even integer), $D^0_\sigma, D^+_\sigma, D^-_\sigma$ of $\mathrm{Sp}(p,q)$. Let us now determine those among them which admit unitarization. According to the results of Refs. 10 and 11 (see also §4 of Chap. 5 in Ref. 4), this can be done in the following manner. We construct the intertwining operators $\Pi(\sigma)$ for all pairs $(\pi^\sigma, \pi^{-\sigma+4p+4q-2})$. Then we determine for which representations the intertwining operator is Hermitian and positive (or negative) definite. These representations can be made unitary. For details we refer to Refs. 10 and 11. The procedure is equivalent to the following steps. We construct the intertwining operators $\Pi(\sigma)$, and then we introduce in the formulas (22) the new basis

$$|J,M,M',r\rangle' = \Pi(\sigma)^{-1/2}|J,M,M',r\rangle. \tag{30}$$

Representations, which admit unitarization, are unitary in the basis (30). Now we have to verify which ones of the representations are unitary, i.e., to verify for which representations the unitarity condition $\pi(Y_j)^* = -\pi(Y_j)$ is fulfilled in the basis (30).

The intertwining operator $\Pi(\sigma)$ for the pair $(\pi^\sigma, \pi^{-\sigma+4p+4q-2})$ is defined as

$$\pi^\sigma\Pi(\sigma) = \Pi(\sigma)\pi^{-\sigma+4p+4q-2}. \tag{31}$$

For the matrix elements of $\Pi(\sigma)$ we have[4,10]

$$\langle J,M,M',r|\Pi(\sigma)|J_1,M_1,M'_1,r_1\rangle = a(J,J,M')\delta_{JJ_1}\delta_{MM_1}\delta_{M'_1M'}\delta_{rr_1}. \tag{32}$$

Let us now consider the relation (31) for the infinitesimal operators of π^σ. Taking matrix elements on both sides of (31), we obtain from (32) and (22) the following system of equations for the matrix elements $a(J,M,M')$:

$$(\sigma+M+M')a(J,M,M') = (-\sigma+M+M'+4p+4q-2)a(J+1,M+1,M'+1),$$
$$(\sigma+M-M'-4q+2)a(J,M,M') = (-\sigma+M-M'+4p)a(J+1,M+1,M'-1),$$
$$(\sigma-M+M'-4p+2)a(J,M,M') = (-\sigma-M+M'+4q)a(J-1,M-1,M'+1),$$
$$(\sigma-M-M'-4p-4q+4)a(J,M,M') = (-\sigma-M-M'+2)a(J-1,M-1,M'-1).$$

It follows from that $a(J,M,M')$ does not depend on J, and that

$$a(M+i,M'+i) = \prod_{j=0}^{i-1}\frac{\sigma+M+M'+2j}{-\sigma+M+M'+4p+4q+2(j-1)}\,a(M,M'), \tag{33}$$

$$a(M+i,M'-i) = \prod_{j=0}^{i-1}\frac{\sigma+M-M'-4q+2(j+1)}{-\sigma+M-M'+4p+2j}\,a(M,M'), \tag{34}$$

$$a(M-i,M'+i) = \prod_{j=0}^{i-1}\frac{\sigma-M+M'-4p+2(j+1)}{-\sigma-M+M'+4q+2j}\,a(M,M'), \tag{35}$$

$$a(M-i,M'-i) = \prod_{j=1}^{i}\frac{\sigma-M-M'-4p-4q+2(j+1)}{-\sigma-M-M'+2j}\,a(M,M'). \tag{36}$$

If we fix $a(M,M')$ for fixed values $M=M_0$, $M'=M'_0$, then we obtain from (33)–(36) the $a(M,M')$ for all M and M'.

Theorem 2: Among the set of irreducible representations π^σ (σ is not an even integer), D^0_σ, D^+_σ, and D^-_σ of $\mathrm{Sp}(p,q)$ the following representations admit unitarization:

(1) the representations π^σ, for which $\sigma-2p-2q+1$ are purely imaginary (principal most degenerate unitary series),

(2) the representations π^σ, $2p+2q-2<\sigma<2p+2q$ (most degenerate supplementary series),

(3) all the representations D^+_σ and D^-_σ (most degenerate discrete series),

(4) the representation $D^0_{2p+2q-2}$ (ladder representation).

The unitarity of the representations of class (1) follows from (22). Direct verification shows that the unitarity condition $\pi(Y_j)^* = -\pi(Y_j)$ is fulfilled for them. In order to obtain unitary representations for classes (2)–(4), we have to use the new basis (30) in (22). This basis is constructed by means of (33)–(36). In the new basis we have

$$\begin{aligned}
\pi^\sigma(Y_j)|J,M,M',r\rangle = &\sum_{r'}[(\sigma+M+M')(-\sigma+M+M'+4p+4q-2)]^{1/2}K^{m_1+1,m_2}_{m'_1+1,m_2}|J+1,M+1,M'+1,r'\rangle\\
&+\sum_{r'}[(\sigma+M+M')(-\sigma+M+M'+4p+4q-2)]^{1/2}K^{m_1,m_2+1}_{m'_1,m'_2+1}|J-1,M+1,M'+1,r'\rangle\\
&-\sum_{r'}[(\sigma-M-M'-4p-4q+4)(-\sigma-M-M'+2)]^{1/2}K^{m_1-1,m_2}_{m'_1-1,m'_2}|J-1,M-1,M'-1,r'\rangle\\
&-\sum_{r'}[(\sigma-M-M'-4p-4q+4)(-\sigma-M-M'+2)]^{1/2}K^{m_1,m_2-1}_{m'_1,m'_2-1}|J+1,M-1,M'-1,r'\rangle\\
&+\sum_{r'}[(\sigma+M-M'-4q+2)(-\sigma+M-M'+4p)]^{1/2}K^{m_1+1,m_2}_{m'_1,m'_2-1}|J+1,M+1,M'-1,r'\rangle\\
&-\sum_{r'}[(\sigma-M+M'-4p+2)(-\sigma-M+M'+4q)]^{1/2}K^{m_1-1,m_2}_{m'_1,m'_2+1}|J-1,M-1,M'+1,r'\rangle\\
&+\sum_{r'}[(\sigma+M-M'-4q+2)(-\sigma+M-M'+4p)]^{1/2}K^{m_1,m_2+1}_{m'_1-1,m'_2}|J-1,M+1,M'-1,r'\rangle\\
&-\sum_{r'}[(\sigma-M+M'-4p+2)(-\sigma-M+M'+4q)]^{1/2}K^{m_1,m_2-1}_{m'_1+1,m'_2}|J+1,M-1,M'+1,r'\rangle.
\end{aligned} \tag{37}$$

Now one can verify directly that the representations of classes (2)–(4) of Theorem 2 satisfy the unitarity condition $\pi(Y_j)^* = -\pi(Y_j)$ in the basis (30).

We want to point out that some of the unitary representations of Theorem 2 were constructed in Ref. 12.

VI. INFINITESIMAL OPERATORS OF UNITARY IRREDUCIBLE REPRESENTATIONS OF Sp($p+q$) IN AN Sp(p)×Sp(q) BASIS

It was pointed out in Sec. 4 that the representations D_σ^F of Sp(p, q) lead to irreducible representations of Sp($p+q$) with highest weights $(-\sigma/2, -\sigma/2, 0,...,0)$. The infinitesimal operators for these representations of Sp($p+q$) can be obtained in the following manner. The Lie algebra Sp(p, q) of Sp(p, q) has the Cartan decomposition

$$\mathrm{sp}(p,q) = \mathfrak{k} + \mathfrak{p}, \quad \mathfrak{k} = \mathrm{sp}(p) + \mathrm{sp}(q).$$

The compact Lie algebra sp($p+q$) corresonds to sp(p, q), which has the decomposition

$$\mathrm{sp}(p+q) = \mathfrak{k} + i\mathfrak{p}, \quad i = \sqrt{-1}.$$

Therefore, multiplying the infinitesimal operators $Y_j \in \mathfrak{p}$ of the representations D_σ^F of Sp(p, q) by $\sqrt{-1}$, we obtain the infinitesimal operators $X_j = \sqrt{-1}\,Y_j \in \sqrt{-1}\,\mathfrak{p}$ of the representations D_σ^F of Sp($p+q$) with highest weights $(-\sigma/2, -\sigma/2, 0,...,0)$. The infinitesimal operators X_j do not satisfy the unitarity condition $X_j^* = -X_j$ if they are obtained from (22). To satisfy the unitarity condition, we have to use the basis (30). As a consequence of this, we obtain the formula for the infinitesimal operators X_j of the irreducible unitary representations of Sp($p+q$) with highest weights $(m,m,0,...,0)$, $m \geqslant 0$ and integer, from formula (37) by replacing $\pi^\sigma(Y_j)$ by X_j, namely,

$$[(\sigma+M+M')(-\sigma+M+M'+4p+4q-2)]^{-1/2} \quad \text{by} \quad -[(2m-M-M')(2m+M+M'+4p+4q-2)]^{1/2},$$

$$[(\sigma-M-M'-4p-4q+4)(-\sigma-M-M'+2)]^{-1/2} \quad \text{by} \quad -[(2m-M-M'+2)(2m+M+M'+4p+4q-4)]^{1/2},$$

$$[(\sigma+M-M'-4q+2)(-\sigma+M-M'+4p)]^{-1/2} \quad \text{by} \quad -[(2m-M+M'+4q-2)(2m+M-M'+4p)]^{1/2},$$

$$[(\sigma-M+M'-4p+2)(-\sigma-M+M'+4q)]^{-1/2} \quad \text{by} \quad -[(2m+M-M'+4p-2)(2m-M+M'+4q]^{1/2}.$$

The resulting formula satisfies the unitary condition $X_j^* = -X_j$.

[1]A. U. Klimyk and A. M. Gavrilik, J. Math. Phys. **20**, 1624 (1979).
[2]A. U. Klimyk and B. Gruber, J. Math. Phys. **20**, 1995 (1979).
[3]B. Gruber and A. U. Klimyk, J. Math. Phys. **22**, 2762 (1981).
[4]A. U. Klimyk, *Matrix Elements and Clebsch–Gordan Coefficients of Group Representations* (Naukova Dumka, Kiev, 1979) (in Russian).
[5]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962).
[6]G. Warner, *Harmonic Analysis on Semisimple Lie Groups* (Springer-Verlag, Berlin, 1972), Vol. 1.
[7]A. O. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN, Warsaw, 1977).
[8]P. I. Holod and A. U. Klimyk, Rep. Math. Phys. (to be published).
[9]D. P. Zhelobenko, *Compact Lie Groups and Their Representations* (Nauka, Moscow, 1970) (in Russian).
[10]A. W. Knapp and E. M. Stein, Ann. Math. **93**, 489 (1971).
[11]A. W. Knapp and E. M. Stein, Invent. Math. **60**, 9 (1980).
[12]P. Pajas, J. Math. Phys. **10**, 1777 (1969).

Realizations of F_4 in SO(3) ⊗ SO(3) bases and structural zeros of the 6 j-symbol

H. De Meyer,[a] G. Vanden Berghe, and J. Van der Jeugt[b]
Seminarie voor Wiskundige Natuurkunde, R. U. G., Krijgslaan 281–S9, B-9000 Gent, Belgium

(Received 8 February 1983; accepted for publication 29 April 1983)

The decompositon of F_4 irreps in the group chains $F_4 \supset SO(3) \otimes G_2 \supset SO(3) \otimes SO(3)$ and $F_4 \supset SO(3) \otimes Sp(6) \supset SO(3) \otimes SO(3)$ is considered. In both cases a realization of the F_4 Lie algebra in terms of SO(3) ⊗ SO(3) tensor operators is established. From these, the nontrivial vanishing of certain 6 j-coefficients is explained.

PACS numbers: 02.20.Qs, 03.65.Fd

I. INTRODUCTION

It has been known for a long time that the standard SO(3) tensor operator formulation of the exceptional Lie algebra G_2 provides an explanation for the nontrivial vanishing of Racah's 6 j-coefficient $\{ {}^{5}_{3}\, {}^{5}_{3}\, {}^{3}_{3} \}$. Many others of the nontrivial or structural zeros of the 6 j-symbol, which do not result from triangle condition violation, have been tabulated in a recent book of Biedenharn and Louck.[1] Moreover, it has been suggested by these authors that the exploration of other exceptional Lie algebras might lead to an analogous explanation of some more of these zeros.

A realization of the by rank next higher exceptional algebra F_4 in terms of SO(3) tensor operators has already been constructed by Wadzinski.[2] Unfortunately, no zeros follow from it, since his solution makes the SO(9) subalgebra contained in F_4 explicit and, therefore, too many SO(3) representation space bases (or, equivalently, in that case too many bosons) are needed. Instead, it has been shown by the present authors[3] that in the maximal decomposition of F_4 into SO(3), a minimal realization in terms of two such bases exists from which two structural zeros can be explained. On account of Regge symmetries, even more zeros followed.

Besides SO(9) and SO(3), F_4 contains three more maximal subalgebras, i.e., SO(3) ⊗ G_2, SO(3) ⊗ Sp(6), and SU(3) ⊗ SU(3). By inspection of branching rule tables such as those of Mckay and Patera,[4] it is readily seen that the chain $F_4 \supset SU(3) \otimes SU(3) \supset SO(3) \otimes SO(3)$ is not a good candidate for explaining structural zeros of the 6 j-symbol, since both the lowest dimensional and the adjoint irrep of F_4 decompose into SO(3) ⊗ SO(3) irreps with degeneracy. On the other hand, the chains $F_4 \supset SO(3) \otimes G_2 \supset SO(3) \otimes SO(3)$ and $F_4 \supset SO(3) \otimes Sp(6) \supset SO(3) \otimes SO(3)$ will prove in the present paper to be an excellent basis for the explanation of such zeros.

In Sec. II, some relevant formulas concerning SO(3) ⊗ SO(3) tensor operators and their use in realizing a classical Lie algebra are reviewed. Sections III and IV, respectively, deal with the two inclusion chains in the order mentioned above. It is clarified how structural zeros appear in the analysis. Finally, the reader should be aware that throughout, the labeling conventions of McKay and Patera[4] are followed, with the exception of SO(3) irreps which we label by half the number they use.

II. SO(3) ⊗ SO(3) TENSOR OPERATORS

An SO(3) ⊗ SO(3) tensor operator is the outer product of two SO(3) tensor operators, each acting upon separate SO(3) representation spaces. Hence we can define such an operator by means of reduced matrix elements[2,5]:

$$\begin{aligned}\langle \tau'_b l'_b;\tau_b l_b \| v^{k'}(\tau'_2 l'_2,\tau'_1 l'_1)\cdot v^{k}(\tau_2 l_2,\tau_1 l_1) \| \tau'_a l'_a;\tau_a l_a \rangle &= \langle \tau'_b l'_b \| v^{k'}(\tau'_2 l'_2,\tau'_1 l'_1) \| \tau'_a l'_a \rangle \langle \tau_b l_b \| v^{k}(\tau_2 l_2,\tau_1 l_1) \| \tau_a l_a \rangle \\ &= \{[k'][k]\}^{1/2} \delta_{\tau'_a\tau'_1} \delta_{\tau'_b\tau'_2} \delta_{\tau_a\tau_1} \delta_{\tau_b\tau_2} \delta_{l'_a l'_1} \delta_{l'_b l'_2} \delta_{l_a l_1} \delta_{l_b l_2}. \end{aligned} \tag{2.1}$$

Herein l and k are SO(3) representation labels which assume integral or half-odd integral values, $[k] = 2k + 1$ is the dimension of the SO(3) tensor representation, and τ is an additional label to distinguish irreps with the same l. The heavy dot and semicolon are used to indicate the outer product of tensors and SO(3) irreps, respectively. The SO(3) ⊗ SO(3) tensor operators satisfy the following commutation relations:

[a] Research Associate N.F.W.O., Belgium.
[b] Research Assistant N.F.W.O., Belgium.

 0022-2488/84/040751-04$02.50

$$
\begin{aligned}
&\left[v^{k'_1}_{q'_1}(\tau'_1 l'_1,\tau'_2 l'_2)\cdot v^{k_1}_{q_1}(\tau_1 l_1,\tau_2 l_2), v^{k'_2}_{q'_2}(\tau'_3 l'_3,\tau'_4 l'_4)\cdot v^{k_2}_{q_2}(\tau_3 l_3,\tau_4 l_4)\right] \\
&\quad = \sum_{k'_3 k_3 q'_3 q_3} \{[k'_1][k_1][k'_2][k_2][k'_3][k_3]\}^{1/2} \begin{pmatrix} k'_1 & k'_2 & k'_3 \\ q'_1 & q'_2 & -q'_3 \end{pmatrix} \begin{pmatrix} k_1 & k_2 & k_3 \\ q_1 & q_2 & -q_3 \end{pmatrix} \\
&\qquad \times (-1)^{2l'_4 + 2l_4 + l'_3 + l_3 - l'_2 - l_2 - q'_3 - q_3} \Big[(-1)^{K+K'+L+L'} \delta_{\tau'_2\tau'_3}\delta_{\tau_2\tau_3}\delta_{l'_2 l'_3}\delta_{l_2 l_3} \\
&\qquad \times \begin{Bmatrix} k'_1 & k'_2 & k'_3 \\ l'_4 & l'_1 & l'_3 \end{Bmatrix} \begin{Bmatrix} k_1 & k_2 & k_3 \\ l_4 & l_1 & l_3 \end{Bmatrix} v^{k'_3}_{q'_3}(\tau'_1 l'_1,\tau'_4 l'_4)\cdot v^{k_3}_{q_3}(\tau_1 l_1,\tau_4 l_4) \\
&\qquad - \delta_{\tau'_1\tau'_4}\delta_{\tau_1\tau_4}\delta_{l'_1 l'_4}\delta_{l_1 l_4} \begin{Bmatrix} k'_1 & k'_2 & k'_3 \\ l'_3 & l'_2 & l'_1 \end{Bmatrix} \begin{Bmatrix} k_1 & k_2 & k_3 \\ l_3 & l_2 & l_1 \end{Bmatrix} v^{k'_3}_{q'_3}(\tau'_3 l'_3,\tau'_2 l'_2)\cdot v^{k_3}_{q_3}(\tau_3 l_3,\tau_2 l_2) \Big],
\end{aligned}
\tag{2.2}
$$

with $K = k_1 + k_2 + k_3$ and $L = l_1 + l_2 + l_3 + l_4$.

If A denotes an irrep of a classical Lie group H, we learn from tables[4] how A decomposes into irreps $(\gamma' k';\gamma k)$ of a particular SO(3) ⊗ SO(3) subgroup, γ being used to distinguish again between similar irreps. Hence the SO(3) ⊗ SO(3) tensor operators which constitute an operator realization of A can be labeled by $(\gamma' k' q';\gamma k q)$ where $q'(q)$ runs in unit steps from $-k'$ to k' $(-k$ to $k)$. In order to define a space spanned by product states $|\tau' l' m';\tau l m\rangle$ in which these operators act, we select an irrep B of H for which B occurs in the decomposition of the Kronecker product[2] $A \times B$. If B decomposes into irreps $(\tau' l';\tau l)$, the set $\{|\tau' l' m';\tau l m\rangle : m' = -l', -l'+1, \ldots, l'; m = -l, -l+1, \ldots, l\}$ is an acceptable basis. Since it is our aim to construct realizations of a Lie algebra, A is identified to the adjoint representation of H. Moreover, it is advantageous as well for computational purposes as for making the probability of finding nontrivial zeros maximal, to choose the dimension of B as low as possible. Without making the choice of B explicit, the generators of H can be written as

$$
G^{\gamma' k';\gamma k}_{q';q} = \sum_{\tau'_1 l'_1 \tau'_2 l'_2} \sum_{\tau_1 l_1 \tau_1 l_2} g\left[{\gamma' k';\tau'_1 l'_1,\tau'_2 l'_2 \atop \gamma k;\tau_1 l_1,\tau_2 l_2} \right] v^{k'}_{q'}(\tau'_1 l'_1,\tau'_2 l'_2)\cdot v^{k}_{q}(\tau_1 l_1,\tau_2 l_2),
\tag{2.3}
$$

whereby the g-coefficients remain to be determined. To do so, we use the property that the set of generators is closed under commutation, which allows us to write

$$
\begin{aligned}
\left[G^{\gamma'_2 k'_2;\gamma_2 k_2}_{q'_2;q_2}, G^{\gamma'_1 k'_1;\gamma_1 k_1}_{q'_1;q_1} \right] &= (-1)^{q_1 + q_2 + q'_1 + q'_2} \\
&\quad \times \sum_{k' k \gamma' \gamma} \{[k'][k][k'_1][k_1][k'_2][k_2]\}^{1/2} \begin{pmatrix} k'_2 & k'_1 & k' \\ q'_2 & q'_1 & -(q'_1 + q'_2) \end{pmatrix} \\
&\quad \times \begin{pmatrix} k_2 & k_1 & k \\ q_2 & q_1 & -(q_1 + q_2) \end{pmatrix} C^{\gamma' k';\gamma k}_{\gamma'_2 k'_2,\gamma'_1 k'_1;\gamma_2 k_2,\gamma_1,k_1} G^{\gamma' k';\gamma k}_{q'_1 + q'_2;q_1 + q_2},
\end{aligned}
\tag{2.4}
$$

where the C-coefficients remain to be determined. Substitution of (2.3) in both sides of (2.4) leads with the application of (2.2) to the following equations with respect to the unknown g- and C-coefficients:

$$
\begin{aligned}
&C^{\gamma' k';\gamma k}_{\gamma'_2 k'_2,\gamma'_1 k'_1;\gamma_2 k_2,\gamma_1 k_1}\, g\left[{\gamma' k';\tau'_2 l'_2,\tau'_1 l'_1 \atop \gamma k;\tau_2 l_2,\tau_1,l_1} \right] \\
&\quad = (-1)^{l'_2 + l_2 - l'_1 - l_1} \sum_{\tau' j' \tau j} \Big\{ (-1)^{K+K'+2j+2j'} \begin{Bmatrix} k'_2 & k'_1 & k' \\ l'_1 & l'_2 & j' \end{Bmatrix} \begin{Bmatrix} k_2 & k_1 & k \\ l_1 & l_2 & j \end{Bmatrix} \\
&\qquad \times g\left[{\gamma'_2 k'_2;\tau'_2 l'_2,\tau' j' \atop \gamma_2 k_2;\tau_2 l_2,\tau j} \right] g\left[{\gamma'_1 k'_1;\tau' j',\tau'_1 l'_1 \atop \gamma_1 k_1;\tau j,\tau_1 l_1} \right] - \begin{Bmatrix} k'_1 & k'_2 & k' \\ l'_1 & l'_2 & j' \end{Bmatrix} \begin{Bmatrix} k_1 & k_2 & k \\ l_1 & l_2 & j \end{Bmatrix} g\left[{\gamma'_1 k'_1;\tau'_2 l'_2,\tau' j' \atop \gamma_1 k_1;\tau_2 l_2,\tau j} \right] g\left[{\gamma'_2 k'_2;\tau' j',\tau'_1 l'_1 \atop \gamma_2 k_2;\tau j,\tau_1 l_1} \right] \Big\},
\end{aligned}
\tag{2.5}
$$

where now $K = k + k_1 + k_2$. If we let all parameters herein vary, we generate a usually overcomplete system of equations linear in the C's and quadratic in the g's. Substituting the solution for the latter coefficients in (2.3) we obtain an operator realization of the group generators. In the following sections, we apply the present technique to F_4 in two different decompositions. Since in neither case degeneracies occur, we systematically drop the labels γ and τ.

III. $F_4 \supset$ SO(3) ⊗ $G_2 \supset$ SO(3) ⊗ SO(3)

Using standard tables,[4] it is verified that in the present chain of reductions, the F_4 adjoint representation (1 0 0 0) decomposes into the SO(3) ⊗ SO(3) irreps (0;1), (0;5), (1;0), and (2;3), whereas the lowest-dimensional F_4 irrep (0 0 0 1) decomposes into (1;3) and (2;0). By means of the procedure outlined at the end of Sec. II, we obtained, after straightforward calculation, the following realization of the F_4 algebra in terms of SO(3) ⊗ SO(3) tensor operators:

$$
\begin{aligned}
&G^{0;1}_{0;q} = v^0_0(1,1)\cdot v^1_q(3,3), \quad G^{0;5}_{0;q} = v^0_0(1,1)\cdot v^5_q(3,3), \\
&G^{1;0}_{q';0} = v^1_{q'}(1,1)\cdot v^0_0(3,3) + (\sqrt{5}/\sqrt{7}) v^1_{q'}(2,2)\cdot v^0_0(0,0), \\
&G^{2;3}_{q';q} = v^2_{q'}(1,1)\cdot v^3_q(3,3) + ((-1)^\alpha/\sqrt{2}) [v^2_{q'}(1,2)\cdot v^3_q(3,0) + v^2_{q'}(2,1)\cdot v^3_q(0,3)].
\end{aligned}
\tag{3.1}
$$

Herein α is a free parameter. Obviously, the generators $G^{0;1}$, $G^{0;5}$, and $G^{1;0}$ of (3.1) form together the SO(3) ⊗ G_2 subalgebra of F_4.

We next demonstrate how the basis (3.1) can be exploited to explain some structural zeros of the $6j$-symbol. Let us first consider the commutator $[G^{0;5}, G^{2;3}]$ in which we substitute the explicit forms (3.1) for the generators. On account of formula (2.2), the commutator can then be reexpressed as a linear combination of SO(3) ⊗ SO(3) tensor operators. Since the algebra (3.1) is closed, it should be possible to recombine the tensor operators in terms of F_4 generators. However, from (2.2) it is clear that one can generate tensor operators which do not even occur in (3.1). Hence the coefficients with which these operators appear have to vanish. Two cases have to be distinguished: either an unwanted tensor is formed more times with different coefficients, or an unwanted tensor can appear only one or more times with the same coefficient. In the first case, there results a relation between products of $6j$-coefficients; in the second case a product of two $6j$-coefficients has to vanish. If then one of these is obviously different from zero, the remaining one has to vanish and we will have explained a structural zero. When this program is executed on the commutator mentioned above, it turns out that a single term proportional to

$$\begin{Bmatrix} 0 & 2 & 2 \\ 1 & 1 & 1 \end{Bmatrix} \begin{Bmatrix} 5 & 3 & 5 \\ 3 & 3 & 3 \end{Bmatrix} v^2(1,1)\cdot v^5(3,3)$$

remains, by which is explained the structural zero $\begin{Bmatrix} 5 & 3 & 5 \\ 3 & 3 & 3 \end{Bmatrix} = 0$ which has already been found in the chain $G_2 \supset SO(3)$. This is not really a surprise since the generator $G^{0;5}$ in the commutator under consideration behaves as a scalar (rank zero tensor) with respect to the SO(3) part of the SO(3) ⊗ G_2 subalgebra. Similarly, in the commutator $[G^{0;5}, G^{0;5}]$, there remains a term proportional to

$$\begin{Bmatrix} 2 & 2 & 0 \\ 1 & 1 & 1 \end{Bmatrix} \begin{Bmatrix} 5 & 5 & 3 \\ 3 & 3 & 3 \end{Bmatrix} v^0(1,1)\cdot v^3(3,3),$$

which is absent in (3.1). The same structural zero is found again.

The previous reasonings show that new nontrivial zeros are most likely to appear in the commutator $[G^{2;3}, G^{2;3}]$. And, indeed, it turns out that a single term proportional to

$$\begin{Bmatrix} 2 & 2 & 3 \\ 2 & 2 & 1 \end{Bmatrix} \begin{Bmatrix} 3 & 3 & 0 \\ 0 & 0 & 3 \end{Bmatrix} v^3(2,2)\cdot v^0(0,0),$$

which originates from the commutators $[v^2(1,2)\cdot v^3(3,0), v^2(2,1)\cdot v^3(0,3)]$ and $[v^2(2,1)\cdot v^3(0,3), v^2(1,2)\cdot v^3(3,0)]$ survives. Hence we have explained the following structural zero of the $6j$-symbol,

$$\begin{Bmatrix} 2 & 2 & 3 \\ 2 & 2 & 1 \end{Bmatrix} = 0. \tag{3.2}$$

Since this $6j$-coefficient is, under Regge symmetry operations, only related to itself, the result (3.2) is the only new zero connected to the chain $F_4 \supset SO(3) \otimes G_2 \supset SO(3) \otimes SO(3)$. On the other hand, also, relations between products of $6j$-symbols follow from that chain, but in the context of the present paper it is not worthwhile to report on them explicitly. Also, they are independent of the choice of the q-subscripts and this is why we constantly omitted that label. Finally, it should be remarked that the result (3.2) has not been used in setting up (3.1), so that the vanishing of the $6j$-coefficient (3.2) is really a consequence of the fact that (3.1) closes under commutation.

IV. $F_4 \supset$ SO(3) ⊗ Sp(6) $\supset$ SO(3) ⊗ SO(3)

In this chain, the adjoint representation of F_4 reduces into (0;1), (0;3), (0;5), (1;0), $(\frac{1}{2};\frac{3}{2})$, and $(\frac{1}{2};\frac{9}{2})$, whereas the 26-dimensional irrep decomposes into (0;2), (0;4), and $(\frac{1}{2};\frac{5}{2})$. A similar but lengthier calculation than in the previous case now leads to the following SO(3) ⊗ SO(3) tensor operator realization of F_4:

$$\begin{aligned}
G^{0;1}_{0;q} &= v^0_0(0,0)\cdot v^1_q(2,2) \\
&\quad + \sqrt{6}v^0_0(0,0)\cdot v^1_q(4,4) + (\sqrt{7}/\sqrt{2})v^0_0(\tfrac{1}{2},\tfrac{1}{2})\cdot v^1_q(\tfrac{5}{2},\tfrac{5}{2}), \\
G^{0;3}_{0;q} &= v^0_0(0,0)\cdot v^3_q(2,2) + (\sqrt{11}/9)v^0_0(0,0)\cdot v^3_q(4,4) \\
&\quad + (-1)^\alpha(5\sqrt{2}/3\sqrt{3})v^0_0(0,0)\cdot[v^3_q(2,4) + v^3_q(4,2)] \\
&\quad - \tfrac{14}{9}\, v^0_0(\tfrac{1}{2},\tfrac{1}{2})\cdot v^3_q(\tfrac{5}{2},\tfrac{5}{2}), \\
G^{0;5}_{0;q} &= v^0_0(0,0)\cdot v^5_q(4,4) - (-1)^\alpha(\sqrt{15}/\sqrt{26})v^0_0(0,0) \\
&\quad \times [v^5_q(2,4) + v^5_q(4,2)] - (\sqrt{14}/\sqrt{13})v^0_0(\tfrac{1}{2},\tfrac{1}{2})\cdot v^5_q(\tfrac{5}{2},\tfrac{5}{2}), \\
G^{1;0}_{q';0} &= v^1_{q'}(\tfrac{1}{2},\tfrac{1}{2})\cdot v^0_0(\tfrac{5}{2},\tfrac{5}{2}), \\
G^{1/2;3/2}_{q';q} &= v^{1/2}_{q'}(\tfrac{1}{2},0)\cdot v^{3/2}_q(\tfrac{5}{2},2) + v^{1/2}_{q'}(0,\tfrac{1}{2})\cdot v^{3/2}_q(2,\tfrac{5}{2}) \\
&\quad + (-1)^\alpha(\sqrt{2}/\sqrt{5})[v^{1/2}_{q'}(\tfrac{1}{2},0)\cdot v^{3/2}_q(\tfrac{5}{2},4) \\
&\quad + v^{1/2}_{q'}(0,\tfrac{1}{2})\cdot v^{3/2}_q(4,\tfrac{5}{2})], \\
G^{1/2;9/2}_{q';q} &= v^{1/2}_{q'}(\tfrac{1}{2},0)\cdot v^{9/2}_q(\tfrac{5}{2},2) + v^{1/2}_{q'}(0,\tfrac{1}{2})\cdot v^{9/2}_q(2,\tfrac{5}{2}) \\
&\quad + (-1)^\alpha(\sqrt{11}/\sqrt{3})[v^{1/2}_{q'}(\tfrac{1}{2},0)\cdot v^{9/2}_q(\tfrac{5}{2},4) \\
&\quad + v^{1/2}_{q'}(0,\tfrac{1}{2})\cdot v^{9/2}_q(4,\tfrac{5}{2})].
\end{aligned} \tag{4.1}$$

Again α is a free parameter, and now one recognizes the SO(3) ⊗ Sp(6) subalgebra by restricting (4.1) to the set of generators $G^{0;1}$, $G^{0;3}$, $G^{0;5}$, and $G^{1;0}$.

In the same way as before, all the commutators between the F_4 generators (4.1) must be systematically analyzed. There is one commutator, namely $[G^{1/2;9/2}, G^{1/2;9/2}]$, by which a new structural zero of the $6j$-symbol can be explained. Indeed, from the commutators $[v^{1/2}(\frac{1}{2},0)\cdot v^{9/2}(\frac{5}{2},4), \cdot v^{1/2}(0,\frac{1}{2})\cdot v^{9/2}(4,\frac{5}{2})]$, $[v^{1/2}(0,\frac{1}{2})\cdot v^{9/2}(4,\frac{5}{2}), v^{1/2}(\frac{1}{2},0)\cdot v^{9/2}(\frac{5}{2},4)]$ originates on account of (2.2) a single term proportional to

$$\begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{Bmatrix} \begin{Bmatrix} \frac{9}{2} & \frac{9}{2} & 7 \\ 4 & 4 & \frac{5}{2} \end{Bmatrix} v^0(0,0)\cdot v^7(4,4),$$

which is not present in the algebra (4.1) and, therefore, has to vanish. Hence the zero

$$\begin{Bmatrix} \frac{9}{2} & \frac{9}{2} & 7 \\ 4 & 4 & \frac{5}{2} \end{Bmatrix} = 0 \tag{4.2}$$

finds an explanation within the chain $F_4 \supset SO(3) \otimes Sp(6) \supset SO(3) \otimes SO(3)$. On account of a Regge symmetry operation carried out on the $6j$-coefficient (4.2) at the same time the zero

$$\begin{Bmatrix} \frac{13}{2} & 5 & \frac{9}{2} \\ 2 & \frac{9}{2} & 4 \end{Bmatrix} = 0 \tag{4.3}$$

becomes explained. Finally, as before, we will not insist here on the many relations between products of $6j$-symbols which result from the present treatment of the F_4 algebra.

V. DISCUSSION

The present paper closes our study of F_4 in the context of explaining structural zeros of the $6j$-symbol. At this stage, it has been shown that F_4 yields eleven such zeros, of which three have been reported on in this paper. The other eight which have been explained[3] in the chain $F_4 \supset SO(3)$ are

$$\begin{Bmatrix} 11 & 11 & 3 \\ 4 & 4 & 8 \end{Bmatrix} = \begin{Bmatrix} 11 & 10 & 2 \\ 4 & 5 & 9 \end{Bmatrix} = 0,$$

$$\begin{Bmatrix} 11 & 11 & 9 \\ 8 & 4 & 8 \end{Bmatrix} = \begin{Bmatrix} 12 & 11 & 8 \\ 5 & 8 & 7 \end{Bmatrix} = \begin{Bmatrix} 11 & 10 & 10 \\ 4 & 9 & 7 \end{Bmatrix}$$
$$= \begin{Bmatrix} 13 & 9 & 9 \\ 6 & 8 & 6 \end{Bmatrix} = \begin{Bmatrix} 13 & 10 & 8 \\ 6 & 7 & 7 \end{Bmatrix}$$
$$= \begin{Bmatrix} 12 & 10 & 9 \\ 5 & 9 & 6 \end{Bmatrix} = 0.$$

A new aspect of the procedure of realizing algebras which has been clarified here is the usefulness of direct products of the familiar SO(3) tensor operators. They are interesting by the way they reduce, for a given algebra, the number of tensors and at the same time the number of different representation space bases to be taken into consideration. These are precisely conditions which favor the finding of nontrivial zeros.

On the other hand, we may not have exhausted F_4 completely yet. Indeed, since we have restricted our attention to minimal representations and to the short reduction chains ending at a single SO(3) or at a direct product of SO(3) groups, extensions can still be envisaged. However, although not zero, the probability of explaining nontrivial zeros becomes very small.

ACKNOWLEDGMENTS

We would like to express our gratitude to Professor L. Biedenharn of Duke University for his encouraging comments at an early stage of this research project.

[1] L. C. Biedenharn and J. D. Louck, *The Racah–Wigner Algebra in Quantum Theory*; *Encyclopedia of Mathematics and its Applications* (Addison-Wesley, London, 1981), Vol. 9, p. 415.
[2] H. T. Wadzinski, Nuovo Cimento B **62**, 247 (1969).
[3] J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Phys. A: Math. Gen. **16**, 1377 (1983).
[4] M. G. McKay and J. Patera, *Tables of dimensions, indices and branching rules for representations of simple Lie algebras*; *Lecture Notes in Pure and Applied Mathematics* (Marcel Dekker, New York, 1981), Vol. 69.
[5] B. R. Judd, *Operator techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963), p. 151.

Matrix elements for indecomposable representations of complex su(2)

Bruno Gruber
Physics Department, Southern Illinois University, Carbondale, Illinois 62901
Anatoli U. Klimyk
Institute for Theoretical Physics, Academy of Sciences of the Ukrainian SSR, Kiev-130, USSR

(Received 20 October 1982; accepted for publication 6 May 1983)

Indecomposable representations of the simple complex Lie algebra A_1 are investigated in this article from a general point of view. First a "master representation" is obtained which is defined on the space of the universal enveloping algebra Ω of A_1. Then, from this master representation other indecomposable representations are derived which are induced on quotient spaces or subduced on invariant subspaces. Finally, it is shown that the familiar finite-dimensional and infinite-dimensional irreducible representations of su(2) and su(1,1) are closely related to certain of the indecomposable representations. Indecomposable representations of A_1 [su(2), su(1,1)] have found increased applications in physical problems, including the unusual "finite multiplicity" indecomposable representations. Emphasis is placed in this article on an analysis of the more unfamiliar indecomposable representations. The matrix elements are obtained in *explicit form* for all representations which are discussed in this article. The methods used are purely algebraic.

PACS numbers: 02.20.Qs

INTRODUCTION

It is the purpose of this article to study unusual representations of the simple complex Lie algebra A_1, which is the complexification of the physically significant Lie algebras su(2) and su(1,1).

Originally it was only the algebra su(2), and its *finite*-dimensional irreducible representations, which was of relevance to physical applications. The original application of su(2) was in atomic physics, and it was Wigner who pioneered this work.[1] The theory behind the application of su(2) to physical problems became known as angular momentum theory.

The noncompact real simple Lie algebra su(1,1), and its *infinite*-dimensional irreducible representations, entered physics at a later date. It was Bargmann[2] who did a systematic study of su(1,1) [actually the group SU(1,1) and its infinite-dimensional irreducible representations]. As in the case of su(2), many authors have discussed the irreducible infinite-dimensional representations of su(1,1) and it is beyond the scope of this introduction to attempt to list even the most important articles. We want, however, to mention the work of Gel'fand, Graev, and Vilenkin[3] and that of Holman and Biedenharn,[4] since they had a profound influence on our thinking.

Finally, it is indecomposable representations (i.e., reducible, but not completely reducible representations) of A_1 which have found physical applications. Chacón, Levi, and Moshinsky[5] have shown that the so-called Armstrong tensors[6] (the radial part of the multipole operators in atomic physics) transform like indecomposable representations of A_1 [or su(1,1)]. Representations of this type were discussed in a general manner, by Gel'fand, Graev, and Vilenkin,[3] while the authors of this article obtained the matrix elements of these representations in explicit form.[7] In this article the authors also demonstrated the existence of an unfamiliar type of indecomposable representations of A_1 [or su(1,1)], which they called *finite-multiplicity* indecomposable representations. These representations have the property that some of their weight subspaces have dimension > 1 (in fact 2). It turns out that these representations are also found in physical applications. Flato and Fronsdal[8] have, in their investigation of singletons, made use of indecomposable representations of the (de Sitter) algebra so(3,2) whose restriction to a certain su(1,1) subalgebra contains finite multiplicity indecomposable representations.

Certain types of indecomposable representations have actually been known in physics for a long time. They are associated with the Poincaré algebra, the algebra of the Euclidean group, and others. We will not discuss these algebras and their representations since we limit our discussion in this article to the algebra A_1. We refer the interested reader to Ref. 9.

In view of the proliferation in physical applications of what was once considered by physicists as unfamiliar and unusual representations, it appears to be warranted to investigate the (indecomposable) representations of A_1 in a general manner. The methods used are purely algebraic and follow along the lines of thought as presented by Dixmier[10] and Jacobson.[11]

While *irreducible* representations of A_1 have been discussed extensively, relative little is known regarding *indecomposable* representations of A_1. In fact the few indecomposable representations which have been discussed are of a rather special type. They are defined on "one-parameter families of states," i.e., they can be realized algebraically on the quotient spaces Ω_+, Ω_-, $\Omega_+ + \Omega_-$ of the universal enveloping algebra Ω of A_1. Indecomposable representations of this type were discussed in Ref. 3, 12, 13, and 14.

In this article we obtain the matrix elements for all representations which are discussed in *explicit form*. Again, very little is known regarding the matrix elements, and with the exception of the indecomposable representations discussed in Refs. 12, 13, and 14, the matrix elements for the indecomposable representations are obtained here for the

 0022-2488/84/040755-10$02.50

first time. Finally, some of the representations, finite- as well as infinite-dimensional, are quite familiar, and all is known about them. These will occur as special cases in our investigation.

In this article we will use the term indecomposable to mean that a representation is reducible, but not completely reducible (i.e., the representation space has invariant subspaces, but is not the direct sum of invariant subspaces).

The representations of A_1 which are discussed in this article are defined on the space of its universal enveloping algebra Ω, on ideals of Ω, on invariant subspaces, and on quotient spaces. We define a very general representation of A_1 on all of Ω, called "master representation," and derive from it other indecomposable representations, until we obtain the familiar representations of su(2) and su(1,1) as simple special cases.

1. REPRESENTATIONS OF su(2) AND su(1,1) ON THE SPACE OF THEIR UNIVERSAL ENVELOPING ALGEBRA

In this section representations of the simple complex Lie algebra A_1 (Cartan notation) are investigated on the space of the universal enveloping algebra Ω of A_1. The algebra A_1 is the complex extension of both, the compact algebra su(2) and the noncompact algebra su(1,1) Thus having obtained an algebraic representation of A_1, an algebraic representation of su(2) and su(1,1) has been obtained too. These latter representations are obtained from a representation of A_1 by means of restriction of the field to the reals $\mathbb{R}$ (and possibly by a change of basis within the complex algebra A_1, unless the basis employed in A_1 happens to correspond to the basis of the real form).

Certain representations of su(1,1) on the space Ω were investigated in Ref. 7. The present article takes a much more general point of view, and the results which were obtained in Ref. 7 will appear as special cases of the results obtained in this article. It should be noted, however, that both the basis chosen for A_1 and for Ω in this article differ from the one which was chosen in Ref. 7. In this article we prefer to choose the more familiar (to the physicist) "angular momentum" basis of su(2) in A_1, while in Ref. 7 we chose a basis which "corresponds" to the noncompact form su(1,1). Moreover, in this article we choose a basis for Ω by ordering the elements l_+, l_-, l_3 in the sequence $l_+l_-l_3$. Simple symmetry considerations will relate the results obtained in different bases. In particular, the basis chosen for A_1 in this article is related to the basis chosen for A_1 in Ref. 7 by the substitution

$$l_3 \to h = l_3, \quad l_+ \to e = il_+, \quad l_- \to f = il_-, \quad i = \sqrt{-1}.$$

The algebra A_1 can be defined by choosing the basis

$$A_1: \{l_3, l_+, l_-\}, \tag{1.1}$$

with the Lie products given by

$$[l_3, l_\pm] = \pm l_\pm, \quad [l_+, l_-] = l_3, \tag{1.2}$$

and the general element $l \in A_1$,

$$l = c_3 l_3 + c_+ l_+ + c_- l_-, \quad c_3, c_\pm \in \mathbb{C}. \tag{1.3}$$

The algebra A_1 has the Casimir element

$$C = 2l_+l_- + l_3(l_3 - 1).$$

A basis for universal enveloping algebra Ω of A_1 can be chosen as

$$\Omega: \{\mathbb{1}, l_+^n l_-^m l_3^r;\ n, m, r = 0, 1, 2, \ldots\}, \tag{1.4}$$

where $\mathbb{1}$ denotes the identity operator and n, m, r are not simultaneously zero. For reason of convenience we will later on frequently represent the identity operator Ω by the values $n = m = r = 0$. The elements of Ω are then all the finite linear combinations over the basis elements given by Eq. (1.4) [the product in Eq. (1.4) is the tensor product].[10,11]

The space Ω can be factored into a product of three subspaces,

$$\Omega = \Omega_+ \Omega_- \mathscr{H}, \tag{1.5}$$

where the bases for the three subspaces can be chosen as

$$\begin{aligned} \Omega_+&: \quad \{\mathbb{1}, l_+^n,\ n = 1, 2, 3, \cdots\}, \\ \Omega_-&: \quad \{\mathbb{1}, l_-^m,\ m = 1, 2, 3, \cdots\}, \\ \mathscr{H}&: \quad \{\mathbb{1}, l_3^r,\ r = 1, 2, 3, \cdots\}. \end{aligned} \tag{1.6}$$

A linear representation of A_1 on the space Ω (or on one of its subspaces V) is a linear map of Ω (or of the subspace V) into itself, such that the Lie product is satisfied. The following Lie products, within the universal enveloping algebra, will be needed in order to calculate the matrix elements for the various representations:

$$\begin{aligned} [l_3, l_\pm^n] &= \pm n l_\pm^n, \\ [l_-, l_+^n] &= -n(l_+^{n-1} l_3 + \tfrac{1}{2}(n-1) l_+^{n-1}). \end{aligned} \tag{1.7}$$

In the following the most general algebraic representation of A_1 on its universal enveloping algebra Ω is obtained. This representation is subsequently made use of to derive from it other representations which are defined on various (invariant) subspaces of Ω, on quotient spaces of Ω with respect to invariant subspaces, on quotient spaces of (invariant) subspaces of Ω with respect to invariant subspaces of the invariant subspaces, etc., leading down all the way to the familiar finite-dimensional irreducible representations of angular momentum theory su(2) and the familiar irreducible infinite-dimensional representations, bounded on one side or unbounded on both sides, of su(1,1). We say that a representation of su(1,1) is bounded if the operator l_3 can be diagonalized and the set of its eigenvalues is bounded.[7]

2. REPRESENTATIONS ON Ω

Making use of Eq. (1.7), it follows that the relations hold:

$$\begin{aligned} \rho(l_3) l_+^n l_-^m l_3^r &= l_+^n l_-^m l_3^{r+1} + (n-m) l_+^n l_-^m l_3^r, \\ \rho(l_+) l_+^n l_-^m l_3^r &= l_+^{n+1} l_-^m l_3^r, \\ \rho(l_-) l_+^n l_-^m l_3^r &= l_+^n l_-^{m+1} l_3^r - n l_+^{n-1} l_-^m l_3^{r+1} \\ &\quad + n[m - \tfrac{1}{2}(n-1)] l_+^{n-1} l_-^m l_3^r, \\ \rho(C) l_+^n l_-^m l_3^r & \\ &= 2 l_+^{n+1} l_-^{m+1} l_3^r + l_+^n l_-^m l_3^{r+2} \\ &\quad - (2m+1) l_+^n l_-^m l_3^{r+1} + m(m+1) l_+^n l_-^m l_3^r. \end{aligned} \tag{2.1}$$

It is easy to prove that the map ρ forms a representation of A_1 on Ω, i.e., that

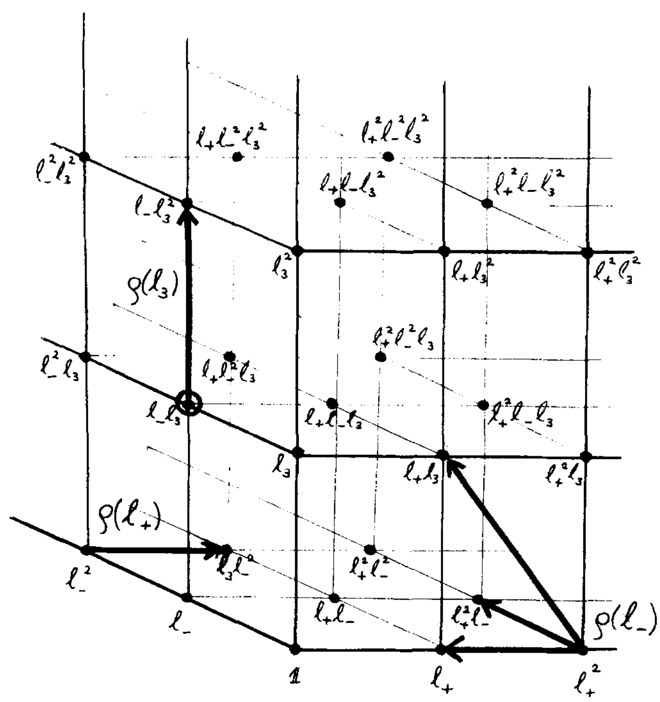


FIG. 1. The basis for Ω is plotted in the form of a crystal lattice. The action of the elements $l_3, l_\pm$ of A_1 upon the basis elements is indicated by (multiple) arrows. The circle at the bottom of the arrow which indicates the action of l_3 indicates that also a multiple of the state is obtained upon which l_3 acts. Note, however, that matrix elements may become zero for the action of l_3 and l_- upon certain basis elements.

$$[\rho(l_3),\rho(l_\pm)] = \pm\rho(l_\pm), \quad [\rho(l_+),\rho(l_-)] = \rho(l_3). \tag{2.2}$$

The representation given by Eq. (2.1) is a rather unusual one. It is indecomposable, and, moreover, the linear operator $\rho(l_3)$ which represents the element l_3 of the Cartan subalgebra of A_1 does not have any eigenvector.

The action of the operators $\rho(l_\pm)$, $\rho(l_3)$ onto the basis elements is shown in Fig. 1. It can be seen from Eq. (2.1) and Fig. 1 that the operators $\rho(l_\pm)$, $\rho(l_3)$ can only increase a power of l_- and a power of l_3, whereas a power of l_+ can be increased and decreased. Therefore, each of the subspaces ${}_MV$, V_R, ${}_MV_R$ with bases

$$\begin{aligned} {}_MV&: \{l_+^n l_-^{M+m} l_3^r,\ n,m,r=0,1,2,\cdots\},\\ V_R&: \{l_+^n l_-^m l_3^{R+r},\ n,m,r=0,1,2,\ldots\},\\ {}_MV_R&: \{l_+^n l_-^{M+m} l_3^{R+r},\ n,m,r=0,1,2,\cdots\},\end{aligned} \tag{2.3}$$

where R, M are positive integers, forms an *invariant* subspace of Ω with respect to the representation equation (2.1). These invariant subspaces carry subrepresentations which are induced by the representation equation (2.1). The representation equation (2.1) also defines representations on quotient spaces of Ω with respect to these invariant subspaces. The representations on the quotient spaces are obtained from Eq. (2.1) by formally setting

$$\begin{aligned} &l_+^n l_-^M l_3^r \to 0, && \text{for } \Omega/{}_MV,\\ &l_+^n l_-^m l_3^R \to 0, && \text{for } \Omega/V_R,\\ &l_+^n l_-^M l_3^r \to 0,\quad l_+^n l_-^m l_3^R \to 0, && \text{for } \Omega/{}_MV_R.\end{aligned}$$

It is clear that the space ${}_MV$ has invariant subspaces ${}_{M'}V$, $M'>M$, and ${}_{M'}V_R$, $M'\geqslant R$, R any positive integer. The space V_R has invariant subspaces $V_{R'}$, $R'>R$, and ${}_MV_{R'}$, $R'\geqslant R$, M any positive integer. The space ${}_MV_R$ has invariant subspaces ${}_{M'}V_{R'}$, $M'\geqslant M$, $R'\geqslant R$, $(M,R)\neq(M',R')$. Thus, we can consider representations induced by the representation on the quotient spaces

$$\begin{aligned} &{}_MV/{}_{M'}V,\ M'>M, \quad {}_MV/{}_{M'}V_R,\ M'\geqslant M,\\ &V_R/V_{R'},\ R'>R,\\ &V_R/{}_MV_{R'},\ R'\geqslant R, \quad {}_MV_R/{}_{M'}V_{R'},\ M'\geqslant M,\ R'\geqslant R.\end{aligned}$$

It is easy to see from Eq. (2.1) and Fig. 1 that the representation ρ of A_1 on Ω is a semidirect sum of the representations realized on ${}_MV/{}_{M+1}V$, $M=0,1,2,\cdots$ (${}_0V=\Omega$), and a semidirect sum of the representations realized on V_R/V_{R+1}, $R=0,1,2,\cdots$ ($V_0=\Omega$). In other words, if ρ_M denotes the representation realized on ${}_MV/{}_{M+1}V$, then the representation ρ, Eq. (2.1), can be represented in the basis (1.4) by matrices

$$\begin{pmatrix} \rho_0 & 0 & 0 & 0 & \cdots\\ * & \rho_1 & 0 & 0 & \cdots\\ 0 & * & \rho_2 & 0 & \cdots\\ 0 & 0 & * & \rho_3 & \cdots\\ \cdots & \cdots & \cdots & & \end{pmatrix},$$

where $*$ is a nonzero matrix.

The representations on quotient spaces will be considered now. First of all we consider the representations on $\Omega/{}_1V$ and Ω/V_1. The representations on ${}_MV/{}_{M+1}V$ and V_R/V_{R+1} can be considered in the same way.

3. REPRESENTATIONS ON $\Omega/{}_1V$

The basis of the quotient space $\Omega/{}_1V$ can be identified with the set

$$\{l_+^n l_3^r,\ n,r=0,1,2,\cdots\}. \tag{3.1}$$

This space can be formally defined by setting $l_+^n l_-^m l_3^r=0$, $m\geqslant 1$. Making use of this relation in Eq. (2.1), one obtains

$$\begin{aligned} \rho(l_3)l_+^n l_3^r &= l_+^n l_3^{r+1} + n l_+^n l_3^r,\\ \rho(l_+)l_+^n l_3^r &= l_+^{n+1} l_3^r,\\ \rho(l_-)l_+^n l_3^r &= -n l_+^{n-1} l_3^{r+1} - \tfrac{1}{2}(n-1)n l_+^{n-1} l_3^r,\\ \rho(C')l_+^n l_3^r &= l_+^n l_3^{r+2} - l_+^n l_3^{r+1}.\end{aligned} \tag{3.2}$$

This formulas are graphically illustrated in Fig. 2.

It can be easily checked that Eq. (3.2) forms a representation. This representation is indecomposable, has invariant subspaces W_R with bases

$$W_R: \{l_+^n l_3^{R+r},\ n,r=0,1,2,\cdots\},$$

with $R\geqslant 0$, integer, and $\rho(l_3)$ has *no* eigenvector. The subspaces W_R carry subrepresentations. Moreover, the representation equation (3.2) induces representations on the quotient spaces $(\Omega/{}_1V)/W_R \sim \Omega/{}_1V_R$. These are obtained from Eq. (3.2) by formally setting $l_+^n l_3^R=0$. For $R=1$ one obtains

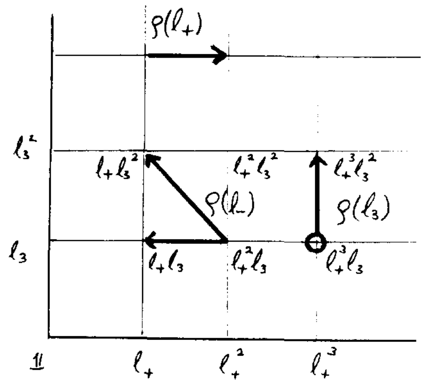


FIG. 2. Action of elements of A_1 upon the basis of the space $\Omega /_1 V$.

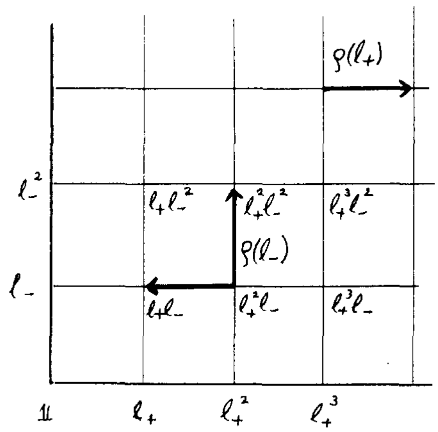


FIG. 3. Action of elements of A_1 upon the basis elements of the space Ω / V_1.

$$\rho(l_3)l_+^n = nl_+^n, \quad \rho(l_+)l_+^n = l_+^{n+1}, \qquad (3.3)$$
$$\rho(l_-)l_+^n = -\tfrac{1}{2}(n-1)nl_+^{n-1}, \quad \rho(C)l_+^n = 0.$$

Equation (3.3) forms an infinite-dimensional indecomposable representation with lowest weight $\Lambda = 0$. It has two extremal vectors, namely, $\mathbb{1}, \rho(l_-)\mathbb{1} = 0$, and $l_+, \rho(l_-)l_+ = 0$ (for the definition of extremal vectors, see Refs. 14, 7). It is a special case of a so-called elementary representation.[14,7] The representation equation (3.3) contains two irreducible representations of A_1: a one-dimensional representation and an infinite-dimensional representation with lowest weight 1.

The space W_R has invariant subspaces $W_{R'}, R' > R$. Equation (3.2) induces on $W_R/W_{R'}$ a representation. Let us consider representations on the quotient spaces W_R/W_{R+1}. The basis of this space can be identified with

$$\{l_+^n l_3^R, \ n = 0, 1, 2, \cdots\}. \qquad (3.4)$$

It follows from Eq. (3.2) that on W_R/W_{R+1}

$$\rho(l_3)l_+^n l_3^R = nl_+^n l_3^R,$$
$$\rho(l_+)l_+^n l_3^R = l_+^{n+1}l_3^R, \qquad (3.5)$$
$$\rho(l_-)l_+^n l_3^R = -\tfrac{1}{2}(n-1)nl_+^{n-1}l_3^R,$$
$$\rho(C)l_+^n l_3^R = 0.$$

Thus, the representation on W_R/W_{r+1} is algebraically equivalent to the representation equation (3.3). Therefore, the representation of A_1 on $\Omega /_1 V$ is a semidirect sum of an infinite number of representations, Eq. (3.3) (denoted by d_0). In the basis $\{l_+^n l_3^r, n, r = 0, 1, 2, \cdots\}$ the representation equation (3.2) can be represented by matrices

$$\begin{pmatrix} d_0 & 0 & 0 & 0 & \cdots \\ * & d_0 & 0 & 0 & \cdots \\ 0 & * & d_0 & 0 & \cdots \\ 0 & 0 & * & d_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where $*$ denotes nonzero matrix.

The representation induced by Eq. (3.2) on the quotient spaces $\Omega/_1 V_R$ has the property that the subset of elements $l_+^n l_3^{R-1}, R > 0, n = 0, 1, 2, \cdots$, are eigenvectors for $\rho(l_3)$ (see Fig. 2).

4. REPRESENTATIONS ON Ω/V_1

The quotient space Ω / V_1 can be defined by setting $l_+^n l_-^m l_3^r = 0, r \geqslant 1$. Applying this condition to Eq. (2.1) yields the following representation on Ω / V_1:

$$\Omega / V_1: \ \{\mathbb{1}, l_+^n l_-^m, \ n, m = 0, 1, 2, \cdots\},$$
$$\rho(l_3)l_+^n l_-^m = (n-m)l_+^n l_-^m,$$
$$\rho(l_+)l_+^n l_-^m = l_+^{n+1}l_-^m, \qquad (4.1)$$
$$\rho(l_-)l_+^n l_-^m = l_+^n l_-^{m+1} + n[m - \tfrac{1}{2}(n-1)]l_+^{n-1}l_-^m,$$
$$\rho(C)l_+^n l_-^m = 2l_+^{n+1}l_-^{m+1} + m(m+1)l_+^n l_-^m.$$

These formulas are graphically represented by Fig. 3.

The representation equation (4.1) is indecomposable, $\rho(l_3)$ is diagonal and its weight subspaces are infinite-dimensional. There are invariant subspaces ${}_R W, R > 0$, integer, with bases

$$\{l_+^n l_-^{R+r}, \ n, r = 0, 1, 2, \cdots\}.$$

The representation equation (4.1) is discussed in detail in Ref. 7 and corresponds to the nonmultiplicity free representation with $\Lambda = 0$.

5. REPRESENTATIONS ON Ω/I_-

The basis of Ω / I_- can be identified with

$$\{\mathbb{1}, l_+^n l_3^r, \ n, r = 0, 1, 2, \cdots\}. \qquad (5.1)$$

It follows from Eq. (2.1), making use of $l_- l_3^r = (l_3 + 1)^r l_-$ and the property $\rho(l_-)\mathbb{1} = \Lambda \mathbb{1}$, that

$$\rho(l_3)l_+^n l_3^r = l_+^n l_3^{r+1} + nl_+^n l_3^r,$$
$$\rho(l_+)l_+^n l_3^r = l_+^{n+1}l_3^r,$$
$$\rho(l_-)l_+^n l_3^r = \Lambda \sum_{k=0}^{r} \frac{r!}{(r-k)!k!} l_+^n l_3^k$$
$$- nl_+^{n-1}l_3^{r+1} - \tfrac{1}{2}(n-1)nl_+^{n-1}l_3^r, \qquad (5.2)$$
$$\rho(C)l_+^n l_3^r = 4\Lambda \sum_{k=0}^{r} \frac{r!}{(r-k)!k!} l_+^{n+1}l_3^k$$
$$+ l_+^n l_3^{r+2} - l_+^n l_3^{r+1}.$$

These relations form a representation of A_1, as can be verified by direct computation. This representation has no invariant

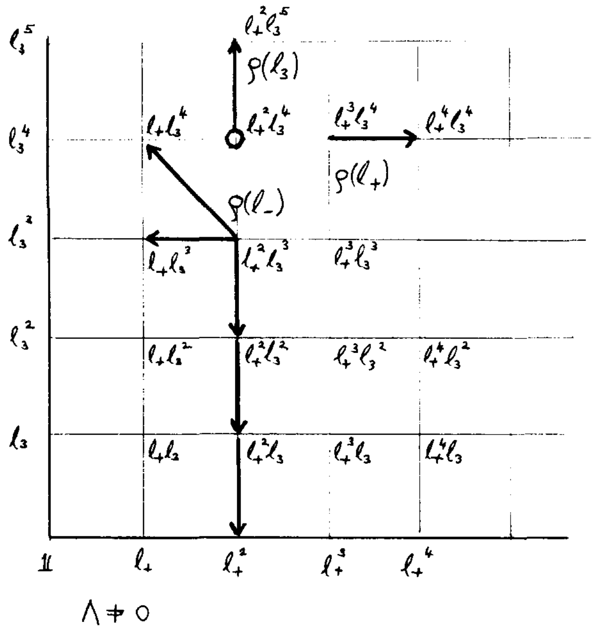


FIG. 4. Action of elements of A_1 upon the basis elements of the space Ω/I_-.

subspaces and $\rho(l_3)$ has no eigenvectors (Fig. 4). For $\Lambda = 0$ the representation Eq. (3.2) is obtained.

6. REPRESENTATIONS ON Ω/I_3

Let us consider the representations induced by Eq. (2.1) on the quotient space Ω/I_3, where I_3 is a left ideal generated by $(l_3 - \Lambda\mathbb{1})$ with fixed $\Lambda \in C$. The basis of Ω/I_3 can be identified with

$$\{\mathbb{1}, l_+^n l_-^m, \ n, m = 0, 1, 2, \cdots\}. \tag{6.1}$$

Using the property $\rho(l_3)\mathbb{1} = \Lambda\mathbb{1}$, one obtaines from Eq. (2.1),

$$\begin{aligned}
\rho(l_3) l_+^n l_-^m &= (\Lambda + n - m) l_+^n l_-^m,\\
\rho(l_+) l_+^n l_-^m &= l_+^{n+1} l_-^m,\\
\rho(l_-) l_+^n l_-^m &= l_+^n l_-^{m+1}\\
&\quad - n(\Lambda - m + \tfrac{1}{2}(n-1)) l_+^{n-1} l_-^m,\\
\rho(C) l_+^n l_-^m &= 2 l_+^{n+1} l_-^{m+1}\\
&\quad + (\Lambda - m)(\Lambda - m - 1) l_+^n l_-^m.
\end{aligned} \tag{6.2}$$

The representation equation (4.1) is obtained from Eq. (6.2) for the special value $\Lambda = 0$. The representation equation (6.2) is graphically shown in Fig. 9 of Ref. 7.

The representation given by Eq. (6.2) was investigated in some detail in Ref. 7 (with a difference in the definition of the bases for A_1 and Ω, as was noted at the beginning of Sec. 1). Here we investigate these representations further. In order to do so, we need to state some of the results of Sec. III of Ref. 7.

The space Ω/I_3 has invariant subspaces H_R with basis

$$\{l_+^n l_-^{R+r}, \ n, r = 0, 1, 2, \cdots\}. \tag{6.3}$$

On the quotient space H_R/H_{R+1} ($R = 0, 1, 2, \cdots; H_0 = \Omega$) the multiplicity-free representation ρ^R is realized and it holds

$$\begin{aligned}
\rho^R(l_3) l_+^n l_-^R &= (\Lambda + n - R) l_+^n l_-^R,\\
\rho^R(l_+) l_+^n l_-^R &= l_+^{n+1} l_-^R,\\
\rho^R(l_-) l_+^n l_-^R &= -n[\Lambda - R + \tfrac{1}{2}(n-1)] l_+^{n-1} l_-^R,\\
\rho^R(C) l_+^n l_-^R &= (\Lambda - R)(\Lambda - R - 1) l_+^n l_-^R.
\end{aligned} \tag{6.4}$$

This is a representation with lowest weight $\Lambda - R$. The representation ρ^R is reducible if $\Lambda - R = -\frac{1}{2}(n-1)$, $n = 1,2,3,\cdots$. In this case ρ^R has a second extremal vector, namely $l_+^{-2(\Lambda - R)+1} l_-^R$, and is indecomposable. This indecomposable representation has two irreducible representations: a finite-dimensional representation of A_1 with highest weight $\frac{1}{2}(n-1) \equiv R - \Lambda$ and an infinite-dimensional representation with lowest $\frac{1}{2}(n+1) \equiv R - \Lambda + 1$, which is realized in an invariant subspace. If $\Lambda - R \neq -\frac{1}{2}(n-1)$, $n = 1,2,3,\cdots$, then the representation ρ^R is irreducible.

The representation ρ, Eq. (6.2), on Ω/I_3 is decomposed into a semidirect sum of the representations ρ^R,

$$\rho = \begin{pmatrix}
\rho^0 & 0 & 0 & 0 & \cdots\\
* & \rho^1 & 0 & 0 & \cdots\\
0 & * & \rho^2 & 0 & \cdots\\
0 & 0 & * & \rho^3 & \cdots\\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix},$$

where each of the constituents ρ^R is realized on the quotient space H_R/H_{R+1}.

The representation ρ, Eq. (6.2), on Ω/I_3 cannot be decomposed into a direct sum of multiplicity-free representations of A_1. It can be decomposed into a direct sum under the condition that we extend Ω/I_3 by including infinite sums of basis elements $l_+^n l_-^m$, $n, m = 0,1,2,\cdots$. Two cases are distinguished:

Case (6a): $\Lambda \neq \frac{1}{2}m$, $m = 1,2,3,\cdots$

First we consider representations on the quotient space $(\Omega/I_3)/H_R$. These representations can be decomposed into a direct sum

$$\sum_{s=0}^{R-1} \oplus \ \bar{\rho}^{(s)} \tag{6.5}$$

of multiplicity-free representations $\bar{\rho}^{(s)}$. The basis for $\bar{\rho}^{(s)}$ is

$$\{l_+^n x^{(s)}, \ s = 0, 1, 2, \cdots\}$$

with

$$x^{(s)} = l_-^s + \sum_{k=1}^{R-1} \left\{ k! \prod_{t=1}^{k} [\Lambda - s - \tfrac{1}{2}(t+1)] \right\}^{-1} l_+^k l_-^{k+s}. \tag{6.6}$$

The representation $\bar{\rho}^{(s)}$ acts on the basis elements as

$$\begin{aligned}
\bar{\rho}^{(s)}(l_3) l_+^n x^{(s)} &= (\Lambda - s + n) l_+^n x^{(s)},\\
\bar{\rho}^{(s)}(l_+) l_+^n x^{(s)} &= l_+^{n+1} x^{(s)},\\
\bar{\rho}^{(s)}(l_-) l_+^n x^{(s)} &= -n(\Lambda - s + \tfrac{1}{2}(n-1)) l_+^{n-1} x^{(s)},\\
\bar{\rho}^{(s)}(C) l_+^n x^{(s)} &= (\Lambda - s)(\Lambda - s - 1) l_+^n x^{(s)},
\end{aligned} \tag{6.7}$$

i.e., the representation $\bar{\rho}^{(s)}$ is isomorphic to the representation ρ^R which acts on H_R/H_{R+1} (see above).

Now we extend the space Ω/I_3 to include infinite (formal) sums of basis elements and extend the action of the representation ρ by linearity. In this case the representation decomposes into a (infinite) direct sum of representations $\rho^{(s)}$, $s = 0, 1, 2, \cdots$, which are multiplicity-free. The basis for the representation $\rho^{(s)}$ is

$$\{l_+^n x^{(s)},\ n = 0, 1, 2, \cdots\}$$

with

$$x^{(s)} = l_-^s + \sum_{k=1}^{\infty} \left\{k!\prod_{t=1}^{k}[\Lambda - s - \tfrac{1}{2}(t+1)]\right\}^{-1} l_+^k l_-^{k+s}. \tag{6.8}$$

For $\rho^{(s)}$ we have

$$\begin{aligned}
&\rho^{(s)}(l_3) l_+^n x^{(s)} = (\Lambda - s + n) l_+^n x^{(s)},\\
&\rho^{(s)}(l_+) l_+^n x^{(s)} = l_+^{n+1} x^{(s)},\\
&\rho^{(s)}(l_-) l_+^n x^{(s)} = -n[\Lambda - s + \tfrac{1}{2}(n-1)] l_+^{n-1} x^{(s)},\\
&\rho^{(s)}(C) l_+^n x^{(s)} = (\Lambda - s)(\Lambda - s - 1) l_+^n x^{(s)}.
\end{aligned} \tag{6.9}$$

Therefore, the representations $\rho^{(s)}$ are given by the same formulas as the representations $\bar{\rho}^{(s)}$ and ρ^R, but they act in different spaces.

Case (6b): $\Lambda = \frac{1}{2}m$, $m = 1, 2, 3, \cdots$

If one attempts to decompose the representation ρ into a direct sum of multiplicity-free representations of A_1 [extending the space Ω/I_3 to include infinite sums as was done for the case (6a)], it turns out to be impossible.[7] Instead, the representation ρ (after an extension of Ω/I_3 to include infinite sums) can be decomposed into a direct sum which includes multiplicity-free representations as well as a finite number of finite-multiplicity representations. In the following we give the decomposition in explicit form. Moreover, we construct explicitly the indecomposable finite-multiplicity representations. These representations have found physical applications.[8]

As was mentioned above, the representation ρ on Ω/I_3 can be decomposed into a semidirect sum of representations ρ^s with lowest weights $\Lambda, \Lambda - 1, \Lambda - 2, \cdots$. We divide these weights into two subsets:

Subset 1:

$$\begin{aligned}
&\Lambda = \tfrac{1}{2}m,\\
&\Lambda - 1 = \tfrac{1}{2}m - 1,\\
&\Lambda - 2 = \tfrac{1}{2}m - 2,\\
&\quad\vdots\\
&\Lambda - (2\Lambda - 2) = -\Lambda + 2 = -\tfrac{1}{2}m + 2,\\
&\Lambda - (2\Lambda - 1) = -\Lambda + 1 = -\tfrac{1}{2}m + 1.
\end{aligned}$$

Subset 2:

$$\begin{aligned}
&\Lambda - 2\Lambda = -\Lambda = -\tfrac{1}{2}m,\\
&\Lambda - (2\Lambda + 1) = -\Lambda - 1 = -\tfrac{1}{2}m - 1,\\
&\Lambda - (2\Lambda + 2) = -\Lambda - 2 = -\tfrac{1}{2}m - 2,\\
&\quad\vdots\ .
\end{aligned}$$

Now we consider the vectors (6.8). In this case (i.e., when $\Lambda = \frac{1}{2}m$, $m = 1,2,3,\cdots$) some of the $x^{(s)}$ may not exist. Indeed, in the sum

$$\begin{aligned}
&x^{(s)} = \sum_{k=0}^{\infty} c_k^{(s)} l_+^k l_-^{k+s},\\
&c_k^{(s)} = \left\{k!\prod_{t=1}^{k}[\Lambda - s - \tfrac{1}{2}(t+1)]\right\}^{-1}, \quad k = 1, 2, 3, \cdots,\\
&c_0^{(s)} = 1,
\end{aligned} \tag{6.10}$$

some of the coefficients $c_k^{(s)}$ may become singular. In fact, for given $\Lambda = \frac{1}{2}m$, $m = 1, 2, 3, \cdots$, and given $s = 0, 1, 2, \cdots$, such that

$$2(\Lambda - s) - 1 = k \geqslant 1, \quad \text{integer}, \tag{6.11}$$

the coefficients

$$c_k^{(s)},\ c_{k+1}^{(s)},\ c_{k+2}^{(s)},\ \cdots$$

become singular. Thus, this $x^{(s)}$ does not exist.

It follows from Eq. (6.11) that the lowest vector $x^{(s)}$ becomes singular if

$$\Lambda - s \geqslant 1. \tag{6.12}$$

It is known that the lowest vector $x^{(s)}$ belongs to the weight $\Lambda - s$. Now we consider the lowest weights of Subsets 1 and 2. It follows from the singularity condition $\Lambda - s \geqslant 1$ that all vectors $x^{(s)}$, $s = 2\Lambda, 2\Lambda + 1, 2\Lambda + 2, \cdots$, which belong to weights of Subset 2, are not singular. The vectors $x^{(s)}$, $s = \Lambda - \frac{1}{2}$ or $\Lambda, \ldots, 2\Lambda - 1$, which belong to the weights $\frac{1}{2}$ or $0, \ldots, -\Lambda + 1$, of Subset 1 are not singular too. The vectors $x^{(s)}$, $s = 0,1,2,\ldots,\Lambda - 1$ or $\Lambda - \frac{3}{2}$, which belong to the weights $\Lambda, \Lambda - 1, \ldots, 1$ or $\frac{3}{2}$ of Subset 1 are singular.

It follows that for the case $\Lambda = \frac{1}{2}$ there is no singular vector. For this case the vectors $x^{(s)}$, $s = 0, 1, 2, \cdots$, can be constructed in exactly the same manner as was done for Case (6a) by making use of Eq. (6.8). Then these vectors are used to construct the bases $\{l_+^n x^{(s)}, n = 0, 1, 2, \cdots\}$, $s = 0, 1, 2, \cdots$. The representations $\rho^{(s)}$, which are given by Eq. (6.9), are realized on these bases. As in Case (6a) the representation ρ, which is realized on the extension of the space Ω/I_3, $\Lambda = \frac{1}{2}$, can be decomposed into a direct sum of representations $\rho^{(s)}$, $s = 0, 1, 2, \cdots$.

Now let $\Lambda = \frac{1}{2}m$, $m = 2, 3, 4, \cdots$. For every s, such that $s \geqslant \Lambda$ we construct the vector $x^{(s)}$, by making use of Eq. (6.8). Then we construct the basis

$$\{l_+^n x^{(s)},\ n = 0, 1, 2, \cdots\}.$$

The representation $\rho^{(s)}$, given by Eq. (6.9), is realized on this basis. We obtain a direct sum

$$\sum_{\substack{s=\Lambda\\ \text{or } s=\Lambda-\frac{1}{2}}}^{\infty} \oplus\, \rho^{(s)}.$$

This representation is a *sub*representation of the representation ρ defined on Ω/I_3. Now we will construct the remainder of the representation ρ.

We return to the singular vectors (6.10). Let us consider the finite part of the sum (6.10),

$$\sum_{k=0}^{2(\Lambda-s-1)} c_k^{(s)} l_+^k\, l_-^{k+s}. \tag{6.13}$$

This sum does not contain singular terms. The following relation holds:

$$\rho(l_-)\sum_{k=0}^{2(\Lambda-s-1)} c_k^{(s)} l_+^k\, l_-^{k+s} = c^{(s)}_{2(\Lambda-s-1)}\, l_+^{2(\Lambda-s-1)} l_-^{2(\Lambda-s-1)+s+1}. \tag{6.14}$$

The element

$$l_+^{2(\Lambda-s-1)} l_-^{2(\Lambda-s-1)+s+1}$$

is the first term ($k=0$) of the basis element

$$l_+^{2(\Lambda-s-1)} x^{(s')} = l_+^{2(\Lambda-s-1)}\left(\sum_{k=0}^{\infty} c_k^{(s')} l_+^k\, l_-^{k+s'}\right)$$

of the (indecomposable) representation $\rho^{(s')}$, where

$$s' = 2(\Lambda - s - 1) + s + 1 = 2\Lambda - s - 1.$$

As we saw, the vector $x^{(s')}$ is not singular. The representation $\rho^{(s')}$ has a lowest weight

$$\Lambda - s' = -\Lambda + s + 1.$$

The representations $\rho^{(s)}$ and $\rho^{(s')}$ have the same eigenvalue of the Casimir operator C. Therefore, these two representations are related to each other through an indecomposable representation.

Instead of the nonexisting element $x^{(s)}$ with nonsingular (regular) part (6.13), we consider the vector

$$y^{(s)} = \alpha \sum_{k=0}^{2(\Lambda-s-1)} c_k^{(s)} l_+^k\, l_-^{k+s} + \sum_{k=0}^{\infty} d_k\, l_+^{s'-s+k} l_-^{s'+k} \tag{6.15}$$

such that

$$\rho(l_-)y^{(s)} = l_+^{2(\Lambda-s-1)} x^{(s')}. \tag{6.16}$$

The constants α and d_k are defined by

$$\alpha = 2^{-2(\Lambda-s-1)}\{[2(\Lambda-s-1)]!\}^2 = (c^s_{2(\Lambda-s-1)})^{-1}, \tag{6.17}$$

$$d_{k+1} = \frac{c^{(s')}_{k+1} - d_k}{(k+1)(\Lambda - s + \frac{1}{2}k)}, \quad k = 0, 1, 2, \cdots,$$

where

$$c^{(s')}_{k+1} = \frac{(-2)^{k+1}[2(\Lambda-s)-1]!}{(k+1)![2(\Lambda-s)+k]!}, \quad k = 0, 1, 2, \cdots, \tag{6.18}$$

$$c_0^{(s')} = 1.$$

The coefficient d_0 is arbitrary and can be chosen to be zero.

By means of the vectors $x^{(s')}$ and $y^{(s)}$ we construct the basis

$$\{l_+^n\, y^{(s)},\ n = 0,1,2,\cdots;$$
$$l_+^m\, x^{(s')},\ m = 0,1,2,\cdots;\ s' = 2\Lambda - s - 1\}.$$

It can be directly shown that Eq. (6.2) induces on this basis an indecomposable representation (we denote it by $\rho^{(s,s')}$) and that

$$\rho(l_3)l_+^n\, y^{(s)} = (\Lambda - s + n)l_+^n\, y^{(s)},$$
$$\rho(l_3)l_+^m\, x^{(s')} = (-\Lambda + m + s + 1)l_+^m\, x^{(s')},$$
$$\rho(l_+)l_+^n\, y^{(s)} = l_+^{n+1} y^{(s)},$$
$$\rho(l_+)l_+^m\, x^{(s')} = l_+^{m+1} x^{(s')},$$
$$\rho(l_-)l_+^n\, y^{(s)} = l_+^{2(\Lambda-s-1)+n} x^{(s')} - n[\Lambda - s + \tfrac{1}{2}(n-1)]l_+^{n-1} y^{(s)},$$
$$\rho(l_-)l_+^m\, x^{(s')} = m[\Lambda - s - \tfrac{1}{2}(m+1)]l_+^{m-1} x^{(s')},$$
$$\rho(C)l_+^n\, y^{(s)} = 2l_+^{s'-s+n} x^{(s')} + (\Lambda - s)(\Lambda - s - 1)l_+^n\, y^{(s)},$$
$$\rho(C)l_+^m\, x^{(s')} = (\Lambda - s)(\Lambda - s - 1)l_+^m\, x^{(s')}. \tag{6.19}$$

To simplify notation, we have dropped the indices s, s' in Eq. (6.19).

The subspace spanned by

$$\{l_+^m\, x^{(s')},\ m = 0, 1, 2, \cdots\}$$

remains invariant, and thus the representation $\rho^{(s,s')}$, Eq. (6.19), yields under restriction to this subspace an indecomposable subrepresentation which is multiplicity-free. This representation has two extremal vectors which are $x^{(s')}$ and $l_+^{2(\Lambda-s)-1} x^{(s')}$ with weights $-\Lambda + s + 1$ and $\Lambda - s$, respectively. Thus, the representation $\rho^{(s,s')}$ "contains" a finite-dimensional irreducible representation of A_1 with highest weight $\Lambda - s - 1$ and two infinite-dimensional irreducible representations with lowest weight $\Lambda - s$. None of these representations can be separated into a direct summand.

Let us remember that the representation $\rho^{(s,s')}$ is characterized by an integer s such that $s \leqslant \Lambda - 1$, $\Lambda = \frac{1}{2}m$, $m = 2, 3, 4, \cdots$, and an integer s' such that $s' = 2\Lambda - s - 1$. Thus, the integers s and s' are not independent. Figure 5 gives a graphi-

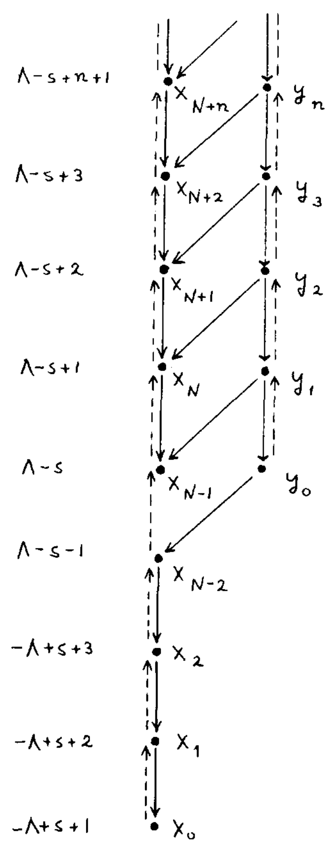


FIG. 5. Finite multiplicity indecomposable representation of A_1 for $\Lambda - s \equiv \frac{1}{2}N = \frac{5}{2}$.

cal description of the representation $\rho^{(s,s')}$. The vectors $l^n_+ y^{(s)} \equiv y_n$, $l^m_+ x^{(s')} \equiv x_m$ are shown by dots. The action of the operator $\rho(l_+)$ is shown by dashed arrows, and the action of $\rho(l_-)$ by solid arrows.

The case of the simplest representation of type $\rho^{(s,s')}$ is graphically described in Ref. 7, Fig. 11.

The situation for Case (6b) with $\Lambda = \frac{1}{2}m$, $m = 2, 3, 4, \cdots$, is thus the following. The representation equation (6.2) can be decomposed into a direct sum of multiplicity-free representations $\rho^{(s)}$ and non-multiplicity-free representations $\rho^{(s,s')}$. The representations $\rho^{(s)}$ correspond to the lowest weights $\Lambda - s$ of Subset 2 (see above). The representations $\rho^{(s,s')}$ are semidirect sums of multiplicity-free representations $\rho^{(s)}$ and $\rho^{(s')}$ with lowest weights $\Lambda - s$ and $\Lambda - s' = -\Lambda + s + 1$, which belong to Subset 1. To the particular case of $\Lambda = \frac{1}{2}$ of Subset 1 corresponds in the decomposition the multiplicity-free representation $\rho^{(s'')}$, $s'' = \Lambda - \frac{1}{2}$.

7. REPRESENTATION ON $\Omega_+ \cup \Omega_-$

Another type of representations of A_1 is obtained as follows. Using the property

$$l_+ l_- = \tfrac{1}{2}[C - l_3(l_3 - 1)], \tag{7.1}$$

where C denotes the Casimir invariant of A_1, it is easy to verify that the basis for Ω, Eq. (1.4), can be reexpressed as

$$\{ l^n_+ f(C,l_3),\ l_-^m g(C,l_3),\ n,m = 0,1,2,\cdots \}, \tag{7.2}$$

where f and g are polynomials. The unit element is given by the trivial polynomial $f = 1$ (or $g = 1$) and $n,m = 0$. With respect to these elements the following relations are obtained from Eq. (2.1):

$$\rho(l_+) l_-^n = n[\Lambda - \tfrac{1}{2}(n-1)] l_-^{n-1}, \quad n \geqslant 1,$$
$$\rho(l_-) l^n_+ = -(n-1)(\Lambda + \tfrac{1}{2}n) l^{n-1}_+, \quad n \geqslant 1.$$

Case (7a)

Making use of Eq. (2.1), we obtain

$$\begin{aligned} \rho(l_3) l^n_\pm &= l^n_\pm (l_3 \pm n), \\ \rho(l_+) l^n_+ &= l^{n+1}_+, \quad \rho(l_-) l_-^n = l_-^{n+1}, \\ \rho(l_-) l^n_+ &= l^{n-1}_+ \{ \tfrac{1}{2}[C - l_3(l_3 - 1)] \\ &\qquad - n[l_3 + \tfrac{1}{2}(n-1)] \}, \quad n \geqslant 1, \\ \rho(l_+) l^n_- &= l_-^{n-1} \{ \tfrac{1}{2}[C - l_3(l_3 + 1)] \\ &\qquad + n[l_3 - \tfrac{1}{2}(n-1)] \}, \quad n \geqslant 1. \end{aligned} \tag{7.3}$$

If we set

$$\rho(C)\mathbb{1} = \lambda \mathbb{1}, \quad \rho(l_3)\mathbb{1} = \Lambda \mathbb{1}, \quad \lambda, \Lambda \in \mathbb{C}, \tag{7.4}$$

then Eq. (7.2) goes over into the basis

$$\{ \mathbb{1}, l^n_+, l_-^m,\ n,m = 1,2,3,\cdots \}. \tag{7.5}$$

This is a basis for the quotient space Ω / E, where E is a left ideal of Ω which is generated by the elements $C - \lambda\mathbb{1}$ and $l_3 - \lambda\mathbb{1}$. The relations (7.3) then induce on Ω / E a representation. For this representation the following holds:

$$\begin{aligned} \rho(l_3) l^n_\pm &= (\Lambda \pm n) l^n_\pm, \\ \rho(l_\pm) l^n_\pm &= l^{n+1}_\pm, \quad n \geqslant 0, \\ \rho(l_\pm) l^n_\mp &= \tfrac{1}{2}[\lambda - (\Lambda \mp n)(\Lambda \mp n \pm 1)] l^{n-1}_\mp, \quad n \geqslant 1, \\ \rho(C) l^n_\pm &= \lambda l^n_\pm. \end{aligned} \tag{7.6}$$

These representations are determined by two numbers λ, Λ. It appears that these representations have not been discussed in the literature. However, for the particular case that $\lambda = \Lambda(\Lambda + 1)$, this type of representation was discussed in Ref. 7.

For the case that $\lambda = \Lambda(\Lambda + 1)$ the following holds:

$$\begin{aligned} \rho(l_+) l_-^n &= n[\Lambda - \tfrac{1}{2}(n-1)] l_-^{n-1}, \quad n \geqslant 1, \\ \rho(l_-) l^n_+ &= -(n-1)(\Lambda + \tfrac{1}{2}n) l^{n-1}_+, \quad n \geqslant 1. \end{aligned} \tag{7.7}$$

Thus, for $\Lambda = \frac{1}{2}m$, $m = 0,1,2,\cdots$, the representation has two infinite-dimensional invariant subspaces which are generated by the extremal vectors l_+, $\rho(l_-) l_+ = 0$, and l_-^{m+1}, $\rho(l_+) l_-^{m+1} = 0$. In addition, there is a finite-dimensional noninvariant subspace with basis

$$\{ \mathbb{1}, l_-, l_-^2, \ldots, l_-^m \},$$

on which a finite-dimensional representation of A_1 is realized. A representation of this type is shown in Fig. 1 of Ref. 7.

If $\Lambda = -\frac{1}{2}m$, $m = 1,2,3,\cdots$, there are two infinite-dimensional invariant subspaces, with one nested in the other. The extremal vectors are l_+, $\rho(l_-) l_+ = 0$, and l_+^m, $\rho(l_-) l_+^m = 0$. Obviously, the invariant subspace generated by l_+^m is contained in the invariant subspace generated by l_+. It is a representation of the type shown in Fig. 7 of Ref. 7.

Case (7b)

Let

$$\{n\}_\pm \equiv \{ \tfrac{1}{2}[C - l_3(l_3 \mp 1)] - n[l_3 \pm \tfrac{1}{2}(n-1)] \}. \tag{7.8}$$

We set again

$$\rho(C)\mathbb{1} = \lambda\mathbb{1}, \quad \rho(l_3)\mathbb{1} = \Lambda\mathbb{1}, \quad \lambda, \Lambda \in \mathbb{C}.$$

Then, defining new basis elements

$$l'^n_\pm = \Big(\prod_{k=1}^{n} \{k\}_\pm \Big)^{-1} l^n_\pm, \tag{7.9}$$

one obtains from Eq. (7.3) the relations

$$\begin{aligned} \rho(l_3) l'^n_\pm &= l'^n_\pm (l_3 \pm n), \\ \rho(l_+) l'^n_+ &= l'^{n+1}_+ \{ \tfrac{1}{2}[C - l_3(l_3 - 1)] - (n+1)(l_3 + \tfrac{1}{2}n) \}, \\ \rho(l_+) l'^n_- &= l'^{n-1}_-, \quad n \geqslant 1, \\ \rho(l_-) l'^n_+ &= l'^{n-1}_+, \quad n \geqslant 1, \\ \rho(l_-) l'^n_- &= l'^{n+1}_- \{ \tfrac{1}{2}[C - l_3(l_3 + 1) + (n+1)(l_3 - \tfrac{1}{2}n)] \}. \end{aligned} \tag{7.10}$$

If in Eqs. (7.9) and (7.10) the operators C, l_3 are replaced by the eigenvalues λ, Λ, then Eq. (7.10) yields a representation.

If in addition

$$\lambda = \Lambda(\Lambda + 1), \tag{7.11}$$

one obtains from Eq. (7.10) the representation

$$\begin{aligned} \rho(l_3) l'^n_\pm &= (\Lambda \pm n) l'^n_\pm, \\ \rho(l_\pm) l'^n_\mp &= l'^{n-1}_\mp, \quad n \geqslant 1, \\ \rho(l_+) l'^n_+ &= -n[\Lambda + \tfrac{1}{2}(n+1)] l'^{n+1}_+, \\ \rho(l_-) l'^n_- &= (n+1)(\Lambda - \tfrac{1}{2}n) l'^{n+1}_-, \\ \rho(C) l'^n_\pm &= \Lambda(\Lambda + 1) l'^n_\pm. \end{aligned} \tag{7.12}$$

This representation has, for $\Lambda = \frac{1}{2}m$, $m = 0,1,2,\cdots$, two infinite-dimensional invariant subspaces, defined by the extremal vectors $\mathbb{1}'$, $\rho(l_+)\mathbb{1}' = 0$, and l'^m_-, $\rho(l_-)l'^m_- = 0$. The two invariant subspaces overlap, resulting in a third, finite-dimensional invariant subspace with basis

$$\{\mathbb{1}', l'^k_-, k = 1,2,3,\ldots,m\}. \tag{7.13}$$

This is a representation of a type shown in Fig. 2 of Ref. 7. General properties of these representations have been discussed by Gel'fand, Graev, and Vilenkin.[3]

For $\Lambda = -\frac{1}{2}m$, $m = 2,3,4,\cdots$, again two infinite-dimensional invariant subspaces are obtained, nested in each other. The first is definied by the extremal vector $\mathbb{1}'$, $\rho(l_+)\mathbb{1}' = 0$, and the second by the extremal vector $l'^{-2\Lambda+1}_+$, $\rho(l_+)l'^{-2\Lambda+1}_+ = 0$. Obviously the invariant subspace generated by the second extremal vector is an invariant subspace of the invariant subspace generated by the first extremal vector. It is a representation of a type shown in Fig. 5 of Ref. 7.

For

$$\Lambda \neq -\tfrac{1}{2}m,\ m = 2,3,4,\cdots, \text{ and } \Lambda \neq \tfrac{1}{2}m,\ m = 0,1,2,\cdots,$$

the representation contains one invariant subspace which is generated by the extremal vector $\mathbb{1}'$.

8. FINITE-DIMENSIONAL IRREDUCIBLE REPRESENTATIONS

The representation equation (7.12) has, for $\Lambda = \frac{1}{2}t$, $t = 0, 1, 2, \ldots$, three invariant subspaces. One of these is finite-dimensional with basis

$$\{\mathbb{1}', l'^n_-,\ n = 1,2,\ldots,2\Lambda\} \tag{8.1}$$

and carries a $(2\Lambda + 1)$-dimensional irreducible representation. These are the familiar angular momentum representations. An additional redefinition of basis elements,

$$n = \Lambda - m, \quad m = \Lambda, \Lambda - 1, \ldots, -\Lambda,$$

$$|\Lambda,\Lambda\rangle = \mathbb{1}',$$

$$|\Lambda,m-1\rangle = \prod_{k=\Lambda}^{m} [\tfrac{1}{2}(\Lambda - k + 1)(\Lambda + k)]^{1/2} l'^{\Lambda-(m-1)}_-,$$

$$m = \Lambda, \Lambda-1, \Lambda-2, \ldots, -\Lambda+1, \tag{8.2}$$

leads, on the invariant subspace equation (8.1), to the standard form for these representations in angular momentum theory

$$\rho(l_3)|\Lambda,m\rangle = m|\Lambda,m\rangle,$$

$$\rho(l_\pm)|\Lambda,m\rangle = \frac{1}{\sqrt{2}}\{\Lambda(\Lambda+1) - m(m\pm 1)\}^{1/2}|\Lambda,m\pm 1\rangle, \tag{8.3}$$

$$\rho(C)|\Lambda,m\rangle = \Lambda(\Lambda+1)|\Lambda,m\rangle.$$

9. ELEMENTARY REPRESENTATION ON Ω_+,Ω_-

The elementary representations[7,14] of A_1 are defined on the space Ω_- or Ω_+. The elementary representations are infinite-dimensional, but bounded on one side. Moreover, the basis for the representation space can be generated from the extremal vector which belongs to the highest weight (lowest weight) Λ, namely $\mathbb{1}$. That is,

$$\{\mathbb{1}, \rho(l_-)^n\mathbb{1},\quad n = 1,2,3,\cdots\} \tag{9.1}$$

or

$$\{\mathbb{1}, \rho(l_+)^m\mathbb{1},\quad m = 1,2,3,\cdots\} \tag{9.2}$$

forms a basis for a representation. If this representation is irreducible, then $\mathbb{1}$ is the only extremal vector. If not, then there is another extremal vector which corresponds to a lower weight $\Lambda - n$ (higher weight $\Lambda + n$). These representations can be obtained in various manners. For example, the representations generated by the extremal vectors $X^{(s)}$ of Case (6a) are elementary representations (with lowest weight). Another example for obtaining elementary representations is provided by Case (7a), for $\lambda = \Lambda(\Lambda + 1)$. The representation equation (7.6) then has an invariant subspace with basis

$$V\colon \{l_+, l_+^2, l_+^3, \cdots\}. \tag{9.3}$$

On the quotient space $\{\Omega_+ \oplus \Omega_-\}/V \sim \Omega_-$ with basis

$$\{\mathbb{1}, l_-, l_-^2, l_-^3, \cdots\}, \tag{9.4}$$

the elementary representations are obtained from Eq. (7.6) as

$$\begin{aligned}\rho(l_3)l_-^n &= (\Lambda - n)l_-^n,\\ \rho(l_+)l_-^n &= n[\Lambda - \tfrac{1}{2}(n-1)]l_-^{n-1}, \quad n \geqslant 0,\\ \rho(l_-)l_-^n &= l_-^{n+1}.\end{aligned} \tag{9.5}$$

If $\Lambda = \frac{1}{2}t$, $t = 0,1,2,\cdots$, then the representation is indecomposable with a second extremal vector l^{t+1}_-, $\rho(l_+)l^{t+1}_- = 0$. If $\Lambda \neq \frac{1}{2}t$, $t = 0,1,2,\cdots$, then the representation is irreducible.

Considering the case of an irreducible elementary representation ($\Lambda \neq \frac{1}{2}t$) and introducing a new index m by

$$n = \Lambda - m,$$

the following holds:

$$n[\Lambda - \tfrac{1}{2}(n-1)] = \tfrac{1}{2}[\Lambda(\Lambda+1) - m(m+1)].$$

Defining new basis elements

$$|\Lambda,\Lambda\rangle = \mathbb{1}, \quad \text{for } m = \Lambda, \tag{9.6}$$

$$|\Lambda,m\rangle = \left\{\prod_{k=m}^{\Lambda-1} \tfrac{1}{2}[\Lambda(\Lambda+1) - k(k+1)]\right\}^{-1/2} l_-^{\Lambda-m},$$

for $m = \Lambda - 1, \Lambda - 2, \ldots$, the matrix elements of an irreducible elementary representation, Eq. (9.5), then take on the form

$$\rho(l_3)|\Lambda,m\rangle = m|\Lambda,m\rangle,$$

$$\rho(l_-)|\Lambda,m\rangle = \{\tfrac{1}{2}[\Lambda(\Lambda+1) - m(m-1)]\}^{1/2}|\Lambda,m-1\rangle, \tag{9.7}$$

$$\rho(l_+)|\Lambda,m\rangle = \{\tfrac{1}{2}[\Lambda(\Lambda+1) - m(m+1)]\}^{1/2}|\Lambda,m+1\rangle,$$

$$m = \Lambda, \Lambda-1, \Lambda-2, \cdots.$$

If the elementary representation is reducible, i.e., if $\Lambda = \frac{1}{2}t$, $t = 0,1,2,\cdots$, then the basis equation (9.6) is not suitable since the basis element $|\Lambda, -\Lambda - 1\rangle$ is singular. If we are, however, interested only in the elementary subrepresentation defined on the invariant subspace generated by the extremal vector $l^{2\Lambda+1}_-$, then we can define the basis elements

$$|\Lambda, -\Lambda - 1\rangle = l_-^{2\Lambda+1}, \quad m = -\Lambda - 1, \tag{9.8}$$

$$|\Lambda, m\rangle = \left\{ \prod_{k=m}^{-\Lambda-2} \tfrac{1}{2}[\Lambda(\Lambda+1) - k(k+1)] \right\}^{-1/2} l_-^{\Lambda-m},$$

$$m = -\Lambda - 2, -\Lambda - 3, \cdots,$$

and it follows that Eqs. (9.7) again hold true. It is clear that this must be the case since, under $\Lambda \leftrightarrow -\Lambda - 1$, the expression $\Lambda(\Lambda + 1)$ remains unchanged. Thus Eq. (9.7) now describes the irreducible elementary representation with highest weight $-\Lambda - 1$.

For $\Lambda = \frac{1}{2}t$, $t = 0,1,2,\cdots$, Eq. (9.5) induces, on the quotient space with respect to the infinite-dimensional invariant subspace generated by the extremal vector $l_-^{2\Lambda+1}$, a finite-dimensional irreducible representation of dimension $2\Lambda + 1$. It follows that Eq. (9.7) then gives the matrix elements of this finite-dimensional representation for $m = \Lambda, \Lambda - 1, \ldots, -\Lambda$. This, of course, is a result which was obtained previously in a different manner [Eq. (8.3)].

10. INDECOMPOSABLE AND IRREDUCIBLE INFINITE-DIMENSIONAL REPRESENTATIONS

A careful investigation of the representations obtained yields a wealth of derived representations, defined on invariant subspaces, as well as on quotient spaces. Some examples are given below.

Case (10a)

These are infinite-dimensional representations which are bounded on one side and which have a *finite*-dimensional invariant subspace (note that the elementary representations may have an invariant subspace, but if they do have one this subspace is infinite-dimensional). Representations of this type can be obtained, from the representation given by Eq. (7.12), by considering the representations which are induced by this representation on the quotient space $\Omega_+ \cup \Omega_- / \Omega_- \sim V$. One obtains (deleting primes)

$$\begin{aligned}
\rho(l_3)l_+^n &= (\Lambda + n)l_+^n, \quad n \geqslant 1,\\
\rho(l_+)l_+^n &= -n[\Lambda + \tfrac{1}{2}(n+1)]l_+^{n+1}, \quad n \geqslant 1,\\
\rho(l_-)l_+^n &= l_+^{n-1}, \quad n \geqslant 2,\\
\rho(l_-)l_+ &= 0,\\
\rho(C)l_+^n &= \Lambda(\Lambda+1)l_+^n, \quad n \geqslant 1.
\end{aligned} \tag{10.1}$$

Then, if $\Lambda = -\frac{1}{2}m$, $m = 2,3,\cdots$, the representation equation (10.1) contains an invariant subspace of dimension $(m - 1)$, with basis

$$\{l_+, l_+^2, \ldots, l_+^{m-1}\}. \tag{10.2}$$

Considering the representation equation (7.12) again, it is easily observed that for $\Lambda = \frac{1}{2}m$, $m = 0,1,2,\cdots$, this representation has an invariant subspace with basis

$$\{l_-^m, l_-^{m-1}, \ldots, l_-, 1, l_+^n, \quad n = 1,2,3,\cdots\}. \tag{10.3}$$

Restriction of the representation to this invariant subspace yields again an infinite-dimensional representation with a finite-dimensional invariant subspace,

$$\begin{aligned}
\rho(l_3)l_+^n &= (\Lambda + n)l_+^n, \quad n = 0,1,2,\cdots,\\
\rho(l_3)l_-^r &= (\Lambda - r)l_-^r, \quad r = 1,2,\ldots,m,\\
\rho(l_+)l_-^r &= l_-^{r-1},\\
\rho(l_+)l_+^n &= -n[\Lambda + \tfrac{1}{2}(n+1)]l_+^{n+1},\\
\rho(l_-)l_-^r &= (r+1)(\Lambda - \tfrac{1}{2}r)l_-^{r+1},\\
\rho(l_-)l_+^n &= l_+^{n-1},\\
\rho(C)l_+^n &= \Lambda(\Lambda+1)l_+^n,\\
\rho(C)l_-^r &= \Lambda(\Lambda+1)l_-^r.
\end{aligned} \tag{10.4}$$

Case (10b)

These are infinite-dimensional, irreducible representations unbounded on both sides. From Eq. (7.3) it follows that if

$$\Lambda + \tfrac{1}{2} \pm \sqrt{\lambda + \tfrac{1}{4}} \neq n, \quad n = 0,1,2,\cdots,$$

and

$$-(\Lambda - \tfrac{1}{2}) \pm \sqrt{\lambda + \tfrac{1}{4}} \neq m, \quad m = 0,1,2,\cdots,$$

where $\rho(l_3)\mathbb{1} = \Lambda\,\mathbb{1}$ and $\rho(C)\mathbb{1} = \lambda\mathbb{1}$, then the representation equation (7.6) is infinite-dimensional, irreducible, and unbounded on both sides.

[1]E. P. Wigner, *Group Theory* (Academic, New York, 1959).
[2]V. Bargmann, Ann. Math. **48**, 568 (1947).
[3]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[4]W. J. Holman III and L. C. Biedenharn, Ann. Phys. **39**, 1 (1966).
[5]E. Chaćon, D. Levi, and M. Moshinsky, J. Math. Phys. **17**, 1919 (1976).
[6]L. Armstrong, J. Math. Phys. **12**, 953 (1971); Phys. Rev. A **3**, 1546 (1971).
[7]B. Gruber and A. U. Klimyk, J. Math. Phys. **19**, 2009 (1978).
[8]M. Flato and C. Fronsdal, J. Math. Phys. **22**, 1100 (1982) and Phys. Lett B **97**, 236 (1980).
[9]A. D. Barut and R. Rączka, *Theory of Group Representations and Applications* (PWN—Polish Scientific Publishers, Warsaw, 1977).
[10]J. Dixmier, *Algebres enveloppantes* (Gauthiers-Villars, Paris, 1974).
[11]N. Jacobson, *Lie Algebras* (Wiley-Interscience, New York, 1962).
[12]J. M. Maillard and D. Sternheimer, C. R. Acad. Sci. Paris Ser. A **280**, 73 (1975).
[13]D. Arnal and G. Pinczon, Bull. Soc. Math. France **101**, 381 (1973).
[14]B. Gruber and A. U. Klimyk, J. Math. Phys. **16**, 1816 (1975).

Formal linearization of nonlinear massive representations of the connected Poincaré group

E. Taflin
Département de Physique Théorique, Université de Genève, 1211 Genève 4, Switzerland

(Received 14 March 1983; accepted for publication 2 September 1983)

Let U^1 be an arbitrary finite direct sum of unitary irreducible representations, each of positive (mass)^2, of the connected Poincaré group $P_0 = \mathbb{R}^4 \otimes \mathrm{SL}(2,\mathbb{C})$. It is proved that each nonlinear representation of P_0 with linear part U^1, on the space of differentiable vectors of U^1, is formally linearizable. Further, it is remarked that, commonly used, nonlinear massless representations of $\mathbb{R}^2 \otimes \mathrm{SO}_0(1,1)$ are nonlinearizable and that the corresponding evolution equations do not have covariant wave operators.

PACS numbers: 02.20.Qs, 11.10.Qr, 11.30.Cp

1. INTRODUCTION

Autonomous (i.e., non-time-dependent) evolution equations, currently appearing in mathematical physics, have often, due to the physical systems considered, certain covariance properties, such as Poincaré covariance. It was proposed in Ref. 1 (cf. Refs.2 and 3) how to take the covariance properties explicitly into consideration, when trying to integrate such equations and a theory of nonlinear representations of Lie groups and Lie algebras was initiated. Heuristically a (nonlinear) analytic representation (U,E) of a real Lie group G in a Fréchet space E is a local analytic action U of G in E with a fix point (which we take to be the origin). Let $U = \Sigma_{n\geq 1} U^n$ be the Taylor development of U around the origin in E, where U^n is a monomial of degree n. U^1 is called the linear part of U. U is said to be analytically linearizable, if there exists an analytic map A: $O \to E$, being analytically invertible on some neighborhood $O \subset E$ of the origin and if A intertwines U and U^1, i.e., $A \circ U = U^1 \circ A$. U is said to be formally linearizable if the last equality is true in the sense of formal power series. Let $\mathfrak{g}$ be the Lie algebra of G. As in the case of linear Lie group representations, one can also for the nonlinear case associate a Lie algebra representation (T,E_∞) to (U,E) on the space of differentiable vectors E_∞ of (U^1,E) by $T_X = dU_{\exp(tX)}/dt\,|_{t=0}$ (for details, see Ref. 2). Let $B = T_X$ for some $X \in \mathfrak{g}$. Then the evolution equation $du(t)/dt = B(u(t))$, $u(t) \in E_\infty$, is said to be covariant under the action of $\mathfrak{g}$.[3] Now, if (U,E) is analytically linearizable, then $u(t) = A^{-1} \circ U^1_{\exp(tX)} \circ A(u_0)$, $u_0 \in E_\infty$, is a nontrivial global solution, if the image of A has a subset (different from $\{0\}$) invariant under $U^1_{\exp(tX)}$, $t \in \mathbb{R}$, and if u_0 is appropriately chosen. Tentatively, to construct A, one can now try to profit from the often rich algebraic structure of G (compared with that of $\mathbb{R}$), and divide the proof of the existence into two parts. First prove the existence of a formal power series A intertwining (U,E) and (U^1,E) and second prove the convergence. The formal linearizability of all nonlinear representations of G with a given linear part (U^1,E) is equivalent to the triviality of the one cohomology spaces $H^1(G,\mathscr{L}_n(E))$, $n\geq 2$ (see Ref. 2, Proposition 1), where $\mathscr{L}_n(E)$ are the continuous symmetric n-linear maps of E to E. To what concerns the Poincaré group $P_0 = \mathbb{R}^4 \otimes \mathrm{SL}(2,\mathbb{C})$,[4] the following has been proved:

(i) If U^1 is a unitary representation having a definite energy sign and a finite discrete strictly positive (mass)^2 spectrum, then (U,E) is analytically linearizable.[2,5]

(ii) The analytic representations of the proper analytic subgroups, of dimension at least 3, of $\mathrm{SO}_0(3,1)$ are under very general conditions formally linearizable [analytically for $\mathrm{SO}_0(3)$]. See Ref. 1, Proposition 5 (resp. Refs. 6–8) for $\mathrm{SO}_0(3)$ [resp. $\mathrm{SO}_0(2.1)$, the Euclidean group of the plane E_2, and the affine group of the complex plane W_2].

In the present article we prove, on the domain of differentiable vectors, the *formal* linearizability of all analytic (and formal) representations of the Poincaré group $P_0 = \mathbb{R}^4 \otimes \mathrm{SL}(2,\mathbb{C})$, for which the linear part (U^1,E) is a unitary representation with a finite discrete strictly positive (mass)^2 — spectrum. The difference of this linear part to the one considered in (i) above is the admission of both positive and negative energies. The proof of the triviality of $H^1(P_0,\mathscr{L}_n(E))$, $n\geq 2$, in Refs. 2 and 5 was based on that the intersection of the (mass)^2 — spectra of the linear part (U^1,E) and its symmetrized tensor product $(\otimes^n U^1, \hat{\otimes}^n_s E)$ was empty $\forall n\geq 2$. This is no longer the case when both energy signs are present, and, for proving the triviality of the cohomology, the following strategy is adapted. The cocycles $R \in Z^1(P_0,\mathscr{L}_n(E))$ can be identified with functions from P_0 into $S(\mathbb{R}^3,\mathbb{C}^M) \otimes (\otimes^n_s S'(\mathbb{R}^3,\mathbb{C}^M))$. As for the present linear part (U^1,E), the space of differentiable vectors is isomorphic to a Schwartz space $S(\mathbb{R}^3,\mathbb{C}^M)$, $M \in \mathbb{N}$. Fix a point τ on one of the orbits of U^1. The stability group of τ is SU(2). Then $\mathbb{R}^4 \otimes \mathrm{SU}(2) \times (\hat{\otimes}^n E) \ni (g,F) \mapsto [R_g(F)](\tau) \in \mathbb{C}^m$, where m is the dimension of the inducing representation, defines an element $R' \in Z^1(\mathbb{R}^4 \otimes \mathrm{SU}(2), \mathbb{C}^m \otimes (\hat{\otimes}^n S'(\mathbb{R}^3,\mathbb{C}^M)))$. We prove that R' is a coboundary by observing that $H^1(\mathbb{R}^4 \otimes \mathrm{SU}(2),\mathbb{C}^m \otimes (\hat{\otimes}^n S'(\mathbb{R}^3,\mathbb{C}^M))) = H^1(\mathbb{R}^4,\mathbb{C}^m \otimes (\hat{\otimes}^n S'(\mathbb{R}^{3n},\mathbb{C}^M)))$, as SU(2) is compact, and then solve the coboundary equation on $\mathbb{R}^4$. This last step is equivalent to solving a system of linear equations, with polynomial coefficients, in S'. The compatibility condition, being exactly that R' is a cocycle on $\mathbb{R}^4$, the system has a solution λ. The action of $\mathrm{SL}(2,\mathbb{C})$ on λ defines now an element $A \in S(\mathbb{R}^3,\mathbb{C}^m) \hat{\otimes} (\hat{\otimes}^n S'(\mathbb{R}^3,\mathbb{C}^M))$. By repeating this procedure to each orbit of U^1, we have proved the existence of $A \in \mathscr{L}_n(E)$ having R as image under the coboundary operator, (i.e., A is a potential for R).

For the purpose of proving analytical linearization,

0022-2488/84/040765-07$02.50

there is a serious drawback in the proof of $H^1(P_0,\mathscr{L}_n(E))=\{0\}$, outlined above. To prove the existence of λ, we have used theorems[9,10] on the solution of the division problem in S', so we do not have any estimates on Λ, which should be necessary for a proof of the convergence of the intertwinning formal power series. The vanishing of the cohomology indicates that it is worthwhile to try to find such estimates. Another, perhaps, less serious, problem is that we have used the compactness of the stability group SU(2) and in some way (which at the moment is not very clear to the author) used the simple connectivity of the orbits of U^1. This makes the generalization, to the case of a linear part (U^1,E) with "mass zero orbits," problematic. Finally, the finiteness of the $(\text{mass})^2$ — spectrum was essential.

Covariant wave operators, for relativistic nonlinear wave equations, intertwine the linear and nonlinear representations, naturally associated. A necessary condition of the existence of C^∞ wave operators is then the formal linearizability of the nonlinear representation (cf. Ref. 3). It is not difficult to see (see Sec. 5) that the nonlinear representations of the Poincaré group $\mathbb{R}^2 \circledS \mathrm{SO}_0(1,1)$ in $1+1$ dimension, naturally associated with a large class of massless local nonlinear wave equations, are not formally linearizable. This class of equations, not having covariant wave operators, contains the equations $(\partial_t^2-\partial_x^2)\varphi(x,t)=F(\varphi(x,t))$ $\quad(t,x\in\mathbb{R}$, F analytic with no constant and linear term).

2. MAIN RESULT AND NOTATION

In order to expose the main result of the paper, we first fix necessary notation and reproduce some definitions. Given Fréchet spaces F^1,F^2,F^3, and F, we denote $\mathscr{F}(F^1,F^2)$, the set of symmetric formal power series $A=\Sigma_{n\geqslant 1}A^n$ from F^1 to F^2, where $A^n\in\mathscr{L}_n(F^1,F^2)$, the algebraic vector space of continuous symmetric n-linear maps from F^1 to F^2. We denote $\mathscr{L}_n(F,F)$ by $\mathscr{L}_n(F)$, $\mathscr{L}_1(F^1,F^2)$ by $\mathscr{L}(F^1,F^2)$ and $\mathscr{F}(F,F)$ by $\mathscr{F}(F)$, $F^1\hat{\otimes}F^2$ denotes the completed projective tensor product of F^1 and F^2 and $\hat{\otimes}_s^n F$ the symmetrized n-fold completed projective tensor product of F. We will not distinguish between $A^n\in\mathscr{L}_n(F^1,F^2)$ and $A^n\in\mathscr{L}(\hat{\otimes}_s^n F^1,F^2)$, due to the canonical algebraic isomorphism between these spaces (Ref. 11, Proposition 43.4). The usual composition law of functions gives a composition law $\mathscr{F}(F^1,F^2)\times\mathscr{F}(F^2,F^3)\mapsto\mathscr{F}(F^1,F^3)$ defined by

$$B\circ A=\sum_{\substack{n\geqslant 1\\ 1\leqslant p\leqslant n}}\ \sum_{n_1+\cdots+n_p=n} B^p(A^{n_1}\otimes\cdots\otimes A^{n_p})\sigma_n, \tag{2.1}$$

where $A=\Sigma_{n\geqslant 1}A^n\in\mathscr{F}(F^1,F^2)$, $A^n\in\mathscr{L}(\hat{\otimes}_s^n F^1,F^2)$, $B=\Sigma_{n\geqslant 1}B^n\in\mathscr{F}(F^2,F^3)$, $B^n\in\mathscr{L}(\hat{\otimes}_s^n F^2,F^3)$, and σ_n is the symmetrization operator $\sigma_n(u_1\otimes\cdots\otimes u_n)=(n!)^{-1}\Sigma_{i\in P_n}u_{i_1}\otimes\cdots\otimes u_{i_n}$, $u_1,\dots,u_n\in F^1$, P_n being the group of permutations of n elements. The invertible elements in $\mathscr{F}(F^1,F^2)$ are those A with a (continuously) invertible linear part $A^1\in\mathscr{L}(F^1,F^2)$. $A=\Sigma_{n\geqslant 1}A^n\in\mathscr{F}(F^1,F^2)$ is said to be analytic if, for $A^n\in\mathscr{L}_n(F^1,F^2)$, $\quad A^n(u)=A^n(u,\dots,u)$, the series $A(u)=\Sigma_{n\geqslant 1}A^n(u)$ converges for all u in a neighborhood of zero in F^1. We could here demand convergence only on smaller sets and obtain a weaker notion of analyticity.

Definition 2.1 [1]: A formal representation (U,F) of a real Lie group G in F is a homomorphism $g\to U_g$ from G to the group of invertible elements in $\mathscr{F}(F)$ such that, if $U_g=\Sigma_{n\geqslant 1}U_g^n$, $U_g^n\in\mathscr{L}_n(F)$, then the map $(g,u_1,\dots,u_n)\to U_g^n(u_1,\dots,u_n)$ is continuous from $G\times(X^n F)$ into F for each $n\geqslant 1$. (U^1,F) is called the linear part of (U,F).

Definition 2.2 [1]: An analytic representation (U,F) of G in F is a formal representation for which there exists an open neighborhood V of the identity in G such that, for each $g\in V$, $U_g=\Sigma_{n\geqslant 1}U_g^n$ is an analytic map (from a neighborhood $\mathscr{O}_g$ of the origin in F).

Definition 2.3 [1]: A representation (U,F), formal or analytic, of G is said to be formally linearizable if there exists A invertible in $\mathscr{F}(F)$ such that $A\circ U_g=U_g^1\circ A$, $\forall g\in G$.

Next introduce a continuous linear unitary representation $(U^1,\mathscr{H})$, involving only a finite number of massive fields, of the Poincaré group $P_0=\mathbb{R}^4\circledS\mathrm{SL}(2,\mathbb{C})$ in a Hilbert space $\mathscr{H}$. More explicitly, let $(U^1,\mathscr{H})$ be a finite direct sum of unitary irreducible representations of P_0. Let $(dU^1,\mathscr{H}_\infty)$ be the corresponding representation of the Poincaré Lie algebra $\mathfrak{p}=\mathbb{R}^4 \oplus \mathrm{SL}(2,\mathbb{C})$[12] in the Fréchet space $\mathscr{H}_\infty$ of differentiable vectors of $(U^1,\mathscr{H})$. $\mathscr{H}_\infty$ is isomorphic to a Schwartz space of test functions. Denote by P_μ, $M_{\mu\nu}$, $\mu,\nu=0,1,2,3$, the elements of a standard basis of $\mathfrak{p}$. We make the following assumption on the $(\text{mass})^2$ operator:

$$\text{The spectrum of } -(dU_{P_0})^2+\sum_{i=1}^{3}(dU_{P_i})^2$$
$$\text{is a subset of }]0,\infty[. \tag{2.2}$$

Denoting by $(U^1,\mathscr{H}_\infty)$ the differentiable representation associated with $(U^1,\mathscr{H})$ (see Ref. 13, p. 254), the main result, which will be proved in Secs. 3 and 4, can be formulated as follows:

Theorem 2.4: Any analytic or formal representation $(U,\mathscr{H}_\infty)$, with linear part $(U^1,\mathscr{H}_\infty)$, of P_0 is formally linearizable.

Remark 2.5: Given any analytic representation $(U,\mathscr{H})$ of P_0, with the above linear part $(U^1,\mathscr{H})$. Then $(U,\mathscr{H})$ is analytically equivalent to a smooth representation $(T,\mathscr{H})$ (see Definition 6 of Ref. 1), with linear part $(U^1,\mathscr{H})$. $(T,\mathscr{H})$ has a restriction $(T,\mathscr{H}_\infty)$ (see Proposition 4 of Ref. 1) to which Theorem 2.4 can be applied. In this sense $(U,\mathscr{H})$ is formally linearizable on a dense invariant subset of $\mathscr{H}$.

Remark 2.6: The formal representation $(U,\mathscr{H}_\infty)$ in Theorem 2.4 is C^∞, in the sense that the mapping $P_0\times\mathscr{L}(\hat{\otimes}_s^n\mathscr{H}_\infty,\mathscr{H}_\infty)\ni(g,u_1\otimes\cdots\otimes u_n)\to U_g^n(u_1\otimes\cdots\otimes u_n)\in\mathscr{L}(\hat{\otimes}_s^n\mathscr{H}_\infty,\mathscr{H}_\infty)$ is C^∞, when $\mathscr{L}(\hat{\otimes}_s^n\mathscr{H}_\infty,\mathscr{H}_\infty)$ is given the topology of bounded convergence.[14]

3. TRIVIALITY OF CERTAIN COHOMOLOGICAL SPACES

Let (S_i,H_i), $0\leqslant i\leqslant n$, $n\geqslant 2$, be unitary irreducible representations of P_0, with mass $m_i>0$, spin $s_i\in\frac{1}{2}\mathbb{N}$, and energy sign $\epsilon_i\in\{1,-1\}$, respectively. We will use an explicit realization of (S_i,H_i). As usual SL(2,$\mathbb{C}$) acts on the hyperboloid $V_{m_i,\epsilon_i} = \{(k_0,\mathbf{k}) \in \mathbb{R}^4 | \mathbf{k} \in \mathbb{R}^3, k_0^2 - |\mathbf{k}|^2 = m_i^2, k_0|k_0|^{-1} = \epsilon_i\}$, by $(B,k) \mapsto \lambda(B)k$, where $B \in \mathrm{SL}(2,\mathbb{C})$, $k \in V_{m_i,\epsilon_i}$, and $\lambda{:}\mathrm{SL}(2,\mathbb{C}) \to \mathrm{SO}_0(3,1)$ is the canonical projection. Let E_i be the (finite dimensional) carrier space of an irreducible representation, with spin s_i, of SU(2), the stability group of $(\epsilon_i m_i,0,0,0) \in V_{m_i,\epsilon_i}$. We choose then (S_i,H_i) in the following form:

$$[S_i(a,B)f](k) = e^{iak}V_i(B)f(\lambda(B^{-1})k)$$
$$\text{for } k \in V_{m_i,\epsilon_i},\ (a,B) \in \mathbb{R}^4 \circledS \mathrm{SL}(2,\mathbb{C}), \tag{3.1}$$
$$f \in H_i = L^2(V_{m_i,\epsilon_i}, E_i, \omega_i(\mathbf{k})^{-1}d\mathbf{k}),\ \omega_i(\mathbf{k}) = (m_i^2 + |\mathbf{k}|^2)^{1/2},$$

where V_i is a representation of SL(2,$\mathbb{C}$) in E_i, compatible with the one of SU(2) and $ak = a^0k^0 - \Sigma_{i=1} a^ik^i$. Occasionally, we will consider H_i as a space of functions $\mathbb{R}^3 \to E_i$, by means of the diffeomorphism $V_{m_i,\epsilon_i} \ni (k_0,\mathbf{k}) \to \mathbf{k} \in \mathbb{R}^3$. The space of differentiable vectors $H_{i\infty}$ of (S_i,H_i) can then be identified with $S(\mathbb{R}^3,E_i)$ and we denote by $(S_i,H_{i\infty})$ the differentiable representation associated with (S_i,H_i).

Introduce the representations $(S^1,F^1) = (S_0,H_{0\infty})$ and $(S^2,F^2) = (\otimes_{1<i<n} S_i, \hat{\otimes}_{1<i<n} H_{i\infty})$. (S^2,F^2) is differentiable (see Proposition 4.4.1.10 of Ref. 13). The representation $(T,\mathscr{L}_b(F^2,F^1))$ of P_0 (b stands for bounded convergence) defined by $(g,A) \mapsto T_g(A) = S_g^1AS_{g^{-1}}^2$ is differentiable. In fact, the representation $(T,\mathscr{L}_c(F^2,F^1))$ (c stands for the topology of convergence on compact subsets in F^1) is differentiable (see Ref. 14, p. 14). The statement follows now that F^1 is a Montel space. The space of one cocycles $Z^1(P_0,\mathscr{L}_b(F^2,F^1))$ [resp. the space of 1-coboundaries $B^1(P_0,\mathscr{L}_b(F^2,F^1))$] is the space of all continuous maps R: $P_0 \to \mathscr{L}_b(F^2,F^1)$, such that

$$R_{gg'} = R_g + S_g^1R_{g'}S_{g^{-1}}^2$$
$$[\text{resp. } R_g = S_g^1AS_{g^{-1}}^2 - A,$$
$$\text{for some } A \in \mathscr{L}(F^2,F^1)]. \tag{3.2}$$

$B^1(P_0,\mathscr{L}_b(F^2,F^1))$ is a subspace of $Z^1(P_0,\mathscr{L}_b(F^2,F^1))$. The quotient space $H^1(P_0,\mathscr{L}_b(F^2,F^1)) = Z^1(P_0,\mathscr{L}_b(F^2,F^1))/B^1(P_0,\mathscr{L}_b(F^2,F^1))$ is called the 1-cohomology space.

Remark 3.1: As $(T,\mathscr{L}_b(F^2,F^1))$ is differentiable, it follows (see Ref. 14, Lemma 2, p. 13) that $Z^1(P_0,\mathscr{L}_b(F^2,F^1)) = Z^1_\infty(P_0,\mathscr{L}_b(F^2,F^1))$, i.e., the map $P_0 \ni g \to R_g \in \mathscr{L}_b(F^2,F^1)$ is C^∞ if $R \in Z^1(P_0,\mathscr{L}_b(F^2,F^1))$. Obviously, $B^1(P_0,\mathscr{L}_b(F^2,F^1)) = B^1_\infty(P_0,\mathscr{L}_b(F^2,F^1))$.

In the remaining part of this paragraph, we prove the triviality of $H^1(P_0,\mathscr{L}_b(F^2,F^1))$.

Introduce the real analytic functions f^i: $\mathbb{R}^{3n} \to \mathbb{R}$, $1 \leqslant i \leqslant 3$, defined by

$$f^i(\mathbf{k}_1,\ldots,\mathbf{k}_n) = -\sum_{j=1}^{n} k_j^i,\quad \mathbf{k}_j = (k_j^1,k_j^2,k_j^3) \in \mathbb{R}^3. \tag{3.3}$$

Lemma 3.2: Given distributions $T^i \in S'(\mathbb{R}^{3n},\mathbb{C})$, satisfying $f^iT^j = f^jT^i$, $i,j = 1,2,3$. Then there exists a distribution $B \in S'(\mathbb{R}^{3n},\mathbb{C})$, such that $f^iB = T^i$, $i = 1,2,3$.

Proof: Let $\mathfrak{a}_s$ be the germ of real analytic functions at $s = (\mathbf{s}_1,\ldots,\mathbf{s}_n) \in (\mathbb{R}^3)^n$ (see Ref. 10, p. 23–01; cf. Ref. 16, Chap. II). For every $s \in \mathbb{R}^{3n}$, we will prove that, if $\gamma_i \in \mathfrak{a}_s$, $i = 1,2,3$, satisfy $\Sigma_{i=1}^3 \gamma_i f^i = 0$ (in $\mathfrak{a}_s$), then $\Sigma_{i=1}^3 \gamma_i T^i = 0$ in a neighborhood of s. Theorem 1, p. 23–02, and Sec. 3, p. 25–04 in Ref. 10 guarantee then the existence of B.

For $1 \leqslant i \leqslant n-1$, introduce new coordinates $s' = (\mathbf{s}_1',\ldots,\mathbf{s}_n')$, $\mathbf{s}_i' = \mathbf{s}_i$, $\mathbf{s}_n' = -\Sigma_{j=1}^n \mathbf{s}_j$, $k' = (\mathbf{k}_1',\ldots,\mathbf{k}_n')$, $\mathbf{k}_i' = \mathbf{k}_i$, and $\mathbf{k}_n' = -\Sigma_{j=1}^n \mathbf{k}_j$, where $k = (\mathbf{k}_1,\ldots,\mathbf{k}_n) \in (\mathbb{R}^3)^n$. Transformed functions f'^i, germs $\gamma_i' \in \mathfrak{a}_{s'}$ and distributions T'^i are defined by $f'^i(\mathbf{k}_1',\ldots,\mathbf{k}_n') = f^i(\mathbf{k}_1,\ldots,\mathbf{k}_n)$, $\gamma_i'(\mathbf{k}_1',\ldots,\mathbf{k}_n') = \gamma_i(\mathbf{k}_1,\ldots,\mathbf{k}_n)$, and $T'^i(\mathbf{k}_1',\ldots,\mathbf{k}_n') = T^i(\mathbf{k}_1,\ldots,\mathbf{k}_n)$, respectively. Then $f'^i(\mathbf{k}_1',\ldots,\mathbf{k}_n') = k_n'^i$.

If one of the $s_n'^i$, $i = 1,2,3$, let us say $s_n'^1$, is different from zero, then f'^1 is a unit in $\mathfrak{a}_{s'}$, so $\gamma_1' = -(f'^1)(\gamma_2'f'^2 + \gamma_3'f'^3)$. Using the hypothesis of Lemma 3.2, we get

$$\begin{aligned}\sum_{i=1}^{3} \gamma_i'T'^i &= -(f'^1)^{-1}(\gamma_2'f'^2 + \gamma_3'f'^3)T'^1 + \gamma_2T'^2 + \gamma_3T'^3\\ &= -(f'^1)^{-1}(\gamma_2'f'^1T'^2 + \gamma_2'f'^1T'^3)\\ &\quad + \gamma_2'T'^2 + \gamma_3'T'^3 = 0.\end{aligned}$$

Let $\mathbf{s}_n' = 0$, and let g_1,g_2,g_3 be elements in the equivalence classes $\gamma_1',\gamma_2',\gamma_3'$, respectively. If $\bar{k} = (\mathbf{k}_1',\ldots,\mathbf{k}_{n-1}')$, then in some neighborhood of $s' = (\mathbf{s}_1',\ldots,\mathbf{s}_{n-1}',0)$, g_i can be expanded into a convergent power series

$$g_i(k') = \sum_{\alpha_1,\alpha_2,\alpha_3 \geqslant 0} h_i^{\alpha_1,\alpha_2,\alpha_3}(\bar{k})(f'^1(k'))^{\alpha_1}(f'^2(k'))^{\alpha_2}(f'^3(k'))^{\alpha_3}, \tag{3.4}$$

where $h_i^{\alpha_1,\alpha_2,\alpha_3}$ are real analytic functions in a neighborhood $\mathscr{O}$ of $\bar{s} = (\mathbf{s}_1',\ldots,\mathbf{s}_{n-1}')$ (independent of $i,\alpha_1,\alpha_2,\alpha_3$). Substitution of (3.4) in $\Sigma_{i=1}^3 g_u f'^i = 0$, in a neighborhood of s', and identification of the coefficients give in $\mathscr{O}$

$$h_1^{\alpha,0,0} = h_2^{0,\alpha,0} = h_3^{0,0,\alpha},\quad \forall\alpha \geqslant 0, \tag{3.5a}$$
$$h_1^{0,0,1} + h_3^{1,0,0} = h_1^{0,1,0} + h_2^{1,0,0} = h_2^{0,0,1} + h_3^{0,1,0} = 0, \tag{3.5b}$$
$$h_1^{\alpha_1-1,\alpha_2,\alpha_3} + h_2^{\alpha_1,\alpha_2-1,\alpha_3,} + h_3^{\alpha_1,\alpha_2,\alpha_3-1} = 0,\quad \forall\alpha_1,\alpha_2,\alpha_3 \geqslant 1. \tag{3.5c}$$

(3.4) and (3.5a) give in a neighborhood of s'

$$\begin{aligned}\sum_{1\leqslant i\leqslant 3} g_iT'^i &= (h_1^{0,1,0}f'^2 + h_1^{0,0,1}f'^3)T'^1 + (h_2^{1,0,0}f'^1\\ &\quad + h_2^{0,0,1}f'^3)T'^2 + (h_3^{1,0,0}f'^1 + h_3^{0,1,0}f'^2)T'^3\\ &\quad + \sum_{\alpha_1,\alpha_2,\alpha_3 \geqslant 1} [h_1^{\alpha_1-1,\alpha_2,\alpha_3}(f'^1)^{\alpha_1-1}\\ &\quad \times (f'^2)^{\alpha_2}(f'^3)^{\alpha_3}T'^1\\ &\quad + h_2^{\alpha_1,\alpha_2-1,\alpha_3}(f'^1)^{\alpha_1}(f'^2)^{\alpha_2-1}(f'^3)^{\alpha_3}T'^2\\ &\quad + h_3^{\alpha_1,\alpha_2,\alpha_3-1}(f'^1)^{\alpha_1}(f'^2)^{\alpha_2}(f'^3)^{\alpha_3-1}T'^3].\end{aligned} \tag{3.6}$$

The hypothesis $f'^iT'^j = f'^jT'^i$, for $i,j = 1,2,3$, and then (3.5b) give

$$\begin{aligned}&(h_1^{0,1,0}f'^2 + h_1^{0,0,1}f'^3)T'^1 + (h_2^{1,0,0}f'^1 + h_2^{0,0,1}f'^3)T'^2\\ &\quad + (h_3^{1,0,0}f'^1 + h_3^{0,1,0}f'^2)T'^3\\ &= (h_1^{0,1,0}f'^2T'^2 + h_1^{0,0,1}f'^3T'^1)\\ &\quad + (h_2^{1,0,0}f'^1T'^2 + h_2^{0,0,1}f'^2T'^3)\\ &\quad + (h_3^{1,0,0}f'^3T'^1 + h_3^{0,1,0}f'^2T'^3)\\ &= (h_1^{0,1,0} + h_2^{1,0,0})f'^1T'^2\\ &\quad + (h_2^{0,0,1} + h_3^{0,1,0})f'^2T'^3\\ &\quad + (h_3^{1,0,0} + h_1^{0,0,1})f'^3T'^1 = 0.\end{aligned} \tag{3.7}$$

The hypothesis and (3.5c) give, for $\alpha_1,\alpha_2,\alpha_3\geqslant 1$,

$$\begin{aligned}(f'^1)^{\alpha_1-1}(f'^2)^{\alpha_2-1}(f'^3)^{\alpha_3-1}\big[h_1^{\alpha_1-1,\alpha_2,\alpha_3}f'^2f'^3T'^1 \\ + h_2^{\alpha_1,\alpha_2-1,\alpha_3}f'^1f'^3T'^2 + h_3^{\alpha_1,\alpha_2,\alpha_3-1}f'^1f'^2T'^3\big] \quad (3.8)\\ = (f'^1)^{\alpha_1-1}(f'^2)^{\alpha_1-1}(f'^3)^{\alpha_3-1}\big[h_1^{\alpha_1-1,\alpha_2,\alpha_3} \\ + h_2^{\alpha_1,\alpha_2-1,\alpha_3} + h_3^{\alpha_1,\alpha_2,\alpha_3-1}\big]f'^1f'^2T'^3 = 0.\end{aligned}$$

(3.6), (3.7), and (3.8) prove that $\Sigma_{1\leqslant i\leqslant 3}\, g_i T'^i = 0$ in a neighborhood of s'. Hence, the existence of B follows from the above cited result of Ref. 10. Q.E.D.

Let f^0: $(\mathbb{R}^3)^n \to \mathbb{R}$ be the real analytic function defined by

$$f^0(\mathbf{k}_1,\ldots,\mathbf{k}_n) = \epsilon_0 m_0 - \sum_{i=1}^{n} \epsilon_i \omega_i(\mathbf{k}_i), \quad n\geqslant 2. \tag{3.9}$$

Lemma 3.3: Given a distribution $T\in S'(\mathbb{R}^{3n},\mathbb{C})$ satisfying $f^i T=0$, for $i=1,2,3$. Then there exists a distribution $B\in S'(\mathbb{R}^{3n},\mathbb{C})$ satisfying $f^0B=T$ and $f^iB=0$, for $i=1,2,3$.

Proof: For $1\leqslant j\leqslant n$, introduce the isomorphism $Q_j: S(\mathbb{R}^3,\mathbb{C})\to S(\mathbb{R}^3,\mathbb{R})\oplus S(\mathbb{R}^3,\mathbb{R})$ by $Q_j\varphi = (2\,\mathrm{Re}\,\varphi, -2\epsilon_j\omega_j\,\mathrm{Im}\,\varphi)$. The image, of the map $S(\mathbb{R}^3,\mathbb{C})\ni\varphi + i\epsilon_j\omega_j\varphi\in S(\mathbb{R}^3,\mathbb{C})$ ($i^2=-1$) under Q_j is the map $S(\mathbb{R}^3,\mathbb{R})\oplus S(\mathbb{R}^3,\mathbb{R})\ni(\psi_1,\psi_2)\to(\psi_2,\omega_j^2\psi_1)$. Note that ω_j^2 is a polynomial. The equation $if^0B=iT$ is equivalent to the equation $(if^0B)(\otimes_{j=1}^n Q_j^{-1}) = (iT)(\otimes_{j=1}^n Q_j^{-1})$, where the rhs and the lhs are complex valued continuous linear functionals on $\hat{\otimes}^n(S(\mathbb{R}^3,\mathbb{R})\oplus S(\mathbb{R}^3,\mathbb{R})) = S(\mathbb{R}^{3n},\mathbb{R})\oplus\cdots\oplus S(\mathbb{R}^{3n},\mathbb{R})$ (2^n copies). A complex valued continuous linear functional F on $S(\mathbb{R}^{3n},\mathbb{R})$, is identified with $F\in S'(\mathbb{R}^{3n},\mathbb{C})$. There are then $T',B'\in S'(\mathbb{R}^{3n},\mathbb{C}^{2n})$, such that $(iT)(\otimes_{j=1}^n Q_j^{-1}) = T'$ and

$$\begin{aligned}(if^0B)\Big(\bigotimes_{j=1}^n Q_j^{-1}\Big) &= (im_0\epsilon_0 B)\Big(\bigotimes_{j=1}^n Q_j^{-1}\Big) \\ &\quad - \sum_{l=1}^n \Omega_l\Big(B\Big(\bigotimes_{j=1}^n Q_j^{-1}\Big)\Big) \\ &= \Big(im_0\epsilon_0 - \sum_{l=1}^n \Omega_l\Big)B\Big(\bigotimes_{j=1}^n Q_j^{-1}\Big) \\ &= \Big(im_0\epsilon_0 I - \sum_{l=1}^n \Omega_l\Big)B',\end{aligned}$$

where

$$\Omega_l = I\otimes\cdots\otimes I\otimes\begin{pmatrix}0 & 1\\ -\omega_j^2 & 0\end{pmatrix}\otimes I\otimes\cdots\otimes I,\quad I=\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

$\Lambda = (im_0\epsilon_0 I - \Sigma_{l=1}^n \Omega_l)$ is a $2^n\times 2^n$ matrix of polynomials in $\mathbf{k}_1,\ldots,\mathbf{k}_n$. Hence we have to prove the existence of $B'\in S'(\mathbb{R}^{3n},\mathbb{C}^{2n})$, such that $\Lambda B'=T'$ and $f^iB'=0$ for $i=1,2,3$, under the condition $f^iT'=0$. Now, $f^iT'=0$ for $i=1,2,3$, so there is (Ref. 15, p. 101) a distribution $G\in S'(\mathbb{R}^{3(n-1)},\mathbb{C}^{2n})$, such that $T'u=Gv$, for all $u\in S(\mathbb{R}^{3n},\mathbb{C})$, where $v\in S(\mathbb{R}^{3(n-1)},\mathbb{C})$ and $v(\mathbf{k}_1,\ldots,\mathbf{k}_{n-1}) = u(\mathbf{k}_1,\ldots,\mathbf{k}_{n-1}, -\Sigma_{j=1}^{n-1}\mathbf{k}_j)$. Similarly put $B'u=Fv$, for $F\in S'(\mathbb{R}^{3(n-1)},\mathbb{C}^{2n})$. The equation $f^iB'=0$, $i=1,2,3$, is then satisfied and the equation $\Lambda B'=T'$ is equivalent to $PF=G$, where $P(\mathbf{k}_1,\ldots,\mathbf{k}_{n-1}) = \Lambda(\mathbf{k}_1,\ldots,\mathbf{k}_{n-1}, -\Sigma_{j=1}^{n-1}\mathbf{k}_j)$. P is a $2^n\times 2^n$ matrix with polynomial coefficients. The determinant of P is not the zero polynomial. In fact, f^0 is not the zero function. By Ref. 9 there exists then a solution F of $PF=G$ in $S'(\mathbb{R}^{3(n-1)},\mathbb{C}^{2n})$. Q.E.D.

Lemma 3.4: Let $T^\mu\in S'(\mathbb{R}^{3n},\mathbb{C})$ satisfy $f^\mu T^\nu = f^\nu T^\mu$, for $\mu,\nu=0,1,2,3$. Then there exists $B\in S'(\mathbb{R}^{3n},\mathbb{C})$ satisfying $f^\mu B=T^\mu$, for $\mu=0,1,2,3$.

Proof: By Lemma 3.2 there exists a solution $C\in S'(\mathbb{R}^{3n},\mathbb{C})$ of the system $f^iC=T^i$, $i=1,2,3$. Define $T'^\mu = T^\mu - f^\mu C$, $\mu=0,1,2,3$. As $T'^i=0$, $i=1,2,3$, there is, by Lemma 3.3, a solution $C'\in S'(\mathbb{R}^{3n},\mathbb{C})$ of the system $f^\mu C' = T'^\mu$, $\mu=0,1,2,3$. $B=C+C'$ has the announced properties. Q.E.D.

Let $\tau=(\epsilon_0 m_0,0,0,0)$. The representation $(\theta,\mathscr{L}_b(F^2,E_0))$, of $\mathbb{R}^4 \circledS \mathrm{SU}(2)$, is defined by $(\mathbb{R}^4\circledS\mathrm{SU}(2))\times\mathscr{L}_b(F^2,E_0)\ni(g,A_0)\to S_g^1(\tau)A_0S_{g^{-1}}^2\in\mathscr{L}_b(F^2,E_0)$, where $S_g^1(\tau)$ is the inducing representation of the stability group $\mathbb{R}^4\circledS\mathrm{SU}(2)$ of τ in E_0, i.e., $S^1_{(a,B)}(\tau)e = V_0(B)e^{i\epsilon_0 m_0 a^0}e$, $(a,B)\in\mathbb{R}^4\circledS\mathrm{SU}(2)$. θ is differentiable. As above, we introduce, for the representation $(\theta,\mathscr{L}_b(F^2,E_0))$, the cohomology $H_1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0))$.

Remark 3.5: Similarly, as in Remark 3.1, it follows that $Z^1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0)) = Z^1_\infty(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0))$ and that $B^1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0)) = B^1_\infty(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0))$.

For the sake of completeness, we prove the following lemma, which is contained in Proposition 5 of Ref. 1:

Lemma 3.6: Let $R\in Z^1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0))$. Then $R_g=0$, mod $B^1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0))$, for $g=(0,B)$, $B\in\mathrm{SU}(2)$.

Proof: Let μ be the normalized invariant measure on SU(2). Define $A=\int_{\mathrm{SU}(2)}R_{(0,B)}\,d\mu(B)\in\mathscr{L}_b(F^2,E_0)$. R is equal, mod B^1 to $g\to R_g+\theta_g(A)-A$. But $R_g+\theta_g(A)-A=0$ if $g=(0,B)$, $B\in\mathrm{SU}(2)$. Q.E.D.

Proposition 3.7: $H^1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0)) = \{0\}$.

Proof: Let $R\in Z^1(\mathbb{R}^4\circledS\mathrm{SU}(2),\mathscr{L}_b(F^2,E_0))$. We have to prove that R is a coboundary. By Lemma 3.6, there is no restriction to choose $R_g=0$ for $g=(0,B)$, $B\in\mathrm{SU}(2)$. The map $\mathbb{R}^4\circledS\mathrm{SU}(2)\ni g\to R_g\in\mathscr{L}_b(F^2,E_0)$ is differentiable. Define $\xi_X = dR_{\exp(tX)}/dt\,|_{t=0}\in\mathscr{L}(F^2,E_0)$ for all $X\in\mathbb{R}^4\oplus\mathfrak{su}(2)$, the Lie algebra of $\mathbb{R}^4\circledS\mathrm{SU}(2)$. As $\mathbb{R}^4\circledS\mathrm{SU}(2)$ is simple connected, R is a coboundary if and only if (see Ref. 17, Proposition 1.2) there exists $A\in\mathscr{L}(F^2,E_0)$ such that

$$d(S^1(\tau))_X A - \mathrm{Ad}(\otimes^n S^2)_X = \xi_X,\quad \forall X\in\mathbb{R}^4\oplus\mathfrak{su}(2). \tag{3.10}$$

Note that $\xi_X=0$ for all $X=(0,b)$, $b\in\mathrm{SU}(2)$, and that if A_0 is a solution of (3.10) on the ideal $\mathbb{R}^4$, then $A=\int_{\mathrm{SU}(2)}d\mu(B)\,\theta_{(0,B)}A_0$, where μ is the normalized invariant measure on SU(2), is a solution on $\mathbb{R}^4\circledS\mathrm{SU}(2)$.

We have the following topological isomorphism:

$$\begin{aligned}\mathscr{L}_b(F^2,E_0) &\cong \mathscr{L}_b\Big(\hat{\bigotimes_{1\leqslant i\leqslant n}} S(\mathbb{R}^3,E_i),E_0\Big) \\ &\cong \Big(\hat{\bigotimes_{1\leqslant i\leqslant n}} S(\mathbb{R}^3,E_i)\Big)'\otimes E_0 \\ &\cong \Big(\hat{\bigotimes_{1\leqslant i\leqslant n}} (S(\mathbb{R}^3,\mathbb{C})\otimes E_i)\Big)'^{\otimes E_0} \\ &\cong \Big(S(\mathbb{R}^{3n},\mathbb{C})\otimes\Big(\bigotimes_{1\leqslant i\leqslant n} E_i\Big)\Big)'\otimes E_0 \\ &\cong S'(\mathbb{R}^{3n},\mathbb{C})\otimes E_0\otimes\Big(\bigotimes_{1\leqslant i\leqslant n} E_i'\Big),\end{aligned}$$

(for isomorphism numbers 2, 3, 4, and 5 see, respectively, pp.

525, 533, 530, and 525 of Ref. 11), where $S'(\mathbb{R}^{2n},\mathbb{C})$ is the strong dual of $S(\mathbb{R}^{3n},\mathbb{C})$. Let $\{e_0^j\}_{j=1}^{\dim E_0}$ and $\{e_i^j\}_{j=1}^{\dim E_i'}$, $1\leqslant i\leqslant n$, be bases of E_0 and E_i', respectively. Due to the above isomorphism, if $C\in\mathscr{L}_b(F^2,E_0)$, then

$$C=\sum_{i\leqslant j_0\leqslant\dim E_0}\sum_{1\leqslant i\leqslant n}\sum_{1\leqslant j_i\leqslant\dim E_i'}C(j_0;j_1,\dots,j_n)e_0^{j_0}e_1^{j_1}\cdots e_n^{j_n},$$

where $C(j_0;j_1,\dots,j_n)\in S'(\mathbb{R}^{3n},\mathbb{C})$. The restriction of (3.10) to the ideal $\mathbb{R}^4$ gives then

$$if^\mu A_0(j_0;j_1,\dots,j_n)=\xi_{P_\mu}(j_0;j_1,\dots,j_n),$$
$$\forall_0\leqslant\mu\leqslant 3,\quad 1\leqslant j_0\leqslant\dim E_0, \tag{3.11}$$

and $1\leqslant j_i\leqslant\dim E_i'$, $1\leqslant i\leqslant n$. By differentiation of the cocycle equation $R_{gg'}=R_g+\theta(g)R_{g'}$, it follows that $f^\mu\xi_{P_\nu}(j)=f^\nu\xi_{P_\mu}(j)$, $\forall j=(j_0;j_1,\dots,j_n)$. There is then, by Lemma 3.4, a solution $A_0(j)\in S'(\mathbb{R}^{3n},\mathbb{C})$, for each possible $j=(j_0;j_1,\dots,j_n)$. The corresponding $A_0\in\mathscr{L}(F^2,E_0)$ solves (3.10) on the ideal $\mathbb{R}^4$. Hence there exists $A\in\mathscr{L}(F^2,E_0)$ solving (3.10) for all $X\in\mathbb{R}^4\otimes\mathfrak{su}(2)$, which proves that R is a coboundary. Q.E.D.

For $k\in V_{m_0,\epsilon_0}$, let $B_k\in SL(2,\mathbb{C})$ be such that $\lambda(B_k)\tau=k$ and let $\mathbb{R}^3\ni\mathbf{k}\to k=(\epsilon_0\omega_0(\mathbf{k}),\mathbf{k})\to B_k\in SL(2,\mathbb{C})$ be C^∞. This is always possible, as seen from the choice [see (6.34) of Ref. 18]:

$$B_k=\frac{\epsilon_0 m_0+\tilde{k}}{(2\epsilon_0 m_0(k^0+\epsilon_0 m_0))^{1/2}},\quad \tilde{k}=\sum_{\mu=0}^{3}k^\mu\sigma_\mu, \tag{3.12}$$

where the Pauli matrices are

$$\sigma_0=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},$$
$$\sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\quad \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix}.$$

The corresponding transformation in $\mathbb{R}^4$ is [see (6.33) of Ref. 18], if $p=(p^0,p^1,p^2,p^3)=(p_0,\mathbf{p})\in\mathbb{R}^4$:

$$(\lambda(B_k)p)^0=(\epsilon_0 m_0)^{-1}\sum_{\mu=0}^{3}k^\mu p^\mu,$$
$$(\lambda(B_k)p)^i=p^i+(\epsilon_0 m_0)^{-1}k^i\Big[(k^0+\epsilon_0 m_0)^{-1}\times\sum_{j=1}^{3}k^jp^j+p^0\Big],\quad 1\leqslant i\leqslant 3. \tag{3.13}$$

Each $R\in Z^1(P_0,\mathscr{L}_b(F^2,F^1))$ has a restriction $R^0\in Z^1(\mathbb{R}^4\otimes SU(2),\mathscr{L}_b(F^2,E_0))$, defined by $R_g^0(f)=[R_g(f)](\tau)$, $f\in F^2$, and g in the subgroup $\mathbb{R}^4\otimes SU(2)$. By Proposition 3.7, there is, for each such R, $A^0\in\mathscr{L}(F^2,E_0)$ satisfying

$$R_g^0=\theta_g A^0-A^0,\quad \forall g\in\mathbb{R}^4\otimes SU(2). \tag{3.14}$$

For the couple (R,A^0) define A by

$$[Af](k)=V_0(B_k)\{[R_{g_k^{-1}}S^2_{g_k^{-1}}f](\tau)+A^0S^2_{g_k^{-1}}f\},\quad \forall f\in F^2,\ k\in V_{m_0,\epsilon_0}, \tag{3.15}$$

where $g_k=(0,B_k)$ and B_k is given by (3.12). By the differentiability properties of S^1, S^2, R, and B it follows that $A\in\mathscr{L}(F^2,C_b^\infty(V_{m_0,\epsilon_0},E_0))$. In the next few lemmas, we prove that $R_g=S_g^1AS^2_{g^{-1}}-A$, $g\in P_0$, when $R_g\ (\in\mathscr{L}(F^2,F^1))$ is considered as an element in $\mathscr{L}(F^2,C_b^\infty(V_{m_0,\epsilon_0},E_0))$. For consistency, we note that the representation (S^1,F^1) of P_0 has a unique extension to a differentiable representation $(S^1,C_b^\infty(V_{m_0,\epsilon_0},E_0))$.

Lemma 3.8: Let $R\in Z^1(P_0,\mathscr{L}_b(F^2,F^1))$ be zero on the subgroup SU(2) and consider R valued in $C_b^\infty(V_{m_0,\epsilon_0},E_0)$. If A^0 is defined by (3.14), $g=(0,B)$, $B\in SL(2,\mathbb{C})$, $h_1^{-1}=(0,B_k^{-1}B^{-1}B_{\lambda(B)k})$ and $h_2=(0,B^{-1}_{\lambda(B)k})$, then

(i) $V_0(B_k)A^0S^2_{g_k^{-1}}S^2_{g^{-1}}=V_0(B^{-1})V_0(B_{\lambda(B)k})A^0S^2_{h_2}$, $\forall k\in V_{m_0,\epsilon_0}$,

(ii) $V_0(B_k)(R_{g_k^{-1}}S^2_{g_k^{-1}}S^2_{g^{-1}}f)(\tau)$
$=V_0(B^{-1})V_0(B_{\lambda(B)k})(R_{h_2}S^2_{h_2}f)(\tau)$
$+(S^1_{g^{-1}}R_gf)(k)$, $\forall f\in F^2$, $k\in V_{m_0,\epsilon_0}$.

Proof: We first note that $h_1=h_2gg_k$ and that $B_k^{-1}B^{-1}B_{\lambda(B)k}\in SU(2)$. In fact, the last property follows from that $\lambda(B_k^{-1})\lambda(B^{-1})\lambda(B_{\lambda(B)k})\tau=\lambda(B_k^{-1})\lambda(B^{-1})\lambda(B)k=\lambda(B_k^{-1})k=\tau$ and that the stability group of τ is SU(2).

(i) $V_0(B_k)A^0S^2_{g_k^{-1}}S^2_{g^{-1}}=V_0(B_k)A^0S^2_{h_1^{-1}}S^2_{h_2}$. The rhs of this expression, by using $A^0S^2_{h_1^{-1}}=V_0(B_k^{-1}B^{-1}B_{\lambda(B)k})(A^0+R^0_{h_1})$ [see (3.14)] can be written $V_0(B^{-1})V^0(B_{\lambda(B)k})(A^0+R^0_{h_1})S^2_{h_2}$. The first part of the lemma follows now from $R^0_{h_1}=0$, as by hypothesis $R=0$ on SU(2).

(ii) Let $f\in F^2$. Then $V_0(B_k)(R_{g_k^{-1}}S^2_{h_k^{-1}}S^2_{g^{-1}}f)(\tau)=V_0(B_k)(R_{g_k^{-1}}S^2_{h_1^{-1}}S^2_{h_2}f)(\tau)$. The rhs of this equality can, by using (3.2) and $R_{h_1}=0$, be written $V_0(B_k)V_0(B_k^{-1}B^{-1}B_{\lambda(B)k})(R_{h_1g_k^{-1}}S^2_{h_2}f)(\tau)$. As $h_1g_k^{-1}=h_2g$ and $R_{h_2g}=R_{h_2}+S^1_{h_2}R_gS^2_{h_2^{-1}}$, this gives, by the definition of S^1, that $V_0(B_k)(R_{g_k^{-1}}S^2_{g_k^{-1}}S^2_{g^{-1}}f)(\tau)=V_0(B^{-1})[V_0(B_{\lambda(B)k})(R_{h_2}S^2_{h_2}g)(\tau)+(R_gf)(\lambda(B)k)]$. Using once more the definition of S^1, one gets the second statement of the lemma. Q.E.D.

Lemma 3.9: Let R be as in Lemma 3.8 and $g=(0,B)$, $B\in SL(2,\mathbb{C})$. If A is given by (3.15), then $R_g=S_g^1AS^2_{g^{-1}}-A$.

Proof: We get by (3.15) and Lemma 3.8, for $f\in F^2$, that $[AS^2_{g^{-1}}f](k)=V_0(B_k)(R_{g_k^{-1}}S^2_{g_k^{-1}}S^2_{g^{-1}}f)(\tau)+V_0(B_k)A^0S^2_{g_k^{-1}}S^2_gf=V_0(B^{-1})V_0(B_{\lambda(B)k})[A^0S^2_{h_2}f+(R_{h_2}S^2_{h_2}f)(\tau)]+(S^1_{g^{-1}}R_gf)(k)=(S^1_{g^{-1}}Af)(k)+(S^1_{g^{-1}}R_gf)(k)$. Q.E.D.

Lemma 3.10: Let R and A^0 be as in Lemma 3.8 and let $a\in\mathbb{R}^4$, $g=(a,I)$, $h=(\lambda(B_k^{-1})a,I)$, $u(a)=\exp(ia\tau)$, where $I\in SL(2,\mathbb{C})$ is the identity element. Then, for $f\in F^2$,

(i) $V_0(B_k)A^0S^2_{g_k^{-1}}S^2_{g^{-1}}f=u(-\lambda(B_k^{-1})a)\times V_0(B_k)(A^0S^2_{g_k^{-1}}f+(R_hS^2_{g_k^{-1}}f)(\tau))$,

(ii) $V_0(B_k)(R_{g_k^{-1}}S^2_{g_k^{-1}}S^2_{g^{-1}}f)(\tau)=u(-\lambda(B_k^{-1})a)\times V_0(B_k)\big(R_{g_k^{-1}}S^2_{g_k^{-1}}f-R_hS^2_{g_k^{-1}}f\big)(\tau)+(S^1_{g^{-1}}R_gf)(k)$.

Proof: We observe that $gg_k=g_kh$. Statement (i) follows then directly from (3.14) and statement (ii) follows from repeated use of (3.2) in the following calculus:

$$
\begin{aligned}
&V_0(B_k)(R_{g_k^{-1}}S^2_{g_k^{-1}}S^2_{g^{-1}} f)(\tau)\\
&= V_0(B_k)(R_{g_k^{-1}}S^2_{h^{-1}}S^2_{g_k^{-1}} f)(\tau)\\
&= V_0(B_k)(S^1_{h^{-1}}R_{hg_k^{-1}}S^2_{g_k^{-1}} f - S^1_{h^{-1}}R_h S^2_{g_k^{-1}} f)(\tau)\\
&= V_0(B_k)(S^1_{h^{-1}}R_{g_k^{-1}}S^2_{g_k^{-1}} f + S^1_{h^{-1}g_k^{-1}}R_g f\\
&\quad - S^1_{h^{-1}}R_h S^2_{g_k^{-1}} f)(\tau) = u(-\lambda(B_k^{-1})a)\\
&\quad \times V_0(B_k)(R_{g_k^{-1}}S^2_{g_k^{-1}} f\\
&\quad - R_h S^2_{g_k^{-1}} f)(\tau) + (S^1_{g^{-1}}R_g f)(k). \qquad \text{Q.E.D.}
\end{aligned}
$$

Lemma 3.11: Let R be as in Lemma 3.8, $g=(a,I)$, $a\in\mathbb{R}^4$, and I the identity in SL(2,$\mathbb{C}$). If A is given by (3.15), then $R_g = S^1_g A S^2_{g^{-1}} - A$.

Proof: For $f\in F^2$, we get, by (3.15) and Lemma 3.10, that $(AS^2_{g^{-1}} f)(k) = u(-\lambda(B_k^{-1}a))V_0(B_k)\big[(R_{g_k^{-1}}S^2_{g_k^{-1}} f)(\tau) - (R_h S^2_{g_k^{-1}} f)(\tau) + A^0 S^2_{g_k^{-1}} f + (R_h S^2_{g_k^{-1}} f)(\tau)\big] + (S^1_{g^{-1}}R_g f)(k) = (S^1_{g^{-1}}(A+R_g)f)(k)$. Q.E.D.

Proposition 3.12: Let $R\in Z^1(P_0,\mathscr{L}_b(F^2,F^1))$ be zero on the subgroup SU(2) and consider R valued in $C_b^\infty(V_{m_0,\epsilon_0},E_0)$. If A is given by (3.15), then $R_g = S^1_g A S^2_{g^{-1}} - A$, for every $g\in P_0$.

Proof: As $g=(a,B)=(a,I)(0,B)$, the results follow from (3.2), Lemma 3.9, and Lemma 3.11. Q.E.D.

Proposition 3.13: Let $R\in Z^1(P_0,\mathscr{L}_b(F^2,F^2))$, be zero on the subgroup SU(2). Then A, defined by (3.15), is an element of $\mathscr{L}(F^2,F^1)$.

Proof: For every seminorm q_1 (resp. q_2) in F^1 (resp. F^2), there is a seminorm p_1 (resp. p_2) in F^1 (resp. F^2) and a positive number c_1 (resp. c_2) depending on $X\in\mathfrak{p}$, such that $q_1(S^1_{\exp(tX)} f_1)\leqslant p_1(f_1)\exp(c_1|t|)$ [resp. $q_2(S^2_{\exp(tX)} f_2)\leqslant p_2(f_2)\exp(c_2|t|)$] for all $t\in\mathbb{R}$ and $f_1\in F^1$ (resp. $f_2\in F^2$). This follows, for example, from the definition of (S^1,F^1) (resp. (S^2,F^2)). By Remark 3.1, we can define $\xi_X = dR_{\exp(tX)}/dt\,|_{t=0}\in\mathscr{L}(F^2,F^1)$. Differentiation of the cocycle equation (1.3) and then integration gives that $R_{\exp(tX)} f = \int_0^t ds\, S^1_{\exp(sX)}\,\xi_X S^2_{\exp(-sX)} f$, for $f\in F^2$. Here the integral is that of the Riemann integral in the topology of F^1. This, integral representation of R, and the above estimates show that, for each seminorm p in F^1 and each $X\in\mathfrak{p}$, there is a seminorm q in F^2 and a positive number c, such that $p(R_{\exp(tX)} f)\leqslant\exp(c|t|)q(f)$, for all $f\in F^2$ and $t\in\mathbb{R}$. In particular choose $p(u)=|u(\tau)|_{E_0}$, $u\in F^1$, $X=(0,\sigma_3)$ and $t=-\ln(m_0^{-1}\epsilon_0 k^3 t(\epsilon_0 k^0))$, where $k=(k^0,0,0,k^3)\in V_{m_0,\epsilon_0}$. Then $g_k^{-1}=\exp(tX)$, which substituted into the estimate of R gives

$$|(V_0(B_k)R_{g_k^{-1}}S^2_{g_k^{-1}} f)(\tau)|_{E_0}\leqslant\tfrac{1}{2}|k^0|^c q(f),\quad \forall f\in F^2, \tag{3.16}$$

for some $c>0$ and seminorm q on F^2. Choosing c and q sufficiently large, it follows from the continuity of A^0 [in (3.15)] that

$$|V_0(B_k)A^0 S^2_{g_k^{-1}} f|\leqslant\tfrac{1}{2}|k^0|^c q(f),\quad \forall f\in F^2. \tag{3.17}$$

From the action of SU(2), we can conclude that estimates (3.16) and (3.17) are true for all $k\in V_{m_0,\epsilon_0}$, and further from (3.15) that

$$|(Af)(k)|_{E_0}\leqslant|k^0|^c q(f),\quad \forall f\in F^2,\quad k\in V_{m_0,\epsilon_0}. \tag{3.18}$$

Next, the equation $dS^1_X A - A dS^2_X = \xi_X$, $X\in\mathfrak{p}$ (see Proposition 3.12) yields $(dS^1_X)^n A = \Sigma_{p=0}^{n-1}(dS^1_X)^p \times\xi_X(dS^2_X)^{n-p-1} + A(dS^2_X)^n$, for $n\geqslant 1$. In particular, for $X=P_0$, we get, after division by $(ik^0)^n$,

$$
\begin{aligned}
(Af)(k) = {}&\left(\sum_{p=0}^{n-1}((dS^1_{P_0})^p)\xi_{P_0}(dS^2_{P_0})^{n-p-1}f\right)(k)\\
&+(A(dS^2_{P_0})^n f)(k)(ik^0)^{-n},
\end{aligned}
\tag{3.19}
$$

for $k\in V_{m_0,\epsilon_0}$, $f\in F^2$, and $n\geqslant 1$. (3.19) shows that $A\in\mathscr{L}(F^2,H_0)$. In fact, the first term on the rhs defines an element in $\mathscr{L}(F^2,F^1)$ and (3.18) gives that $|(A(dS^2_{P_0})^n f)(k)(ik^0)^{-n}|_{E_0}\leqslant|k^0|^{c-n}q((dS^2_{P_0})^n f)$. It follows then, from the measurability of $V_{m_0,\epsilon_0}\ni k\to(A(dS^2_{P_0})^n f)(k)\in E_0$, that the second member on the rhs of (3.19) defines an element in $\mathscr{L}(F^2,H_0)$, for n sufficiently big.

Further the cobord equation (see Proposition 3.12), $S^1_g Af = R_g S^2_g f + AS^2_g f$, has a rhs which is a C^∞ function from P_0 into H_0. Hence $Af\in F^1$. Finally, by the same equation, it follows that the map $P_0\ni g\to S^1_g Af\in H_0$ converges to zero in $C^\infty(P_0,H_0)$ (with topology of convergence on compact subsets of P_0), when f goes to zero in F^2. Hence $A\in\mathscr{L}(F^2,F^1)$ (cf. Ref. 13, p. 253). Q.E.D.

Proposition 3.14: $H^1(P_0,\mathscr{L}_b(F^2,F^1))=\{0\}$.

Proof: We can, for every equivalence class in $H^1(P_0,\mathscr{L}_b(F^2,F^1))$, choose a representative R, being zero on the subgroup SU(2) (see the proof of Lemma 3.6). R is then a coboundary by Proposition 3.12 and Proposition 3.13.

4. PROOF OF THE LINEARIZATION THEOREM

By hypothesis, there is a positive integer N and there are continuous unitary irreducible representations $(U^1_i,\mathscr{H}_i)$, $1\leqslant i\leqslant N$ of P_0, such that $U^1=\oplus_{i=1}^N U^1_i$ and $\mathscr{H}=\oplus_{i=1}^N\mathscr{H}_i$. The 1-cohomology $H^1(P_0,\mathscr{L}_b(\hat{\otimes}^n\mathscr{H}_\infty,\mathscr{H}_\infty))$, $n\geqslant 2$, for the representation $P_0\times\mathscr{L}(\hat{\otimes}^n\mathscr{H}_\infty,\mathscr{H}_\infty)\ni(g,A)\to U^1_g A(\hat{\otimes}^n U^1_{g^{-1}})\in\mathscr{L}(\hat{\otimes}^n\mathscr{H}_\infty,\mathscr{H}_\infty)$, is introduced similarly to the cohomology $H^1(P_0,\mathscr{L}_b(F^2,F^1))$ in Paragraph 3.

Lemma 4.1: $H^1(P_0,\mathscr{L}_b(\hat{\otimes}^n\mathscr{H}_\infty,\mathscr{H}_\infty)) = \oplus_i H^1(P_0,\mathscr{L}_b((\mathscr{H}_{i_1})_\infty\hat{\otimes}\cdots\hat{\otimes}(\mathscr{H}_{i_n})_\infty,(\mathscr{H}_{i_0})_\infty))$, where the direct sum is taken over all $i=(i_0,i_1,\dots,i_n)\in\{1,\dots,N\}^{n+1}$.

Proof: For $1\leqslant j\leqslant N$, let $\pi_j:\mathscr{H}_\infty\to(\mathscr{H}_j)_\infty$, be the canonical projection $\pi_j(u_1\oplus\cdots\oplus u_N)=u_j$ and let R: $P_0\to\mathscr{L}_b(\hat{\otimes}^n\mathscr{H}_\infty,\mathscr{H}_\infty)$ be a map. Then $R_g=\Sigma_{i\in\{1,\dots,N\}^{n+1}}\pi_{i_0}R_g(\pi_{i_1}\otimes\cdots\otimes\pi_{i_n}) = \hat{\oplus}_{i\in\{1,\dots,N\}^{n+1}}(R_{i_0;i_1,\dots,i_n})$, where $(R_{i_0;i_1,\dots,i_n})_g\in\mathscr{L}((\mathscr{H}_{i_1})_\infty\hat{\otimes}\cdots\hat{\otimes}(\mathscr{H}_{i_n})_\infty,(\mathscr{H}_{i_0})_\infty)$. As $\pi_i U'_g = U^1_{ig}\pi_i$, it follows now that R is a cocycle (resp. coboundary) if and only if $R_{i_0;i_1,\dots,i_n}$ is a cocycle (resp. coboundary), for all $i\in\{1,\dots,N\}^{n+1}$. Q.E.D.

Proof of Theorem 2.4: There exists an invertible $A\in\mathscr{F}(F)$, such $A\circ U_g = U^1_g\circ A$, if $H^1(P_0,\mathscr{L}_b(\hat{\otimes}^n\mathscr{H}_\infty,\mathscr{H}_\infty))=\{0\}$, for every $n\geqslant 2$ (cf. Proposition 1 of Ref. 2). It follows from Lemma 4.1 and Proposition 3.14 that this cohomology is trivial. Q.E.D.

5. A REMARK ON THE NONLINEARIZABILITY OF CERTAIN MASSLESS REPRESENTATIONS OF $\mathbb{R}^2\otimes\mathfrak{so}(1,1)$

Let $D(r)$ be the open disk with radius $r>0$ and center 0 in $\mathbb{C}$ and let J: $D(r)\to\mathbb{C}$ be a holomorphic function, with

$J(0) = J'(0) = 0$. We consider the $(1+1)$-dimensional wave equation $(\partial^2/\partial t^2 - \partial^2/\partial x^2)\varphi = J(\varphi)$ in its evolution form:

$$\frac{d}{dt}\begin{pmatrix}\varphi(t)\\ \dot\varphi(t)\end{pmatrix} = \begin{pmatrix}0 & 1\\ \partial^2 & 0\end{pmatrix}\begin{pmatrix}\varphi(t)\\ \dot\varphi(t)\end{pmatrix} + \begin{pmatrix}0\\ J(\varphi(t))\end{pmatrix}$$
$$\equiv T_{P_0}\begin{pmatrix}\varphi(t)\\ \dot\varphi(t)\end{pmatrix} \quad \left(\partial = \frac{\partial}{\partial x}\right). \tag{5.1}$$

The space of initial conditions is chosen to be $E = L^2_\infty(\mathbb{R}) \oplus L^2_\infty(\mathbb{R})$, where $L^2_\infty(\mathbb{R})$ is the Fréchet space of complex valued C^∞ functions on $\mathbb{R}$, being in L^2 together with all its derivatives. T_{P_0} and $T_{P_1} = \partial \oplus \partial$ defines a nonlinear representation T of the Lie algebra $\mathbb{R}^2$. This representation can naturally be extended to a representation of $\mathbb{R}^2 \oplus \mathfrak{so}(1,1)$. However, we will not use the boosts. The transformation $a_+(k) = \hat{\dot\varphi}(k) - i|k|\hat\varphi(k)$, $a_-(k) = \hat{\dot\varphi}(k) + i|k|\hat\varphi(k)$, ($k \in \mathbb{R}$ and $\hat\varphi$ the Fourier transform of φ), for $(\varphi,\dot\varphi) \in E$, takes Eq. (5.1) into

$$\frac{d}{dt}\begin{pmatrix}a_+(t)\\ a_-(t)\end{pmatrix} = A_0\begin{pmatrix}a_+(t)\\ a_-(t)\end{pmatrix} + \begin{pmatrix}\widetilde J(a_+(t) - a_-(t))\\ \widetilde J(a_+(t) - a_-(t))\end{pmatrix}$$
$$\equiv \widetilde T_{P_0}\begin{pmatrix}a_+(t)\\ a_-(t)\end{pmatrix}. \tag{5.2}$$

Here $[A_0\binom{u_+}{u_-}](k) = \begin{pmatrix}-i|k| & 0\\ 0 & i|k|\end{pmatrix}\binom{u_+(k)}{u_-(k)}$ and $\widetilde J(a_+(t) - a_-(t)) = J(\varphi(t))\hat{}$. Introduce also $[\widetilde T_{P_1}\binom{u_+}{u_-}](k) = \binom{-iku_+(k)}{-iku_-(k)}$. $\widetilde T$ is the image of T under the above transformation. Denote by $\widetilde E$, the image of E and by $\widetilde E_1$ the LF-space $C_0^\infty(\mathbb{R} - \{0\}) \oplus C_0^\infty(\mathbb{R} - \{0\})$, where $C_0^\infty(\mathbb{R} - \{0\})$ is the space of C^∞ functions with compact support in $\mathbb{R} - \{0\}$.

Proposition 5.1: If $J \neq 0$, there is no formal linearization $A \in \mathscr{F}(\widetilde E_1, \widetilde E)$ of $\widetilde T$ (the linear term of A is the canonical injection).

Proof: Let $J(z) = \Sigma_{n>m} c_n z^n$, $c_m \neq 0$ for $|z| < r$. Denote by $\widetilde T^n$ the term of degree n in $\widetilde T$. A necessary condition for the existence of a formal linearization $A \in \mathscr{F}(\widetilde E_1, \widetilde E)$ is that the cohomological equations, $\widetilde T^1_{P_1} A^m - A^m(\Sigma_{q=0}^{m-1} I_q \otimes \widetilde T^1_{P_1} \otimes I_{m-q-1}) = 0$ and $\widetilde T^1_{P_0} A^m - A^m(\Sigma_{q=0}^{m} I_q \otimes \widetilde T^1_{P_0} \otimes I_{m-q-1}) = -\widetilde T^m_{P_0}$, where I_q is the identity on $\hat\otimes^q_s E_1$, have a solution $A^m \in \mathscr{L}(\hat\otimes^m_s \widetilde E_1, \widetilde E)$ (cf. Ref. 2). Introduce the projection operators $Q_+\binom{a_+}{a_-} = \binom{a_+}{0}$, $Q_-\binom{a_+}{a_-} = \binom{0}{a_-}$, $[Q^+\binom{a_+}{a_-}](k) = \binom{a_+}{a_-}(k)$ if $k \geqslant 0$, $[Q^+\binom{a_+}{a_-}](k) = 0$ if $k < 0$, $Q^- = I - Q^+$, and $Q^{\epsilon_1}_{\epsilon_2} = Q^{\epsilon_1} Q_{\epsilon_2}$ for $\epsilon_1, \epsilon_2 = \pm$. We note that $Q^{\epsilon_1}_{\epsilon_2}\, \widetilde T^1_{P_0} = \epsilon_1\epsilon_2\, \widetilde T^1_{P_1}$. The cohomological equations give the equality

$$Q^{\epsilon_1}_{\epsilon_2} A^m(\otimes^m Q^{\alpha_1}_{\alpha_2})\left(\sum_{q=0}^{m-1} I_q \otimes \widetilde T^1_{P_1} \otimes I_{m-q-1}\right)$$
$$(\epsilon_1\epsilon_2 - \alpha_1\alpha_2) = -Q^{\epsilon_1}_{\epsilon_2}\widetilde T^m_{P_0}(\otimes^m Q^{\alpha_1}_{\alpha_2}). \tag{5.3}$$

Choosing $\epsilon_1 = \alpha_1$, $\epsilon_2 = \alpha_2$, the nonvanishing component of the rhs of (5.3) is, evaluated at $\otimes^m\binom{a_+}{a_-}$: $c_m(i)^m(X^{-1}a^{\epsilon_1}_{\epsilon_2}) * (X^{-1}a^{\epsilon_1}_{\epsilon_2}) * \cdots * (X^{-1}a^{\epsilon_1}_{\epsilon_2})$ (m factors), where $*$ is the convolution and $(X^{-1}a^{\epsilon_1}_{\epsilon_2})(k) = k^{-1}a_{\epsilon_2}(k)$ for $\epsilon_1 k \geqslant 0$ and $(X^{-1}a^{\epsilon_1}_{\epsilon_2})(k) = 0$ for $\epsilon_1 k < 0$. Hence the rhs of (5.3) is different from zero, whereas the lhs vanishes. Q.E.D.

Remark 5.2: It follows from the proof that the nonexistence of A is insensible under change of the spaces $\widetilde E_1$ and $\widetilde E$. The present choice of $\widetilde E_1$ and $\widetilde E_2$ was made in order to state a precise result. The result is also valid for coupled equations $\Box\varphi_i = J_i(\varphi_1,\ldots,\varphi_n)$, $1 \leqslant i \leqslant n$, where J_i are holomorphic in a neighborhood of 0 in $\mathbb{C}^n$.

ACKNOWLEDGMENTS

I want to thank Professor M. Flato, Professor L. Hörmander, Professor G. Pinczon, and Professor J. Simon for useful discussions and helpful suggestions.

[1] M. Flato, G. Pinczon, and J. Simon, "Non-linear representations of Lie groups," Ann. Sci. Ec. Norm. Sup. **10**, 405–18 (1977).

[2] M. Flato and J. Simon, "Non-linear equations and covariance," Lett. Math. Phys. **3**, 279–83 (1979).

[3] M. Flato and J. Simon, "On a linearization program of nonlinear field equations," Phys. Lett. B **94**, 518–22 (1980).

[4] $\ltimes$ stands for semidirect product.

[5] J. Simon, "Non-linear representations of the Poincaré group and global solutions of relativistic wave equations," Pac. J.Math. (to be published).

[6] G. Pinczon and J. Simon, "Models of non-linear representations and examples of linearization techniques," J. Math. Phys. (to be published).

[7] J. Simon, "Non-linear representations of the Euclidean group of the plane," preprint, Université de Dijon, France.

[8] J. Simon, "Non-linear representations of the affine group of the complex plane," preprint, Université de Dijon, France.

[9] L. Hörmander, "On the division of distributions by polynomials," Ark. Mat. **3**, 555–68 (1958).

[10] B. Malgrange, "Division des distributions," Seminar Schwartz 4ème année, 1959/60, No. 23–24.

[11] F. Treves, *Topological Vectorspaces, Distributions and Kernels* (Academic, New York, 1967).

[12] $\oplus$ stands for semidirect sum.

[13] G. Warner, "Harmonic analysis on semi-simple Lie groups I," *Die Grundlehren der mathematischen Wissenschafter in Einzeldarstellungen* (Springer, Berlin, 1972), Band 188.

[14] G. Pinczon, lecture notes, Dijon, 1983.

[15] L. Schwartz, *Théorie des distributions* (Hermann, Paris, 1966; 2nd ed. 1973).

[16] R. C. Gunning and H. Rossi, *Analytic Functions of Several Complex Variables* (Prentice-Hall, Englewood Cliffs, NJ, 1965).

[17] G. Pinczon and J. Simon, "Extension of representations and cohomology," Rep. Math. Phys. **16**, 49–77 (1979).

[18] N. N. Bogolubov, A. A. Logunov, and I. T. Todorov, *Introduction to axiomatic quantum field theory*, Mathematical Physics Monograph Series 18 (Benjamin, New York, 1975).

On Lie algebras built from SO(3) tensor operators

G. Vanden Berghe and H. De Meyer[a)]
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B-9000 Gent, Belgium

(Received 16 May 1983; accepted for publication 17 June 1983)

Concerning the Lie algebras generated by certain linear combinations of SO(3) tensor operators two theorems are given which disprove some widely accepted results in the standard literature.

PACS numbers: 02.20.Sv

In his description of collective motion in the nuclear shell model, Elliott[1] introduced SO(3) tensor operators $v_q^k(ll')$ $(q=-k,-k+1,\ldots,k)$ of rank k mapping the $(2l'+1)$-dimensional space with angular momentum basis $\{|l',m'\rangle : m'=-l',-l'+1,\ldots,l'\}$ into the $(2l+1)$-dimensional space with basis $\{|l,m\rangle : m=-l,-l+1,\ldots,l\}$. From the work of Racah[2] such tensor operators are known to be completely defined by means of their reduced matrix elements:

$$\langle l_2\|v^k(ll')\|l_1\rangle=(2k+1)^{1/2}\delta_{ll_2}\delta_{l'l_1}\,. \tag{1}$$

Herein l and l' assume both positive integer values or both half-odd integer values.

It has been explicitly shown by Judd[3] that the set of tensor operators $\{v_q^k(ll)|k$ odd and $0<k<2l\}$ spans the Lie algebra of the rotational groups $SO(2l+1)$ and this for all $l\in\mathbb{N}_0$. On the other hand, it is known from Elliott's model or more recently from the I. B. A. model that the set $\{v_q^1(dd)$, $v_{q'}^3(dd)$, $v_{q''}^2(sd)+v_{q''}^2(ds)|s=0,d=2\}$ is an SO(6) operator basis, from which it has been taken for granted[1,3] that similarly $\{v_q^k(ll)$, $v_{q'}^{k'}(l'l')$, $v_{q''}^{k''}(ll')+v_{q''}^{k''}(l'l)|k$ and k' odd$\}$ constitutes a basis for the Lie algebra of $SO(2l+2l'+2)$. The incorrectness of this statement follows from:

Theorem 1: The set of tensor operators $\mathscr{L}=\{v_q^k(ll)$, $v_{q'}^{k'}(l'l')$, $v_{q''}^{k''}(ll')+\epsilon(-1)^{k''}v_{q''}^{k''}(l'l)|k,k',k''\in\mathbb{N}_0$, k and k' odd, $0<k\leqslant 2l$, $0<k'\leqslant 2l'$, $|l-l'|\leqslant k''\leqslant l+l'$, $\epsilon\in\{-1,+1\}\}$ forms a Lie algebra basis. The proof is obtained by direct verification, taking into account that (Judd[3])

$$\begin{aligned}
[v_{q_1}^{k_1}(l_al_b),v_{q_2}^{k_2}(l_cl_d)]={}&\sum_{k_3,q_3}[(2k_1+1)(2k_2+1)(2k_3+1)]^{1/2}\\
&\times\begin{pmatrix}k_1 & k_2 & k_3\\ q_1 & q_2 & -q_3\end{pmatrix}(-1)^{2l_d+l_c-l_b-q_3}\Bigg[\delta_{l_bl_c}(-1)^{k_1+k_2+k_3+l_a+l_b+l_c+l_d}\\
&\times\begin{Bmatrix}k_1 & k_2 & k_3\\ l_d & l_a & l_b\end{Bmatrix}v_{q_3}^{k_3}(l_al_d)-\delta_{l_al_d}\begin{Bmatrix}k_1 & k_2 & k_3\\ l_c & l_b & l_a\end{Bmatrix}v_{q_3}^{k_3}(l_cl_b)\Bigg]\,.
\end{aligned}\tag{2}$$

The important fact is the presence of a phase factor $(-1)^{k''}$ in front of one of the mixed type tensors. Since in the (s,d) case there is only one type of such tensors the need for an alternating phase factor did not become apparent there. Presumably, the identification with an $SO(2l+2l'+2)$ algebra in the general case has been erroneously made by simply counting the number of independent tensor operator combinations. Clearly, when $l=2$ and $l'=0$ it is the $\epsilon=+1$ choice which leads to SO(6). What kind of algebra is generated in general by $\mathscr{L}$ for $\epsilon=+1$ or $\epsilon=-1$ is answered in:

Theorem 2: The set of tensor operators $\mathscr{L}$ spans the compact $SO(2l+2l'+2)$ algebra when $\epsilon=+1$ and the noncompact $SO(2l+1,2l'+1)$ algebra when $\epsilon=-1$.

We first define operators $E_{mm'}^{ll'}$,

$$E_{mm'}^{ll'}=\sum_{k,q}(-1)^{l-m}(2k+1)^{1/2}\begin{pmatrix}l & k & l'\\ -m & q & m'\end{pmatrix}v_q^k(ll')\,, \tag{3}$$

which on account of (2) themselves satisfy the commutation relations

$$\left[E_{m_1m_1'}^{l_1l_1'},E_{m_2m_2'}^{l_2l_2'}\right]=\delta_{l_1'l_2}\delta_{m_1'm_2}E_{m_1m_2'}^{l_1l_2'}-\delta_{l_1l_2'}\delta_{m_1m_2'}E_{m_2m_1'}^{l_2l_1'}\,. \tag{4}$$

As a next step we form the combinations

$$\begin{aligned}
W_{\nu\mu}={}&E_{\nu\mu}^{ll}-(-1)^{\mu+\nu}E_{-\mu-\nu}^{ll}+\rho\,\{\delta_{\mu0}\,[E_{\nu0}^{ll'}+\epsilon(-1)^{\mu+\nu}E_{0-\nu}^{l'l}]\\
&+\delta_{\nu0}\,[\epsilon E_{0\mu}^{l'l}+(-1)^{\mu+\nu}E_{-\mu0}^{ll'}]-\delta_{\mu0}\delta_{\nu0}\,[E_{00}^{ll'}+\epsilon E_{00}^{l'l}]\}\\
&(-l\leqslant\mu,\nu\leqslant l)\,,\\
\overline{W}_{\nu\mu}={}&E_{\nu\mu}^{l'l'}-(-1)^{\mu+\nu}E_{-\mu-\nu}^{l'l'}+\rho\,\{\delta_{\mu0}\,[\epsilon E_{\nu0}^{l'l}+(-1)^{\mu+\nu}E_{0-\nu}^{ll'}]\\
&+\delta_{\nu0}\,[E_{0\mu}^{ll'}+\epsilon(-1)^{\mu+\nu}E_{-\mu0}^{l'l}]-\delta_{\mu0}\delta_{\nu0}\,[E_{00}^{ll'}+\epsilon E_{00}^{l'l}]\}\\
&(-l'\leqslant\mu,\nu\leqslant l')\,,\\
\overline{\overline{W}}_{\nu\mu}={}&\rho(1-\delta_{\mu0})(1-\delta_{\nu0})\,[E_{\nu\mu}^{ll'}+\epsilon(-1)^{\mu+\nu}E_{-\mu-\nu}^{l'l}]\\
&(-l'\leqslant\mu\leqslant l',-l\leqslant\nu\leqslant l)\,,
\end{aligned}\tag{5}$$

[a)] Research Associate N.F.W.O. (Belgium).

where ρ is an arbitrary complex constant of unit modulus to be determined later. Notice that for $\rho \in \mathbb{R}_0$ ($\rho \in \mathbb{C}_0$) the set (5) is the image of $\mathscr{L}$ under a real (complex) nonsingular linear mapping. Also $W_{00} = \overline{W}_{00}$, and if l or l' equals zero, (5) reduces to a single line. The operator set (5) is very well suited to construct the Cartan–Weyl basis of the algebra spanned by $\mathscr{L}$. Indeed, $\{W_{ii}(i = 1,...,l), \overline{W}_{jj}(j = 1,...,l'), W_{00} = \overline{W}_{00}\}$ is a set of $l + l' + 1$ intercommuting operators which may serve as Cartan's H-generators. The remaining operators in (5) can play the role of the E-generators since on account of (4)

$$\begin{aligned}
&[W_{ii},W_{\nu\mu}] = (\delta_{\nu i} - \delta_{\mu i} - \delta_{-\nu i} + \delta_{-\mu i})\, W_{\nu\mu}\,,\\
&[\overline{W}_{ii},\overline{W}_{\nu\mu}] = (\delta_{\nu i} - \delta_{\mu i} - \delta_{-\nu i} + \delta_{-\mu i})\, \overline{W}_{\nu\mu}\,,\\
&[W_{ii},\overline{W}_{\nu\mu}] = [\overline{W}_{ii},W_{\nu\mu}] = 0\,,\\
&[W_{ii},\overline{\overline{W}}_{\nu\mu}] = (\delta_{\nu i} - \delta_{-\nu i})\, \overline{\overline{W}}_{\nu\mu}\,,\\
&[\overline{W}_{ii},\overline{\overline{W}}_{\nu\mu}] = (-\delta_{\mu i} + \delta_{-\mu i})\, \overline{\overline{W}}_{\nu\mu}\,,\\
&[W_{00},\overline{\overline{W}}_{\nu\mu}] = 0\,,
\end{aligned} \tag{6}$$

and if $\epsilon\rho^2 = +1$,

$$\begin{aligned}
&[W_{00},W_{\nu\mu}] = (-\delta_{\mu 0} + \delta_{\nu 0})\, W_{\nu\mu}\,,\\
&[W_{00},\overline{W}_{\nu\mu}] = (-\delta_{\mu 0} + \delta_{\nu 0})\, \overline{W}_{\nu\mu}\,,
\end{aligned} \tag{7a}$$

or if $\epsilon\rho^2 = 1$,

$$\begin{aligned}
&[W_{00},W_{\nu\mu}] = (-\delta_{\mu 0} + \delta_{\nu 0})(-1)^{\mu+\nu}\, W_{-\mu-\nu}\,,\\
&[W_{00},\overline{W}_{\nu\mu}] = (-\delta_{\mu 0} + \delta_{\nu 0})(-1)^{\mu+\nu}\, \overline{W}_{-\mu-\nu}\,.
\end{aligned} \tag{7b}$$

From these commutators we can read off the root structure of the algebra spanned by $\mathscr{L}$ only if $\epsilon\rho^2 = +1$, in which case the roots are of the form $\pm \mathbf{e}_j \pm \mathbf{e}_j$ whereby $\{\mathbf{e}_i | i = 1,...,l + l' + 1\}$ is the set of canonical orthogonal unit vectors in an $(l + l' + 1)$-dimensional vector space. The root structure is that of the SO$(2l + 2l' + 2)$ Lie algebra. Hence if $\epsilon = +1$ we can choose $\rho = +1$ (the choice $\rho = -1$ only replaces indices μ by $-\nu$ and vice versa) proving that $\mathscr{L}$ consists of the compact generators of SO$(2l + 2l' + 2)$. If $\epsilon = -1$, we can choose $\rho = i$ (or equivalently $-i$), showing that in order to preserve the root structure all $(2l + 1)\times(2l' + 1)$ mixed type tensors $v^{k''}_{q''}(ll') - (-1)^{k''} v^{k''}_{q''}(l'l)$ should be provided with an i-factor in front. Hence the algebra spanned by $\mathscr{L}$ is the noncompact SO$(2l + 1, 2l' + 1)$ Lie algebra which has $(2l + 1)\times(2l' + 1)$ noncompact generators and which contains SO$(2l + 1) \otimes$ SO$(2l' + 1)$ as maximal compact subalgebra.[4] The latter subalgebra is clearly generated by $\{v^k_q(ll), v^{k'}_q(l'l')\}$. Finally, if $\epsilon < 0$ we see from (7b) that with $\rho = \pm 1$ we do not obtain the Cartan–Weyl form and that it is impossible to establish the Cartan basis by taking real combinations of the $W_{\nu\mu}$, $\overline{W}_{\nu\mu}$, and $\overline{\overline{W}}_{\nu\mu}$. This completes the proof of Theorem 2.

[1]J. P. Elliott, Proc. R. Soc. London Ser. A **245**, 128 (1958).

[2]G. Racah, Phys. Rev. **62**, 438 (1942).

[3]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).

[4]R. Gilmore, *Lie groups, Lie algebras and some of their Applications* (Wiley, New York, 1974).

The positivity of the Lyapunov exponent and the absence of the absolutely continuous spectrum for the almost-Mathieu equation

A. L. Figotin and L. A. Pastur
Physico-Technical Institute of Low Temperatures, UkrSSR Academy of Sciences, Kharkov, USSR

(Received 4 February 1983; accepted for publication 25 February 1983)

This paper contains the rigorous proof of the following statement formulated by Andre and Aubry: the Cauchy solutions of the discrete Schrödinger equation with the potential $q_n = g\cos(2\pi n\theta + \varphi)$ grow exponentially for every irrational θ, $g > 1$ and almost every $\varphi \epsilon [0,2\pi)$. According to known results this fact implies the absence of the absolutely continuous component of the spectrum for the corresponding operator.

PACS numbers: 02.30.Hq

I. INTRODUCTION

Reference 1 contained elegant heuristic arguments which show that the Lyapunov exponent of the almost-Mathieu equation

$$\tfrac{1}{2}(\psi_{n+1} + \psi_{n-1}) + g\cos(2\pi n\theta + \varphi)\psi_n = \lambda\psi_n \tag{1}$$

is positive for $g > 1$ and any irrational θ. According to Refs. 2 and 3, from the positivity of the Lyapunov exponent follows the absence of the absolutely continuous component in the spectrum of the operator $H(\theta,g,\varphi)$ defined by Eq. (1) in $l_2(\mathbb{Z})$. Till now such results could be only obtained by imposing the conditions not on the amplitude of the potential q_n [in Eq. (1), $q_n = g\cos(2\pi n\theta + \varphi)$], but on its statistical properties. Thus, Ref. 4 proved the positivity of the Lyapunov exponent for q_n which are independent and identically distributed random variables, and Ref. 3 did it for the sequence q_n which is an ergodic Markov chain. Therefore the statement formulated in Ref. 1 for an ergodic sequence

$$q_n = g\cos(2\pi n\theta + \varphi) \tag{2}$$

not having any mixing properties seems rather unexpected and interesting.

This paper contains the rigorous proof of the statement formulated in Ref. 1. Avron and Simon[5] also announced a rigorous proof. However, as we can judge by their short note, our proof is different from theirs, though it is also based on the principal idea suggested in Ref. 1.

Note that if the number θ is well enough approximated by rational numbers, so that if for a certain sequence $\{p_k, q_k\}$ of integers,

$$|\theta - p_k/q_k| \leqslant Cq^{-k},$$

then, as was shown in Ref. 5 on the basis of the results of Ref. 6, the operator $H(\theta,g,\varphi)$ has no eigenfunctions. Combined with what was said above, this means that for $g > 1$ the spectrum of $H(\theta,g,\varphi)$ for such θ and almost all $\varphi \epsilon [0,2\pi)$ is singularly continuous. However, the set of such θ's has zero Lebesgue measure.

On the other hand, as is shown in Ref. 7, for the set of irrational θ's of full Lebesgue measure, the spectrum of $H(\theta,g,\varphi)$ for sufficiently small g has a nonempty absolutely continuous component. Hence, by virtue of the duality of Eq. (1) it follows that for sufficiently large g, H has a point component of a positive Lebesgue measure.

2. MAIN RESULTS

Thus, we are to prove the following two theorems.

Theorem 1: Let u_n be the solution of Eq. (1) satisfying the condition

$$u_{-1} = \cos\alpha, \quad u_0 = \sin\alpha, \quad \alpha \epsilon [0,\pi]. \tag{3}$$

Then the limit

$$\lim_{n\to\infty} n^{-1}\ln(u_n^2 + u_{n-1}^2)^{1/2} = \gamma(\lambda;g) \tag{4}$$

does exist and is nonnegative for almost every $\varphi \epsilon [0,2\pi)$, for almost all λ's. This limit, for any irrational θ and $g > 1$, is strictly positive.

Theorem 2: Let $H(\theta,g,\varphi)$ be the operator in $l_2(\mathbb{Z})$ corresponding to Eq. (1). Then, for any irrational θ and $g > 1$, its spectrum for almost every $\varphi \epsilon [0,2\pi)$ has no absolutely continuous component.

Theorem 2 is a corollary of Theorem 1. Indeed, in case of fixed irrational θ and real g, the operator $H(\theta,g,\varphi)$, regarded as a function of the parameter $\varphi \epsilon [0,2\pi)$ is metrically transitive.[3] Namely, if the role of the probability space Ω is played by a circle of a unit radius with the probability measure $(2\pi)^{-1} d\varphi$ on it, where $d\varphi$ is the Lebesgue measure, and the group of shifts of Ω coincides with the group of transformations $\varphi \to \varphi + 2\pi n\theta$, $n \epsilon \mathbb{Z}$, then the sequence (2) will be, as is readily seen, a metrically transitive one,[8] and H accordingly a metrically transitive operator.[3] But as was shown in Ref. 3, positivity on a set of full measure in Ω of the Lyapunov exponent of equation

$$\tfrac{1}{2}(\psi_{n+1} + \psi_{n-1}) + q_n(\omega)\psi_n = \lambda\psi_n, \tag{5}$$

where $q_n(\omega)$, $\omega \epsilon \Omega$ is a metrically transitive sequence, is a sufficient condition of the absence of the absolutely continuous spectrum of almost every operator $H(\omega)$ defined by Eq. (5).

Thus, the main purpose of this work is to prove Theorem 1. But due to the multiplicative ergodic theorem,[12] the limit (4) exists for any metrically transitive q_n in Eq. (5). Therefore, we are only to prove the positivity of the limit (4). According to the central idea of Ref. 1, the latter fact may be proved in the following way. Let $N(\lambda;g)$ be a normalized eigenvalue distribution function (integrated density of states) of the operator $H(\theta,g,\varphi)$ (see its definition in Ref. 3 and below). The existence of such a function for a wide class of metrically transitive operators in $l_2(\mathbb{Z})$ including operators corresponding to Eq. (5) with an arbitrary metrically transitive q_n was proved in Ref. 9. Then, as was argued in Ref. 1, for an irrational θ, the following equation is true (Aubry duality):

$$N(\lambda;g) = N(\lambda g^{-1},g^{-1}). \tag{6}$$

Then, according to Ref. 10, for second-order operators of the general form of Eq. (5), it should be true that (Thouless formula)

$$\gamma(\lambda) = \int_{-\infty}^{\infty} \ln|\lambda - \lambda'|dN(\lambda'). \tag{7}$$

From equalities (6) and (7) it immediately follows that

$$\gamma(\lambda;g) = \ln q + \gamma(\lambda g^{-1},g^{-1}). \tag{8}$$

Since, as has been mentioned, for any λ and g, $\gamma(\lambda,g)$ is nonnegative, then it follows from the latter equality with $g > 1$ that $\gamma(\lambda,g) > 0$.

Thus we see that according to Ref. 1 the proof of Theorem 1 is reduced to the proof of the Aubry duality (6) and demonstration of the validity of the Thouless formula (7). The proof of these facts, formulated rigorously in Theorems 3 and 4, is carried out in the two remaining sections.

3. AUBRY DUALITY

Theorem 3: Let θ in Eq. (1) be an irrational number. Then for any real λ and $g \neq 0$, equality (6) is true.

We shall give two proofs of the theorem. They both use the representation of the function $N(\lambda)$ in terms of the resolution of identity of the corresponding metrically transitive operator (see Ref. 9), which in the case of the operator $H(\theta,g,\varphi)$ of Eq. (1) is

$$N(\lambda;g) = \frac{1}{2\pi}\int_0^{2\pi} d\varphi\, E_{00}(\lambda;g,\varphi). \tag{9}$$

Here $E(\lambda;g,\varphi)$ is the resolution of identity of the operator $H(\theta,g,\varphi)$, $E_{00} = (e_0, Ee_0)$ and e_0 is the vector from $l_2(\mathbb{Z})$, defined by equations $(e_0)_0 = 1$, $(e_0)_k = 0$ if $k \neq 0$.

Note that without loss of generality, g may be always assumed as positive.

Proof 1: First give the definition of the function $N(\lambda)$, i.e., the integrated density of states. If the operator H, defined by Eq. (5), is written in the matrix form $H = \{h_{km}\}$, k, $m \in \mathbb{Z}$ then we may consider, along with it, a sequence of "truncated" matrices $H_n = \{h_{km}\}$, $|k|$, $|m| \leqslant n$. For every matrix H_n define a normalized eigenvalue distribution function $N_n(\lambda)$ so that $(2n+1)N_n(\lambda)$ is the number of H_n-matrix eigenvalues less than λ. As was shown in Ref. 9, there exists a nonrandom continuous and nondecreasing function $N(\lambda)$ such that with probability 1

$$\lim_{n\to\infty} N_n(\lambda) = N(\lambda).$$

This function $N(\lambda)$ will be termed the normalized eigenvalue distribution function or the integrated density of states of the metrically transitive operator of the form of Eq. (2).

Approximate the irrational number θ by a sequence of rational numbers p, q, where p and q are relatively prime integers, i.e., $\theta = \lim_{q\to\infty} p/q$. Consider the operators H_q with periodic potentials, specified by Eq. (1), with the irrational number θ replaced by its rational approximation p/q. Introduce normalized eigenvalue distribution functions $N_q(\lambda;g,\varphi)$ of the operators H_q.

Lemma 1: The equality

$$N(\lambda;g) = \lim_{q\to\infty} \frac{q}{2\pi}\int_{-\pi/q}^{\pi/q} d\varphi\, N_q(\lambda;g,\varphi)$$

is true.

Note that for the operators H_q with a periodic potential, the following analog of Eq. (9) holds. For any fixed φ,

$$N_q(\lambda;g,\varphi) = q^{-1}\sum_{r=1}^{q} (e_0, E_q(\lambda;g,\varphi + 2\pi r/q)e_0),$$

where $E_q(\lambda;g,\varphi)$ is resolution of identity of the operator H_q. Hence, it immediately follows that

$$\frac{q}{2\pi}\int_{-\pi/q}^{\pi/q} d\varphi\, N_q(\lambda;g,\varphi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\varphi\, (e_0, E_q(\lambda;g,\varphi)e_0). \tag{10}$$

Since, when $q\to\infty$, the sequence of operators H_q strongly converges to H: $H_q \xrightarrow{s} H$, then $E_q(\lambda;g,\varphi) \xrightarrow{s} E(\lambda;g,\varphi)$ at the continuity points of the last projector.[11] Therefore,

$$\lim_{q\to\infty} (e_0 E_q(\lambda;g,\varphi)e_0) = (e_0, E(\lambda;g,\varphi)e_0). \tag{11}$$

From Eqs. (9)–(11), the validity of the lemma immediately follows.

Lemma 2: For any $g \neq 0$,

$$\frac{q}{2\pi}\int_{-\pi/q}^{\pi/q} d\varphi\, N_q(\lambda;g,\varphi) = \frac{q}{2\pi}\int_{-\pi/q}^{\pi/q} d\varphi\, N_q(\lambda g^{-1},g^{-1}).$$

Since the operator H_q has a periodic potential, the Floquet theorem is applicable, according to which the H_q generalized eigenfunctions are

$$\psi_n = e^{i\kappa n}v_n, \quad v_{n+q} = v_n, \quad n \in \mathbb{Z}, \quad |\kappa| \leqslant \pi/q.$$

Having rewritten the equation for ψ_n in terms of v_n, we see that v_n is a solution of equation

$$\tfrac{1}{2}(e^{i\kappa}v_{n+1} + e^{-i\kappa}v_{n-1}) + g\cos(2\pi(p/q)n + \varphi)v_n = \lambda v_n. \tag{12}$$

The solvability condition for this equation gives the eigenvalues $\lambda_\alpha(\kappa,g,\varphi)$, $\alpha = 1,2,\dots,q$. Using the finite Fourier transform for v_n,

$$v_n = q^{-1/2}\sum_{l=1}^{q} w_l \exp\left(\frac{2\pi i l n}{q}\right), \quad w_{l+q} = w_l,$$

we obtain from Eq. (12) the following equation for w_l,

$$\tfrac{1}{2}(e^{-i\varphi}w_{l+p} + e^{i\varphi}w_{l-p}) + g^{-1}\cos(2\pi(l/q) + \kappa)w_l = g^{-1}\lambda_\alpha w_l.$$

Taking into account the periodicity of w_l and setting $l = rp$, $r = 1,2,\dots,q$, substitute $w_{rp}\to w_r$ (a different numbering) in the latter equation. Then

$$\tfrac{1}{2}(e^{-i\varphi}w_{r+1} + e^{i\varphi}w_{r-1}) + g^{-1}\cos(2\pi(p/q)r + \kappa)w_r = g^{-1}\lambda_\alpha w_r.$$

Now, comparing this equation with Eq. (12), we arrive at the equality

$$\lambda_\alpha(-\varphi,g^{-1},\kappa) = g^{-1}\lambda_\alpha(\kappa,g,\varphi). \tag{13}$$

Introducing the function $\chi_\lambda(t)$ equal to 1 for $\lambda > t$ and 0 for $\lambda \leqslant t$, we obtain

$$N_q(\lambda;g,\varphi) = \frac{1}{2\pi}\int_{-\pi/q}^{\pi/q} d\kappa \sum_{\alpha=1}^{n} \chi_\lambda(\lambda_\alpha(\kappa,g,\varphi)).$$

Hence, remembering Eq. (13), we obtain

$$\frac{q}{2\pi}\int_{-\pi/q}^{\pi/q} d\varphi\, N_q(\lambda;g,\varphi)$$
$$= \frac{q}{(2\pi)^2}\int_{-\pi/q}^{\pi/q} d\varphi\, d\kappa \sum_{\alpha=1}^{q} \chi_\lambda(g,\lambda_\alpha(-\varphi,q^{-1},\kappa))$$
$$= \frac{q}{(2\pi)^2}\int_{-\pi/q}^{\pi/q} d\varphi\, d\kappa \sum_{\alpha=1}^{q} \chi_{\lambda g^{-1}}(\lambda_\alpha(\kappa,g^{-1},\varphi))$$
$$= \frac{q}{2\pi}\int_{-\pi/q}^{\pi/q} N_q(\lambda g^{-1};g^{-1},\varphi)\,d\varphi.$$

The latter equality concludes the proof of the lemma.

Theorem 2 follows evidently from Lemmas 1 and 2.

Proof 2: Note that the duality relation (6) may be treated as equality of positive measures with finite supports. Therefore it is equivalent to equality of the moments of the measures. Therefore, if we denote

$$P_r(g) = \int_{-\infty}^{\infty} \lambda^r\, dN(\lambda;g), \quad r = 0,1,...,$$

then Eq. (6) will be equivalent to the set of equalities

$$P_r(g) = g^r P_r(g^{-1}), \quad r = 0,1,... . \tag{14}$$

If we make use of Eq. (9), then for $P_r(g)$ one may write the following representation:

$$P_r(g) = \frac{1}{2\pi}\int_0^{2\pi} d\varphi\, [H^r(\theta,g,\varphi)]_{00}, \tag{15}$$

where $[H^r]_{00}$ is the respective matrix element of the rth degree of the operator $H(\theta,g,\varphi)$, specified by Eq. (1). In this case, evidently,

$$[H^r]_{00} = \sum_{n_1,...,n_{r-1}} h_{on_1} h_{n_1 n_2} \cdots h_{n_{r-1}0}, \tag{16}$$

where h_{mn} are the matrix elements of $H(\theta,g,\varphi)$. Since h_{mn} is nonzero only when $|m-n| \leqslant 1$, then the sum in Eq. (16) may be represented as follows. Denote by $\mathscr{E}_r$ the set of sequences $\epsilon = (\epsilon_1,...,\epsilon_r)$, where $\epsilon_j = \pm 1,0$ and $\Sigma_{j=1}^r \epsilon_j = 0$, and by $E_k = \Sigma_{j=1}^k \epsilon_j$. Then

$$[H^r]_{00} = \sum_{\epsilon\in\mathscr{E}_r} h_{0E_1} h_{E_1E_2} \cdots h_{E_{r-1}E_r}. \tag{17}$$

Note that the right-hand side of Eq. (17) is a polynomial in g and the degree of every term in it is evidently equal to the number of zeroes in the set $\epsilon = (\epsilon_1,...,\epsilon_r)$. Moreover, if we denote by $\mathscr{E}_{rl}$ the subset of the set $\mathscr{E}_r$, consisting of the sequences $\epsilon\in\mathscr{E}_r$ which contain exactly l zeroes, and by $k_1,...,k_l$ the indices j for which $\epsilon_j = 0$ in given sequence ϵ, then the coefficient for the degree g^l in the sum (17) is

$$2^{l-r}\sum_{\epsilon\in\mathscr{E}_{rl}} \cos(2\pi E_{k_1}\theta + \varphi)\cdots\cos(2\pi E_{k_l}\theta + \varphi). \tag{18}$$

In order to find the coefficient of the degree g^l of the polynomial $P_r(g)$ of Eq. (15), Eq. (18) should be integrated with respect to φ, i.e.,

$$P_r(g) = \sum_{l=0}^{r} p_{rl} g^l,$$

$$p_{rl} = 2^{-r}\sum_{\substack{t\in\mathscr{T}_l^r\\ \epsilon\in\mathscr{E}_{rl}}} \exp\Big\{2\pi i\theta \sum_{m=1}^{l} \tau_m E_{k_m}\Big\}, \tag{19}$$

where $\mathscr{T}_l^r$ is the set of sequences $\tau = (\tau_1,...,\tau_l)$ with $\tau_m = \pm 1$ and $\Sigma_{m=1}^l \tau_m = 0$. Represent the sum under $\exp\{\cdots\}$ in equality (19) in a form which contains ϵ and τ in a more symmetrical way. To do that, introduce the set $\mathscr{T}_r$ consisting of sequences $\tau = (\tau_1,...,\tau_r)$, where $\tau_j = \pm 1,0$ and $\Sigma_{j=1}^r \tau_j = 0$. Let us now consider the set $(\mathscr{E}\mathscr{T})_{rl}$ of the pairs (ϵ,τ) of the sequences $\epsilon,\tau\in\mathscr{E}_r$ such that if $\epsilon_j = \pm 1$, then $\tau_j = 0$, and if $\tau_j = \pm 1$, then $\epsilon_j = 0$, and besides $\epsilon\in\mathscr{E}_{rl}$; that is, the set ϵ contains l zeroes exactly. Then Eq. (19) may be rewritten as

$$p_{r,l} = 2^{-r}\sum_{(\epsilon,\tau)\in(\mathscr{E}\mathscr{T})_{rl}} \exp\Big\{2\pi i\theta \sum_{j=1}^{r} \tau_j E_j\Big\}. \tag{20}$$

If we introduce $T_k = \Sigma_{j=1}^k \tau_j$ and remember that, by the definition of the set $\tau = (\tau_1,...,\tau_r)$, $\Sigma_{j=1}^r \tau_j = 0$, and also that the inclusion $(\epsilon_1\tau)\in(\mathscr{E}\mathscr{T})_{rl}$ is equivalent to $(\epsilon,-\tau)\in(\mathscr{E}\mathscr{T})_{rl}$, then Eq. (20) may be rewritten as

$$p_{r,l} = 2^{-r}\sum_{(\epsilon,\tau)\in(\mathscr{E}\mathscr{T})_{rl}} \exp\Big\{2\pi i\theta \sum_{j=1}^{r} \epsilon_j T_j\Big\}. \tag{21}$$

Since the condition that ϵ contains exactly l zeroes for $(\epsilon,\tau)\in(\mathscr{E}\mathscr{T})_{rl}$ is equivalent to τ contains $r-l$ zeroes, then we have from equalities (20) and (21),

$$p_{r,l} = p_{r,r-l}.$$

Hence follows evidently the validity of Eq. (14) and accordingly of Theorem 2.

4. THOULESS FORMULA

Theorem 4: Let H be operators as in Eq. (5) and for a certain nonrandom constant C: $|q_n| \leqslant C$. Equation (7) is then true for almost all λ values with respect to the Lebesgue measure.

Let us consider the matrices $\widetilde{H}_n$ obtained by "truncation" of the basic matrix H, somewhat different from that used for H_n: $H_n = \{\tilde{h}_{km}\}$, $-1 \leqslant k,m \leqslant n$, $\tilde{h}_{-1,-1} = \sin\alpha$, $\tilde{h}_{-1,0} = -\cos\alpha$, $\tilde{h}_{-1,m} = 0$, $0 \leqslant m \leqslant n$, $\tilde{h}_{km} = h_{km}$, $0 \leqslant k,m \leqslant n$. Let $n\widetilde{N}_n(\lambda)$ be the number of the eigenvalues of H_n less than λ.

Lemma 3: For every $\alpha\in[0,\pi]$ *with probability* 1 *at all continuity points* of $N(\lambda)$,

$$\lim_{n\to\infty} \widetilde{N}_n(\lambda) = \lim_{n\to\infty} N_n(\lambda) = N(\lambda). \tag{22}$$

The proof of this lemma may be obtained by considering the moments of the measures $d\widetilde{N}_n(\lambda)$ and $dN(\lambda)$ in the same way as it was done in Ref. 9.

Denote

$$\Delta_n(\lambda) = \det(\widetilde{H}_n - \lambda I_n),$$

where I_n is the identity matrix of the nth order. The sequence $\Delta_n(\lambda)$ is clearly the solution of Eq. (5) with the initial conditions (3). Thus, according to the equality (4) with probability 1 for every fixed λ, the limit

$$\gamma(\lambda,\alpha,\omega) = \lim_{n\to\infty} (2n)^{-1}\ln(\Delta_n^2(\lambda) + \Delta_{n-1}^2(\lambda)) \quad (23)$$

exists if q_n is a metrically transitive sequence and as a function of α takes two values $\pm\gamma(\lambda)$, where

$$\gamma(\lambda) = \lim_{n\to\infty} \sup_{\alpha} (2n)^{-1}\ln(\Delta_n^2 + \Delta_{n-1}^2) \geqslant 0$$

is the Lyapunov exponent of Eq. (5) (see Ref. 7).

Lemma 4: (a) $|n^{-1}\ln(\Delta_n^2 + \Delta_{n-1}^2)| \leqslant C(|\lambda| + 1)$, where C is a nonrandom positive constant; (b) if a and b are positive constants and Λ is any Borel set of the interval $[-a,a]$, then the family of functions of μ, $f_\Lambda(\mu) = \int_\Lambda \ln|\lambda - \mu|\, d\lambda$, considered in the interval $[-b,b]$ is equicontinuous and uniformly bounded; (c) for any Borel set Λ, the equality

$$\int_\Lambda \gamma(\lambda)d\lambda = \int_\Lambda d\lambda \int_{-\infty}^{\infty} \ln|\lambda - \lambda'|dN(\lambda')$$

is true for all the realizations of the stochastic process for which Eq. (22) holds.

The proof of (a) and (b) is quite elementary. Go on to the proof of (c). Note that from the definition of $\Delta_n(\lambda)$ and $\widetilde{N}(\lambda)$ it follows that

$$n^{-1}\ln|\Delta_n(\lambda)| = \int_{-\infty}^{\infty} \ln|\lambda - \lambda'|d\widetilde{N}_n(\lambda'). \quad (24)$$

Let Λ be an arbitrary bounded Borel set from $\mathbb{R}$. Consider decomposition $\Lambda = \Lambda_n \cap \Lambda_n'$, where $\Lambda_n = \{\lambda : \lambda\in\Lambda, |\Delta_n(\lambda)| \geqslant |\Delta_{n-1}(\lambda)|\}$ and $\Lambda_n' = \Lambda - \Lambda_n$. Then

$$\int_{\Lambda_n} d\lambda\, n^{-1}\ln|\Delta_n(\lambda)| \leqslant \int_{\Lambda_n} d\lambda\, n^{-1}\ln(\Delta_n^2(\lambda) + \Delta_{n-1}^2(\lambda))$$

$$\leqslant \int_{\Lambda_n} d\lambda\, n^{-1}\ln|\Delta_n(\lambda)| + n^{-1}\ln\sqrt{2}\cdot\text{mes}\,\Lambda_n,$$

$$\int_{\Lambda_n'} d\lambda\, n^{-1}\ln|\Delta_{n-1}(\lambda)|$$

$$\leqslant \int_{\Lambda_n'} d\lambda\, n^{-1}\ln(\Delta_n^2(\lambda) + \Delta_{n-1}^2(\lambda))$$

$$\leqslant \int_{\Delta_n'} d\lambda\, n^{-1}\ln|\Delta_{n-1}(\lambda)| + n^{-1}\ln\sqrt{2}\cdot\text{mes}\,\Lambda_n'.$$

Using again decomposition $\Lambda = \Lambda_n \cap \Lambda_n'$ and Eq. (24), via the latter inequalities, we readily arrive at the following asymptotic equality when $n\to\infty$:

$$\int_\Lambda d\lambda\, n^{-1}\ln(\Delta_n^2 + \Delta_{n-1}^2)^{1/2} = O(n^{-1})$$

$$+ \int_{-\infty}^{\infty}\left[\int_{\Lambda_n}\ln|\lambda - \lambda'|d\lambda\right]dN_n(\lambda')$$

$$+ \int_{-\infty}^{\infty}\left[\int_{\Lambda_n'}\ln|\lambda - \lambda'|d\lambda\right]d$$

$$\times\left[N_n(\lambda') + \frac{n-1}{n}dN_{n-1}(\lambda')\right]. \quad (25)$$

Note that since $|q_n| \leqslant C$, the supports of $dN_n(\lambda)$, and therefore, due to the relation (22), the support of $dN(\lambda)$ too are contained in the interval $[-2-C, 2+C]$. Making use of this fact and applying Lemmas 3, 4(a) and (b), and equality (23), we obtain from Eq. (25) for $n\to\infty$,

$$\int_\Lambda d\lambda\, \gamma(\lambda,\alpha,\omega) = \int_{-\infty}^{\infty}\left[\int_\Lambda \ln|\lambda - \lambda'|d\lambda\right]dN(\lambda'). \quad (26)$$

In doing so, we used the estimate $|\gamma(\lambda,\alpha,\omega)| \leqslant \text{const} \times(|\lambda| + 1)$ which follows from Eq. (23) and Lemma 4(a). But the right-hand side of Eq. (26) does not depend on α and ω and the equality $\gamma(\lambda,\alpha,\omega) = -\gamma(\lambda)$ is possible only for a unique value of α. As a result the rhs of Eq. (26) is nonnegative, and with probability 1, in the lhs of this equation, $\gamma(\lambda,\alpha,\omega)$ can be replaced by $\gamma(\lambda)$. Since the support of the measure $dN(\lambda)$ is bounded and the function $\int_\Lambda |\ln|\lambda - \lambda'||d\lambda$ of λ' is continuous, then the function $|\ln|\lambda - \lambda'||$ is absolutely integrable with respect to the measure $d\lambda\, dN(\lambda')$ in the region $\Lambda\times\mathbb{R}$, and therefore, in the rhs of Eq. (26) the integration order may be interchanged. This completes the proof.

From Lemma 4(c) follows immediately the validity of Theorem 3.

Remark 1: The arguments similar to that in the proof of Lemma 4(c) show that the Lyapunov exponent corresponding to the limiting process $n\to-\infty$ coincides with $\gamma(\lambda)$.

Remark 2: The Thouless formula (7) is true for all, rather than almost all, λ values. One can prove that this is so, e.g., in the case when the sequence q_n in Eq. (5) is an ergodic Markov chain with sufficiently smooth structure.

Remark 3: A formula analogous to Eq. (7), that is, establishing the relation between the Lyapunov exponent and the integrated density of states, exists also in the case of an equation more general than Eq. (5),

$$a_n\psi_{n+1} + a_{n-1}\psi_{n-1} + b_n\psi_n = \lambda\psi_n$$

with metrically transitive coefficients b_n and a_n being sign-constant. In this case it has the form

$$\gamma(\lambda) = -\bar{a} + \int_{-\infty}^{\infty}\ln|\lambda - \lambda'|dN(\lambda'),$$

where

$$\bar{a} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\ln|a_k|.$$

[1]G. Andre and S. Aubrey, Ann. Israel Phys. Soc. **3**, 133 (1980).
[2]A. Casher and J. Lebowitz, J. Math. Phys. **12**, 1701–1711 (1970).
[3]L. A. Pastur, Commun. Math. Phys. **75**, 179–196 (1980).
[4]K. Ishii, Prog. Theor. Phys. Suppl. **53**, 77–138 (1973).
[5]J. Avron and B. Simon, Bull. Amer. Math. Soc. **6**, 81–85 (1982).
[6]A. Ja. Gordon, Usp. Mat. Nauk **31**, 257–258 (1976).
[7]J. Bellisard, R. Lima, and D. Testard, "On the spectrum of the almost Mathieu Hamiltonian," CNRS preprint, Marseille, 1982.
[8]J. L. Dobb, *Stochastic Processes* (Wiley, New York, 1953).
[9]L. A. Pastur and A. L. Figotin, in *Differential equations and methods of functional analysis* (Naukova Dumka, Kiev, 1978), pp. 117–133, and Selecta Mat. Sovetica (in press).
[10]D. Thouless, J. Phys. C **5**, 77–81 (1972).
[11]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976).
[12]V. I. Oseledec, Trudy Moskov. Mat. Obsc. **19**, 197–221 (1968).

The geometric foundations of the integrability property of differential equations and physical systems. II. Mechanics on affinely-connected manifolds and the work of Kowalewski and Painlevé[a)]

Robert Hermann
Association for Physical and Systems Mathematics, 53 Jordan Road, Brookline, Massachusetts 02146

(Received 7 September 1982; accepted for publication 6 October 1983)

In the 1880's Sophie Kowalewski proposed to study "integrability" of differential equations in terms of the analyticity properties of their solutions. The model she studied, the spinning top moving under gravity, has an interesting geometric and algebraic structure, which will be generalized. The differential geometric properties of such models will be investigated in terms of the theory of affinely-connected manifolds and Lie theory.

PACS numbers: 02.30.Hq, 02.40. − k

1. HISTORICAL BACKGROUND

In recent years there has been a revival within the context of mathematical physics and applied mathematics of 19th and early 20th century work on "integrability" of differential and physical system equations. The purpose of this series of papers is to study the differential geometric foundations of various methods which have been successfully used in either the classical or modern work to distinguish the integrable systems from the nonintegrable ones. Thus, my aim is to develop the strategy rather than the tactics of the integrability game.

In the 19th century literature, one can distinguish two such strategic approaches. The first arose from mechanics and Hamilton–Jacobi theory. Given a mechanical system with configuration and momentum coordinates $q = (q^i)$ and $p = (p_j)$, $1 \leqslant i,\ j \leqslant n$, and Hamiltonian $H(q,p)$, one tries to find functions $f(q,p)$ of specified type such that

$$\{H,f\} = 0, \tag{1.1}$$

where $\{\ ,\ \}$ is the usual Poisson bracket.

For example, if H is of the form (which is the usual type encountered in Newtonian mechanics)

$$H = g^{ij}(q)\, p_i p_j + V(q) \tag{1.2}$$

and f is linear in the momentum variables,

$$f = a^i(q)\, p, \tag{1.3}$$

the geometric condition that (1.1) is satisfied is the following one:

The vector field

$$a^i(q)\,\frac{\partial}{\partial q^i}$$

is a Killing vector field for the Riemannian metric $g_{ij}dq^i dq^j$, i.e., generates a one-parameter group of isometries of the metric, and the group leaves invariant the potential function

$$q \to V(q). \tag{1.4}$$

Many of the elementary examples of integrable systems discussed in Whittaker's treatise[1] (he uses the term "soluble" rather than "integrable") are of this type. The next more complicated examples require consideration of f's which are polynomials in p of higher degree:

$$f = a^{i_1 \dots i_m}(q) p_{i_1} \cdots p_{i_m}. \tag{1.5}$$

Now one can write down directly the conditions that (1.1) be satisfied, if H is of the form (1.2) and f of form (1.5). For example, if $m = 2$, among these conditions are those which assert that f is a Killing tensor with respect to the metric (1.2). One can study solutions of the resulting partial differential equations. As described in Part III of this series,[2] Calogero's work involves a generalization of these classical ideas in a certain direction.

However, we will be concerned in this paper with another set of conditions. Find m functions of p and q which are linear in p, with the following conditions satisfied:

$$H = I^{ab} h_a h_b, \tag{1.6}$$

$$\{h_a, h_b\} = F_{ab}(h_1, \dots, h_n), \tag{1.7}$$

where $F_{ab}(\ ,\dots,\ )$ are functions of m real variables, $1 \leqslant a, b \leqslant m$.

$$f = A^{a_1 \dots a_m} h_{a_1} \dots h_{a_m} \tag{1.8}$$

with constants

$$A^{a_1 \dots a_m}.$$

(I^{ab}) is a constant symmetric tensor.

Now condition (1.7) means that the h's form a function group in the sense of Sophus Lie.[3] The simplest type of function groups are the linear ones, when the h_a's satisfy the following conditions:

$$\{h_a, h_b\} = \lambda^{c}_{ab} h_c, \tag{1.9}$$

with constants λ^{c}_{ab}. The h_a then form a Lie subalgebra of the Lie algebra under Poisson bracket of all functions of configuration and momentum variables. (In modern terms, the Lie algebra of C^∞, real-valued functions on the cotangent bundle $T^d(Q)$ of the configuration manifold Q.)

[a)] Supported by Ames Research Center (NASA), Grant NSG-2404; U.S. Army Research Office, Contract #ILI61102RH57-02 Math; NSF #MCS8003227.

The second classical strategy for studying integrability is the analyticity approach of Weirstrass, Poincaré, Kowalewski[4] and Painlevé.[5] (The treatises by Haginara,[6] Ince,[7] and Hille[8] are the best references to this classical work.) To illustrate this point of view in the class of mechanical systems considered above, consider the force-free rotating rigid body. This involves the following specialization of the above data:

$$Q = \mathrm{SO}(3, R),$$
$$m = n = 3, \tag{1.10}$$
$$\lambda^{c}_{ab} = \epsilon^{c}_{ab} = \epsilon^{k}_{ij} = \epsilon_{ijk},$$

the completely skew-symmetric tensor on three indices

$$h^{i} = h^{ij}(q)\, p_j. \tag{1.11}$$

Let $t \to (q(t), p(t))$ be a solution of Hamilton's equation with Hamiltonian H given by (1.6) and (1.11). Then, it is well known (and can be verified directly from the above formulas) that the functions

$$t \to h^{i}(q(t), p(t)) \tag{1.12}$$

are Jacobi elliptic functions.

In particular, they can be extended to be meromorphic functions of the complex variable

$$z = t + i\tau \tag{1.13}$$

and algebraic functions of the initial conditions

$$h^{i}(q(0), p(0)).$$

The $q^{i}(t)$ (which are the Euler angles) are then solutions of linear ordinary differential equations, with coefficients depending on the h^{i}, with singularities determined by the singularities of the functions (1.12).

This example (with other examples of integrable mechanical systems known in the 19th century) led Kowalewski[4] and Painlevé[5] to search for the systems with analogous analyticity properties for their general solutions. The simplest question of this type is then:

> Given a real analytic manifold Q, which real analytic functions H on the cotangent bundle have the property that there are real analytic maps
>
> $$f\colon\ T^{d}(Q) \to \mathbb{C}^{s}, \tag{1.14}$$
>
> such that the maps $t \to f(x(t))$ can be extended to be meromorphic functions of the complex variable $z = t + i\tau$, where $t \to x(t)$ are the solution curves in $T^{d}(Q)$ of the Hamilton equations with Hamiltonian H.

It is this type of question which we will be examining in this paper, and those to follow. Note particularly Painlevé's early work in this direction,[9] which is a prime candidate for a modernization and application to topics of current interest in science and technology.

Another historical thread is the work of V. Arnold,[10] which gives an elegant way of describing mechanical systems which arise as geodesic equations for left-invariant metrics on Lie groups in what came later to be known as "Lax form."[11] In Ref. 12, Chapter 33, I showed that Arnold's work could be formulated in terms of the affine connections which can be canonically defined on Lie groups. In this paper, I plan to develop these ideas further, with the goal of a link-up with the analyticity approaches.

There are, of course, many relations between this work and others in this hyperactive field of "integrable systems." I will only mention Refs. 13–15, which are particularly relevant.

Finally, I want to mention a general geometric insight that arises from this work that might be useful in the current search for the mathematical difference between "chaotic" and "integrable" systems. Let X be a manifold and let $\mathscr{G}$ be a Lie algebra of vector fields on X. (X and/or $\mathscr{G}$ might be infinite-dimensional.) Consider an ordinary differential equation for curves $t \to A(t)$ in $\mathscr{G}$, of the quadratic type

$$\frac{dA}{dt} = B(A, A) + D(A). \tag{1.15}$$

Each solution $t \to A(t)$ of (1.15) generates a flow on X, i.e., a one-parameter family of (possibly local) diffeomorphisms of X by means of the following ordinary differential equation:

The curves $t \to x(t) = g(t)(x_0)$ are the solutions of

$$\frac{dx}{dt} = A(t)(x(t)). \tag{1.16}$$

One obtains in this way a family of curves $t \to x(t)$ on X, which can be identified with the orbits of certain physical systems. Carrying out the Kowalewski–Painlevé program involves studying the analyticity property of the curves $t \to A(t)$, $t \to g(t)$, $t \to x(t)$, where t is extended to be a complex variable.

2. MECHANICAL SYSTEMS ON AFFINELY-CONNECTED MANIFOLDS

I will now carry over and generalize slightly a method presented in Ref. 12, Chapter 33, for describing mechanical systems in terms of affinely-connected manifolds. For the sake of simplicity, we shall deal in this paper only with systems of a finite number of degrees of freedom, although (since it is done in a coordinate-free way) much can be generalized.

Let Q be a finite-dimensional, C^{∞}, parametrized manifold. Let $\mathscr{F}(Q)$ denote the algebra of C^{∞}, real-valued functions on Q. Let $\mathscr{V}(Q)$ be the C^{∞} vector fields on Q considered with the following structures:

> Identified with the cross sections $\Gamma(T(Q))$ of the tangent bundle $T(Q)$ to Q,
> An $\mathscr{F}(Q)$-module structure,
> A real Lie algebra structure $(V_1, V_2) \to [V_1, V_2]$ called Jacobi (or Lie) bracket,
> A Lie derivative derivation action $(V, f) \to V(f)$ of $\mathscr{L}_V(f)$ on $\mathscr{F}(Q)$.

In this paper we shall mainly deal with affine connections in the sense of Koszul, as R-bilinear maps

$$\nabla\colon\ (V_1, V_2) \to \nabla_{V_1} V_2$$

of $\mathscr{V}(X) \times \mathscr{V}(X) \to \mathscr{V}(X)$,

such that

$$\nabla_{fV_1}(V_2) = f\nabla_{V_1} V_2,$$

$$\nabla_{V_1}(fV_2) = V_1(f)V_2 + f\nabla_{V_1}V_2 \tag{2.1}$$

for $f\in\mathscr{F}(X)$; $V_1, V_2\in\mathscr{V}(X)$.

Let ∇ be an affine connection on Q. Given a curve $t\to q(t)$ and a tangent vector field $t\to v(t)\in Q_{q(t)}$ along the curve, we can define[12] a covariant derivative tangent vector field

$$t\to\nabla_{dq/dt}\, v(t).$$

If V_1, V_2 are vector fields, and if $t\to q(t)$ is an orbit curve of V_1 and if $t\to v(t)$ is the restriction of V_2 to the curve, i.e., $V_2(q(t)) = v(t)$, then

$$\begin{aligned}\nabla v(t) &= (\nabla_{dq/dt}\, V_2)(t)\\ &= (\nabla_{V_1(q(t))} V_2)(t).\end{aligned} \tag{2.2}$$

Let $T^d(Q)$ be the cotangent bundle to Q; the fiber above a point $q\in Q$, denoted as Q^d_q, is the dual space to the fiber of $T(Q)$ above q.

Definition: A *mass tensor* is a linear bundle map

$$M\colon\ T(Q)\to T^d(Q), \tag{2.3}$$

i.e., M assigns to a (q, v), $q\in Q$, $v\in Q_q$, a covector $M(q, v)\in Q^d_q$.

A *force tensor* is a map (not necessarily linear)

$$F{:}T(Q)\to T^d(Q), \tag{2.4}$$

which assigns to a point

$$(q, v);\quad q\in Q,\quad v\in Q_q$$

an element

$$F(q, v)\in Q^d_q. \tag{2.5}$$

Definition: A *physical system* with Q as configuration space is a triple

$$(\nabla, M, F)$$

consisting of an affine connection ∇ for Q, a mass tensor M, and force tensor F.

A *trajectory* of the physical system (∇, M, F) is a curve $t\to q(t)$ in Q such that

$$M\left(\nabla_{(dq/dt)}\left(\frac{dq}{dt}\right)\right) = F\left(q, \frac{dq}{dt}\right). \tag{2.6}$$

We can identify the affine connections and mass tensors which appear in Hamilton's equations with Hamiltonian H of "Newtonian" type. Let

$$q = (q^i),\quad 1\leqslant i, j\leqslant n$$

be coordinates of Q,

$$p = (p_i)$$

the corresponding momentum coordinates, and

$$H = \tfrac{1}{2}g^{ij}(q)\,p_i\, p_j + V(q) + a^j(q)\,p_j, \tag{2.7}$$

where

$$\det(g^{ij})\neq 0. \tag{2.8}$$

Let (g_{ij}) be the inverse matrix to (g^{ij}).

$$ds^2 = g_{ij}dq^i dq^j \tag{2.9}$$

is then a Riemannian metric. Let ∇ be the corresponding (torsion-zero) Levi-Cività affine connection.[12]

At any point q^0 of Q, one can choose Riemann normal coordinates (q^i), such that

$$g_{ij}(q^0) = \delta_{ij}, \tag{2.10}$$

$$\nabla_{\partial/\partial q^i}\left(\frac{\partial}{\partial q^j}\right)(q^0) = 0, \tag{2.11}$$

$$\frac{\partial q^{ij}}{\partial q^k}(q^0) = 0. \tag{2.12}$$

Let us use these normal coordinates and the relation (2.10) to express Hamilton's equation for Hamiltonian (2.7) in the terms of the Levi-Cività affine connection. Suppose $t\to(q(t), p(t))$ is a solution with

$$q(0) = q^0. \tag{2.13}$$

$$\begin{aligned}\frac{d}{dt}q^i(t) &= \frac{\partial H}{\partial p^i}\\ &= q^{ij}p_j + a^i,\end{aligned}$$

hence

$$\frac{d}{dt}(q^i(t))\bigg|_{t=0} = p_i(0) + a^i(q^0) \tag{2.14}$$

$$\frac{d}{dt}(p^i(t))\bigg|_{t=0} = -\frac{\partial V}{\partial q^i}(q^0) + \nabla_{\partial/\partial q^i}(A)(q^0), \tag{2.15}$$

where

$$A = a^j\frac{\partial}{\partial q^j}\in\mathscr{V}(Q). \tag{2.16}$$

Now, let

$$M\colon\ T(Q)\to T^d(Q) \tag{2.17}$$

be the linear-bundle isomorphism which is the isomorphism induced by the Riemann metric ds^2.

We see that we have proved the following result:

Theorem 2.1: The solutions of Hamilton's equations with Hamiltonian (2.7) are identical with Eqs. (2.6), where ∇ is the Levi-Cività affine connection, $F{:}\,T(Q)\to T^d(Q)$ is given by the following formula:

$$F(q, v) = d\,f(q) + \nabla_v(A)(q). \tag{2.18}$$

This result is mentioned here mainly to link up the developments to be presented below with the traditional approaches via Hamilton-Jacobi theory mentioned in Sec. 1 (and which take their modern form in the theory of "symplectic manifolds" and "Lagrangian submanifolds"). The material we turn to now is more Lie-theoretic in nature and is, I believe, better adapted to studying the "analyticity" questions of the classical literature.[4,5]

3. FLOWS AND THEIR INFINITESIMAL GENERATORS

Let Q continue as a manifold with an affine connection ∇. Let $\mathscr{V}(Q)$ be the Lie algebra of C^∞ vector fields on Q. Let t be a real variable, $0\leqslant t<\infty$, parametrizing a one-dimensional manifold that we call R_+.

Let $t\to V(t)$ be a curve in $\mathscr{V}(Q)$, which is C^∞ in the sense that the mapping

$$(t, q)\to(t, V(t)(q))$$

of $R_+\times Q\to T(Q)$ is C^∞.

A curve $t\to q(t)$ in Q is an orbit of $t\to V(t)$ if

$$\frac{dq}{dt} = V(t)(q(t)) \tag{3.1}$$

for $t \in R_+$.

Assume for simplicity that a solution of (3.1) exists for all t. (If this condition is not satisfied, the analysis can be extended in the context of Ehresmann's theory of pseudogroups.) Solving the ordinary differential equation (3.1) with initial condition q_0 at $t = 0$ determines a one-parameter family

$$t \rightarrow \phi_t \colon Q \rightarrow Q \tag{3.2}$$

of diffeomorphisms such that

$$\phi_t(q_0) = q(t), \tag{3.3}$$

$$\phi_0 = \text{identity}. \tag{3.4}$$

For $f \in \mathscr{F}(Q)$,

$$\frac{\partial}{\partial t} f(\phi_t(q_0)) = V(t)(\phi_t(q_0))(f)$$

$$= V(t)(f)(\phi_t(q_0)),$$

or

$$\frac{\partial}{\partial t} \phi_t^*(f)(q_0) = \phi_t^*(V(t)(f))(q_0),$$

or, since q_0 is arbitrary,

$$V(t)(f) = \phi_t^{-1*} \frac{\partial}{\partial t} \phi_t^*(f) \tag{3.5}$$

for $f \in \mathscr{F}(X)$.

The one-parameter family $t \rightarrow \phi_t$ of diffeomorphisms of Q is called a flow; the curve $t \rightarrow V(t)$ in $\mathscr{V}(Q)$ is called its infinitesimal generator.

Remark: This is a generalization of material in fluid mechanics. The velocity field in the Eulerian picture is $t \rightarrow V(t)$. It is also basically the Heisenberg picture of quantum mechanics.

Now, return to (3.1). For the purposes of mechanics we want to compute the covariant derivative, with respect to the affine connection ∇, of the "velocity" field $t \rightarrow (dq/dt)(t)$ of the curve $t \rightarrow q(t)$, in terms of the infinitesimal generator $t \rightarrow V(t)$.

Theorem 3.1: Let $t \rightarrow q(t)$ be a curve in Q, with tangent vector field $t \rightarrow dq/dt$. Let $t \rightarrow V(t)$ be a curve in $\mathscr{V}(Q)$. Let $t \rightarrow v(t)$ be the vector field along q given by the following formula:

$$v(t) = V(t)(q(t)). \tag{3.6}$$

Then,

$$\nabla_{dq/dt}\, v(t) = \left(\frac{dV}{dt}\right)(q(t)) + (\nabla_{dq/dt} V(t))(q(t)) \tag{3.7}$$

Proof: (3.7) is the covariant-derivative version of the Leibniz rule for differentiating products. dV/dt is the derivative of the curve $t \rightarrow V(t)$ as a curve in the vector space $\mathscr{V}(Q)$, i.e., for $f \in \mathscr{F}(Q)$:

$$\frac{dV}{dt}(f) = \lim_{\Delta t \rightarrow 0} \frac{V(t + \Delta t)(f) - V(t)(f)}{\Delta t}. \tag{3.8}$$

The affine connection defines a parallel transport along the curve $t \rightarrow q(t)$:

$$\nabla_{dq/dt}\, v = \lim_{\Delta t \rightarrow 0} \frac{v(t + \Delta t) - v(t)}{\Delta t}. \tag{3.9}$$

To interpret (3.9), i.e., to be able to subtract $v(t)$ from $v(t + \Delta t)$, which lie in different vector spaces, namely $Q_{q(t)}$ and $Q_{q(t+\Delta t)}$, it must be understood that they are parallel transported along the curve $t \rightarrow q(t)$. Formula (3.7) now results from putting (3.9) together with (3.7) in the usual Leibnizian way.

Corollary 3.2: If $t \rightarrow q(t)$ is a curve in Q which is an orbit of the flow generated by the curve $t \rightarrow V(t)$ in $\mathscr{V}(Q)$ (in fluid mechanics, termed a velocity field), then the acceleration vector field

$$t \rightarrow a(t) = \nabla_{dq/dt} \frac{dq}{dt} \tag{3.10}$$

is given by the following formula:

$$a(t) = \left(\frac{dV}{dt}\right)(q(t)) + \nabla_{dq/dt}(V(t))(q(t)). \tag{3.11}$$

This formula (which must go back to Euler) plays a basic role in many physical theories. I have used it in Chapter 33 of Ref. 6 to derive the Eulerian equation of motion of a rotating rigid body. It could also be used in the context of Arnold's work[10] to derive the ideal fluid motion equation.

4. THE EQUATIONS OF MOTION OF A PHYSICAL SYSTEM IN EULERIAN FORM

Continue with the notation of Sec. 3. Let

$$M\colon T(Q) \rightarrow T^d(Q),$$

$$F\colon T(Q) \rightarrow T^d(Q),$$

be mass and force tensors, as described above. The (∇, M, F) determine a physical system. The trajectories are the curves $t \rightarrow q(t)$ such that:

$$M\left(\nabla_{dq/dt}\left(\frac{dq}{dt}\right)\right) = F\left(\frac{dq}{dt}\right). \tag{4.1}$$

Suppose that $t \rightarrow V(t)$ is a curve in $\mathscr{V}(Q)$ such that all orbits

$$t \rightarrow \phi_t(q_0) = q(t)$$

of the flow it generates are trajectories of the physical system. Then, we have, using Theorem 3.1

$$M\frac{dV}{dt}(q, t) + \nabla_{V(t)} V(t)(q) = F(V(t)(q)) \tag{4.2}$$

for all $q \in Q$.

These are the Eulerian equations of motion. They are a set of ordinary nonlinear differential equations for the curve $t \rightarrow V(t)$ in $\mathscr{V}(Q)$, an infinite-dimensional vector space.

Now, we can interpret these equations as some sort of generalized Lax equations by means of the following theorem.

Theorem 4.1: Let $\mathscr{G}$ be a Lie algebra of vector fields on the manifold X. Let $t \rightarrow V(t)$ be a curve in $\mathscr{G}$ which satisfies (4.2). Consider the flow on Q generated by $t \rightarrow V(t)$. Then, each orbit of this flow, i.e., each curve $t \rightarrow q(t)$ which satisfies

$$\frac{dq}{dt} = V(t)(q(t)) \tag{4.3}$$

is a solution of Eq. (4.1). Thus, we have set a correspondence between the Eulerian equation (4.3) considered as a set of ordinary differential equations for curves in $\mathscr{G}$ and the ordinary differential equations whose solutions lie in the manifold Q.

Now, Eqs. (4.2) are precisely of the Lax form[11] if $F = 0$, and if $\nabla_V V$ can be expressed as

$$\nabla_V V = [B, V], \tag{4.4}$$

where B is another element of $\mathscr{G}$. For the case where ∇ is the Levi-Cività affine connection associated with a left-invariant Riemannian metric on a semisimple Lie group G (so that $X = G$), and $\mathscr{G}$ is the Lie algebra of left-invariant vector fields, it is proved in Ref. 12 that (4.4) holds,

$$B = \beta(V),$$

where β: $\mathscr{G} \to \mathscr{G}$ is a linear map.

5. SOPHIE KOWALEWSKI'S SYSTEM AND EXTENDED LAX SYSTEMS

Sophie Kowalewski's work[7] on the integrability of a rotating-under-gravity rigid body is one of the most famous works of mathematical physics in the 19th century, linking two main theories of the 19th century mathematics—analytical mechanics and complex function theory—in a beautiful way. This link between "integrability" and "analyticity" is still of great interest for physics.

This paper will begin to link up her work on the connection between "integrability" and "analyticity", and the modern work on Lax systems. As I have already point out briefly in Ref. 16, her paper suggests a generalization of the Lax framework.

Using the notation of her papers,[4] the differential equations are as follows:

$$\begin{aligned}
A\frac{dp}{dt} &= (B - C)qr + Mg(g_0\gamma'' - a_0\gamma'),\\
B\frac{dq}{dt} &= (C - A)rp + Mg(z_0\gamma - x_0\gamma''),\\
C\frac{dr}{dt} &= (A - B)pq + Mg(x_0\gamma' - y_0\gamma),\\
\frac{d\gamma}{dt} &= r\gamma' - q\gamma'',\\
\frac{d\gamma'}{dt} &= p\gamma'' - r\gamma,\\
\frac{d\gamma''}{dt} &= q\gamma - p\gamma'.
\end{aligned} \tag{5.1}$$

(x_0, y_0, z_0) are the coordinates of the center of gravity of the body. A, B, C, U, g are parameters. p, q, r are the components of the angular momentum with respect to an orthonormal frame moving on the body. $\gamma, \gamma', \gamma''$ are the direction cosines of the moving frame with respect to an orthonormal frame fixed in space. The gravitational constant is g.

When $z_0 = y_0 = z_0 = 0$, Eqs. (5.1) decouple to the following set:

$$\begin{aligned}
A\frac{dp}{dt} &= (B - C)qr,\\
B\frac{dq}{dt} &= (C - A)rp,\\
C\frac{dr}{dt} &= (A - B)pq,
\end{aligned} \tag{5.2}$$

$$\begin{aligned}
\frac{d\gamma}{dt} &= r\gamma' - q\gamma'',\\
\frac{d\gamma'}{dt} &= p\gamma'' - r\gamma,\\
\frac{d\gamma''}{dt} &= q\gamma - p\gamma'.
\end{aligned} \tag{5.3}$$

These may be solved by first solving (5.2), which are the Euler equations for a force-free rotating rigid body, and then substituting the result into the right-hand side of (5.3), which is a set of linear, variable-coefficient ordinary differential equations for the $(\gamma, \gamma', \gamma'')$. Thus, if $t \to (p(t), q(t), r(t))$ is a solution of (5.2), it can be extended to a meromorphic function of $z = t + i\tau$, whose poles depend on the initial data $(p(0), q(0), r(0))$. (5.3) then is a set of linear differential equations for $t \to (\gamma(t), \gamma'(t), \gamma''(t))$ whose coefficients are meromorphic. By the usual theorems for linear ordinary differential equations and analytical continuation, the solutions of (5.2) and (5.3) can be extended to be analytic functions in a Riemann surface sitting over the complex z plane with certain points removed. In particular, there are no branch cuts necessary. It is this property of the force-free rigid body which Painlevé focussed on in his work[5,9] and that first arose in Kowalewski's work.

Now, Arnold showed[10] how to write (5.2) in what later became known as "Lax form." Set

$$(p, q, r) = \begin{pmatrix} 0 & r & -q \\ -r & 0 & p \\ q & -p & 0 \end{pmatrix}. \tag{5.4}$$

Let G be the Lie group SO(3, R) of 3×3 real orthogonal matrices of determinant $+1$. Let $\mathscr{G}$ be its Lie algebra, namely the Lie algebra of 3×3 real skew-symmetric matrices. Then, if (p, q, r) is interpreted as a point in R^3, α may be considered as a linear map

$$\alpha: R^3 \to \mathscr{G}.$$

Theorem 5.1: There is another linear map β: $R^3 \to \mathscr{G}$ such that the Eqs. (5.2) take the following form:

$$\frac{d}{dt}\alpha(p, q, r) = [\beta(p, q, r), \alpha]. \tag{5.5}$$

Proof: Let us find β so that Eqs. (5.2) are equivalent to (5.4)

$$\beta = \begin{pmatrix} 0 & \beta_{12} & \beta_{13} \\ -\beta_{12} & 0 & \beta_{23} \\ -\beta_{13} & -\beta_{23} & 0 \end{pmatrix}. \tag{5.6}$$

We require that

$$[\beta,\alpha]=\begin{pmatrix}0 & \frac{A-B}{C}pq & \frac{A-C}{B}rp\\ \frac{B-A}{C}pq & 0 & \frac{B-C}{A}qr\\ \frac{C-A}{B}rp & \frac{C-B}{A}qr & 0\end{pmatrix}. \tag{5.7}$$

If the coefficients β_{ij} of (5.6) are written as linear functions of (p, q, r), (5.7) constitutes a set of linear equations for the coefficients that can readily be solved.

Having found α and β, we can now write (5.1) as a set of matrix differential equations. Set

$$\Gamma=\begin{pmatrix}0 & \gamma & \gamma'\\ -\gamma & 0 & \gamma''\\ -\gamma' & \gamma'' & 0\end{pmatrix}, \tag{5.8}$$

$$\frac{d\alpha}{dt}=[\beta(\alpha),\alpha]+[\Lambda,\Gamma], \tag{5.9}$$

where Λ is a constant 3×3 skew-symmetric matrix, determined by the constants M, g, x_0, y_0, z_0 in (5.1).

$$\frac{d\Gamma}{dt}=[\alpha,\Gamma].$$

These equations are a special case of what I have called[16] an extended Lax system, which we will now define independently.

6. EXTENDED LAX SYSTEMS AND QUADRATIC DIFFERENTIAL EQUATIONS

Let $\mathscr{G}$ be a Lie algebra, V a real vector space, λ a linear map from V to $\mathscr{G}$, β a map $\mathscr{G}\to\mathscr{G}$, C a map from $\mathscr{G}$ to $L(V, V)$, the space of linear maps $V\to V$. The following differential equations:

$$\frac{dA}{dt}=[\beta(A),A]+\lambda(v(t)),$$

$$\frac{dv}{dt}=C(A)(v), \tag{6.1}$$

will be said to form an extended Lax system.

Notice that when the "coupling" λ is zero, the system (6.1) decouples to the following one:

$$\frac{dA}{dt}=[\beta(A),A], \tag{6.2}$$

$$\frac{dv}{dt}=\alpha(A)v, \tag{6.3}$$

i.e., a Lax system (6.2) that is then solved and fed into the system (6.3) to obtain a linear, time-varying system.

We can now combine the two components of the system (6.1) in the following way: Let

$$\mathscr{A}' = \text{vector space direct sum of } \mathscr{G} \text{ and } V, \tag{6.4}$$

i.e., $\mathscr{G}\oplus V$.

Let B be the following bilinear map: $\mathscr{A}'\times\mathscr{A}'\to\mathscr{A}'$

$$B(A\oplus v, A_1\oplus v_1)=[\beta(A),A_1]\oplus C(A)(v_1). \tag{6.5}$$

Let D be the following linear map: $\mathscr{A}\to\mathscr{A}$

$$D(A\oplus v)=\lambda(v)\oplus 0. \tag{6.6}$$

Theorem 6.1: With these definitions, the extended Lax systems can be written in the following form:

$$\frac{dA'}{dt}=B(A',A')+D(A') \tag{6.7}$$

for $A'\in\mathscr{A}\equiv\mathscr{G}+V$.

I will not pursue at the moment the analysis of these equations from the Painlevé point of view. (That is on the agenda for a later paper in this series.) As a first step, I will repeat the course followed by Kowalewski in her initial work[4] and look for solutions which seem likely to be meromorphic. Thus, if the "general solution" can be proved to consist of meromorphic functions, it might be possible to prove the Painlevé property directly in certain cases.

7. FORMAL LAURENT SERIES SOLUTIONS OF BILINEAR EQUATIONS

We can now partially algebracize some material in Sophie Kowalewski's papers.[4] Let $\mathscr{A}$ be a real or complex vector space,

$$B:\ \mathscr{A}\times\mathscr{A}\to\mathscr{A} \tag{7.1}$$

a bilinear map. Let

$$D:\ \mathscr{A}\to\mathscr{A} \tag{7.2}$$

be a linear map.

Consider the following differential equation:

$$\frac{dA}{dt}=B(A,A)+D(A) \tag{7.3}$$

to be solved for a curve $t\to A(t)$ in $\mathscr{A}$. The most straightforward way of investigating the analyticity of (7.3) is to assume a solution exists in the form of a Laurent series, substitute it into the differential equations, and investigate the compatibility of the resulting algebraic equations for the coefficients of the series. This was the method used by Kowalewski[4] in her first approach to the weighted spinning top.

Let us use this approach on Eq. (7.3). Suppose $t\to A(t)$ is a solution of (7.3) of the following form:

$$A(t)=t^m\alpha(t), \tag{7.4}$$

where m is a complex number, and $m\neq 0, 1, 2, \ldots$

$$\alpha(t)=A_0+A_1t+\cdots,\quad A_0\neq 0 \tag{7.5}$$

is a formal power series. Then

$$\frac{dA}{dt}=mt^{m-1}\alpha+t^m\frac{d\alpha}{dt}, \tag{7.6}$$

$$B(A,A)=t^{2m}B(\alpha,\alpha), \tag{7.7}$$

$$\frac{dA}{dt}-B(A,A)-D(A)=mt^{m-1}\alpha+t^m\frac{d\alpha}{dt}-t^{2m}B(A,A)-t^mD(\alpha(t)). \tag{7.8}$$

Theorem 7.1: With the above notation, suppose

$$\frac{dA}{dt}-B(A,A)-D(A)$$

is a formal power series in t containing only nonnegative powers. Then,

$$m = -1, \tag{7.9}$$

$$A_0 = B(A_0, A_0). \tag{7.10}$$

Proof: Using (7.7) and (7.8),

$$\begin{aligned} \frac{dA}{dt} &- B(A, A) - D(A) \\ &= mt^{m-1}(A_0 + tA_1 t\ldots) \\ &\quad + t^m(A_1 + 2tA_2 + \cdots) \\ &\quad - t^{2m}B(A_0 + tA_1 + \cdots, A_0 + tA_1 + \cdots) \\ &\quad - t^m D(A_0 + tA_1 + \cdots). \end{aligned} \tag{7.11}$$

Using (7.4) and (7.9), (7.10) follows from (7.11), and equating to zero the coefficients of t^{-2} on the right-hand side of (7.11).

Let us now insert conditions (7.8)–(7.9) into $A(t)$, and find the recursion relations for the coefficients of the power series:

$$A(t) = \sum_{j=-1}^{\infty} A_{j+1} t^j, \tag{7.12}$$

$$\begin{aligned} \frac{dA}{dt} &= \sum_{j=-1}^{\infty} jA_{j+1} t^{j-1} \\ &= \sum_{k=-2}^{\infty} (k+1)A_{k+2} t^k, \end{aligned} \tag{7.13}$$

$$B(A, A) = \sum_{k=-2}^{\infty} t^k \sum_{\substack{i+j=k \\ i,j\geqslant -1}} [B(A_i), A_j], \tag{7.14}$$

$$\begin{aligned} \frac{dA}{dt} &- B(A, A) - D(A) \\ &= \sum_{k=-2}^{\infty} t^k \left[(k+1)A_{k+2} - \sum_{\substack{i+j=k+2 \\ i,j\geqslant 0}} B(A, A) \right. \\ &\quad \left. - \sum_{k=-2}^{\infty} D(A_{k+2})\right]. \end{aligned} \tag{7.15}$$

It now follows from (7.15) that:

Theorem 7.2: The t^{-1} coefficient in

$$\frac{dA}{dt} - [B(A), A] - D(A)$$

vanishes if and only if the following condition is satisfied:

$$D(A_1) + B(A_0, A_1) + B(A_1, A_0) = 0. \tag{7.16}$$

Let us now set equal to zero, for $k\geqslant 0$, the terms in (7.15)

$$A_2 = B(A_0, A_2) + B(A_1, A_1) + B(A_2, A_0) + D(A_2)$$

$$\vdots$$

$$\begin{aligned} (k-1)A_k &= \sum_{\substack{i+j=k \\ i,j\geqslant 0}} B(A_i, A_j) \\ &= B(A_0, A_k) + B(A_1, A_{k-1}) \\ &\quad + \cdots + B(A_k, A_0) + D(A_k). \end{aligned} \tag{7.17}$$

Let us set this up more algebraically. For $A_0 \in \mathscr{A}$, let $\gamma_{A_0,k}$ be the following linear map: $\mathscr{A} \to \mathscr{A}$

$$\gamma_{A_0,k}(A) = (k-1)A - B(A_0, A) - B(A, A_0) + D(A). \tag{7.18}$$

Then, we can write the recursion relation (7.9) in the following way:

$$\gamma_{A_0,k}(A_k) = \sum_{\substack{i+j=k \\ i,j\geqslant 1}} [B(A_i), A_j]. \tag{7.19}$$

Thus, we have proved the following result:

Theorem 7.3: Suppose that A_0, A_1 are elements of $\mathscr{A}$ satisfying (7.10) and (7.16). Suppose also that $\gamma_{A_0,k}$, for $k\geqslant 2$, is an onto linear map of $\mathscr{A} \to \mathscr{A}$. Then, a formal Laurent series solution

$$A(t) = A_0 t^{-1} + A_1 + A_1 t + \cdots$$

of Eq. (7.2) exists.

Let us now investigate further a simple situation involving an ansatz for B.

8. SOLUTIONS ASSOCIATED WITH INVOLUTIVE AUTOMORPHISMS OF LIE ALGEBRAS

Suppose now that $\mathscr{A}$ has a Lie algebra structure

$$(A_1, A_2) + [A_1, A_2]. \tag{8.1}$$

Suppose, in this section, that

$$\beta: \mathscr{A} \to \mathscr{A}$$

is a linear map satisfying the following conditions:

$$\beta([A_1, A_2]) = [\beta(A_1), \beta(A_2)], \tag{8.2}$$

i.e., β is an automorphism of the Lie algebra structure (8.1). Suppose also that

$$\beta^2 = 1, \tag{8.3}$$

i.e., β is an involutive automorphism of the Lie algebra structure (8.1). (Recall that such objects play an important role in the theory of symmetric spaces.)

We can decompose $\mathscr{A}$ as the direct sum

$$\mathscr{A} = \mathscr{A}^+ + \mathscr{A}^- \tag{8.4}$$

of linear subspaces such that

$$\beta(A) = \pm A \tag{8.5}$$

for $A \in \mathscr{G}^\pm$.

Also,

$$[\mathscr{A}^+, \mathscr{A}^\pm] \subset \mathscr{A}^\pm, \tag{8.6}$$

$$[\mathscr{A}^-, \mathscr{A}^\pm] \subset \mathscr{A}^\mp. \tag{8.7}$$

We can proceed now to construct the "general solution" of the Arnold–Lax equation, in this case. Suppose

$$t \to A(t)$$

is an analytic solution of

$$\frac{dA}{dt} = [\beta(A), A], \tag{8.8}$$

where β is an involutive automorphism of the Lie algebra $\mathscr{A}$. Then,

$$\begin{aligned} \frac{d}{dt}(A + \beta(A)) &= [\beta(A), A] \pm \beta([\beta(A), A]) \\ &= [\beta(A), A] + [A, \beta(A)] \\ &= 0. \end{aligned} \tag{8.9}$$

In other words,

$$A + \beta(A) \tag{8.10}$$

is independent of t. The projection map

$$\begin{aligned} \pi^{+}&: \mathscr{A} \rightarrow \mathscr{A}^{+}, \\ A &\rightarrow A+\beta(A) \end{aligned} \tag{8.11}$$

is a "conserved quantity" for the differential equation (8.8), i.e., the orbits of the vector field

$$V(A)=[\beta(A), A] \tag{8.12}$$

on $\mathscr{A}$ are contained in the fibers of π^{+}, and the vector field V is tangent to the fibers of π^{+}.

Now,

$$\begin{aligned} \frac{d}{dt}(A(t)) &= [\beta(A), A] \\ &= [\pi^{+}(A), A]. \end{aligned} \tag{8.13}$$

Since $\pi^{+}(A)$ is independent of t, this is a constant-coefficient linear-differential equation, which can be solved explicitly as follows:

$$\begin{aligned} A(t) &= \exp(\operatorname{ad} \pi^{+}(A(0)) t)(A(0)) \\ &= A(0)+t[\pi^{+}(A(0)), A(0)]. \end{aligned} \tag{8.14}$$

In particular, note that all solutions of the Arnold–Lax equation are analytic in this case.

[1]E. T. Whittaker, *A Treatise on Analytical Dynamics of Particles and Rigid Bodies*, (Cambridge University, London, 1959).

[2]R. Hermann, "The geometric foundation of the integrability property of differential equations and physical systems," I–IV, Part I, J. Math. Phys. **24**, 2422 (1983); Parts III and IV (to be published).

[3]S. Lie, *Transformationsgruppen*, Vol. II (Chelsea, New York, 1967).

[4]S. Kowalewski, "Sur le probleme de la rotation d'un corps solide autour d'un point fixe; Sur une propriete du systeme d'equations differentielles qui definit la rotation d'un corps solide autour d'un pointe fixe," Acta Math. **12**, 177–232 (1889); **14**, 81–93 (1891).

[5]P. Painlevé, *Lecons de Stockholm, Oeuvres* (CNRS, Paris, 1972), Vol. 1, pp. 199–791.

[6]Y. Hagihara, *Celestial Mechanics*, Vol. 1 (MIT, Cambridge, Mass., 1970).

[7]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1947).

[8]E. Hille, *Ordinary Differential Equations in the Complex Domain* (Wiley, New York, 1976).

[9]P. Painlevé, "Memoire sur les équations différentielles dont l'integrale générale est uniforme," Bull. Soc. Math. **28**, 201–261 (1900). Ref. 5, Vol. 3, pp. 123–183.

[10]V. Arnold, Ann. Inst. Grenoble **16**, 319 (1966).

[11]P. D. Lax, "Integrals of nonlinear equations of evolution and solitary waves," Comm. Pure Appl. Math. **21**, 467–490 (1968).

[12]R. Hermann, *Differential Geometry and the Calculus of Variations* (Academic, New York, 1967).

[13]A. Weinstein, "The local structure of Poisson manifolds," preprint, University of California, Berkeley, 1983.

[14]J. Marsden, T. Ratiu, and A. Weinstein, "Semi-direct products and reduction in mechanics" (to appear).

[15]A. M. Perolomov, Comm. Math. Phys. **81**, 239–241 (1981).

[16]R. Hermann, "The Kowalewski top and extended Lax systems" (unpublished).

The Schrödinger–Langevin equation: Special solutions and nonexistence of solitary waves

L. Brüll and H. Lange
Mathematisches Institut, Universität Köln, Weyertal 86-90, D-5000 Köln 41, West Germany

(Received 9 May 1983; accepted for publication 17 June 1983)

Two types of solutions of the Schrödinger–Langevin equation are investigated. It is proved that a special type of Gaussian solutions exist globally in time for the harmonic oscillator Hamiltonian. Furthermore, it is shown that the Schrödinger–Langevin equation can have no solitary wave type solutions in the damped free-particle case which lie in L^2.

PACS numbers: 02.30.Jr, 03.65. – w

I. INTRODUCTION

The theory of quantum-mechanical treatment of dissipative systems has received attention recently due to the interest in heavy ion physics and frictional problems in fission; furthermore, we may mention the quantum theory of lasers, the theory of detection of photons, and the theory of fission of atomic nuclei (see Messer,[1] Hasse,[2] Dekker,[3] Svinin,[4,5] and the literature cited in these references).

A model to describe a quantal dissipative system acted upon by a frictional force was introduced by Griffin and Kan,[6] and Kostin[7]: It is a nonlinear one-dimensional time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t}\psi(x,t) = \left\{-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right\}\psi(x,t) + \frac{\hbar\gamma}{2i}\left\{\log\frac{\psi(x,t)}{\bar\psi(x,t)} - \left\langle \log\frac{\psi(x,t)}{\bar\psi(x,t)}\right\rangle\right\}\psi(x,t) \tag{1.1}$$

with a potential $V=V(x)$ (which may be a stochastic potential, but we consider the nonstochastic case only), and

$$\left\langle \log\frac{\psi(x,t)}{\bar\psi(x,t)}\right\rangle = \int_{-\infty}^{\infty} \log\frac{\psi(x,t)}{\bar\psi(x,t)}\,|\psi(x,t)|^2\,dx$$

is a frictional or damping coefficient. Since $(1/2i)\log z/\bar z = \arg z$, we prefer to write (1.1) in the form

$$i\psi_t = -\partial_x^2\psi + V\psi + \gamma(\arg\psi - W(\psi))\psi \tag{1.2}$$

with some obvious normalization, and $W(\psi) = \int_{-\infty}^{\infty}\arg\psi|\psi|^2\,dx$; (1.1) is usually called the Schrödinger–Langevin equation. In connection with (1.2), we also consider the nonlinear Schrödinger equation

$$i\psi_t = -\partial_x^2\psi + V\psi + \gamma\arg\psi\psi, \tag{1.3}$$

which we call the argument-Schrödinger equation or simply the argument equation. Equation (1.1) [and (1.3), too] has several remarkable properties. First, the L^2-norm of the wave function ψ is preserved for all time (see Sec. IV for a further conservation law). Secondly, eigenfunctions of the Hamiltonian $H = -(h^2/2m)\partial_x^2 + V$ are, in general, (undamped) solutions to (1.1) [for the phase $s(x,t) = -(\hbar/2i)\times\log\psi(x,t)/\bar\psi(x,t)$ is usually independent of x; see Skagerstam[8,9]]. Furthermore, the nonlinearity is discontinuous if we consider it as a single-valued mapping.

Several interesting special solutions to (1.1) are known, so for the case $V=0$, (1.1) has solutions of the form

$$\psi(x,t) = Ne^{ikx}\exp[-(x-X(t))^2/2(\alpha_R + i\alpha_I)], \tag{1.4}$$

where $k = k_0e^{-\gamma t}$, $X(t) = X_0 + (\hbar k_0/\gamma)(1-e^{-\gamma t})$. The real functions α_R, α_I have to satisfy the following coupled system of first-order differential equations in the time variable,

$$\dot\alpha_R = \gamma\alpha_I 2\alpha_R\alpha_I/(\alpha_R^2+\alpha_I^2),$$
$$\dot\alpha_I = \hbar/m - \gamma\alpha_I(\alpha_R^2-\alpha_I^2)/(\alpha_R^2+\alpha_I^2) \tag{1.5}$$

(see Immele, Kan, and Griffin[10]; there seems to be numerical evidence that (1.5) has nontrivial solutions existing globally in time).

In the case of the damped harmonic oscillator (i.e., $V(x) = \frac12 m\omega^2x^2$), the following solutions are known:

$$\psi(x,t) = \chi_0(x-X(t))\exp\Big(-\frac12\omega t\Big)\exp\Big\{\frac{i}{\hbar}\Big(mx\dot X(t) + \frac12 m\gamma(X(t)^2-a^2) - \int_0^t L_0(s)ds\Big)\Big\}, \tag{1.6}$$

where

$$\chi_0(y) = (m\omega/\pi\hbar)^{1/4}\exp(-(m\omega/2\hbar)y^2),$$
$$L_0(s) = \tfrac12 m\dot X(t)^2 - \tfrac12 m\omega^2X(t)^2,$$
$$X(t) = ae^{-(1/2)\gamma t}(\cos\Omega t + (\gamma/2\Omega)\sin\Omega t),$$
$$\Omega^2 = \omega^2 - \tfrac14\gamma^2 > 0.$$

This solution can be generalized to

$$\psi(x,t) = \phi_n(x-X(t))\exp(-i(E_n/\hbar)tx\dot X(t)) \times\exp(ig(X(t),\dot X(t),t)), \tag{1.7}$$

where ϕ_n are eigenfunctions of the Hamiltonian

$$H = -(\hbar^2/2m)\partial_x^2 + \tfrac12 m\omega^2x^2$$

of the harmonic oscillator with eigenvalues

$$E_n = (n+\tfrac12)\hbar\omega,\quad \ddot X(t) + \gamma\dot X(t) + \omega^2X(t) = 0,$$

and $g(X,\dot X,t)$ is given by

$$\hbar\frac{dg(X(t),\dot X(t),t)}{dt} = m\gamma X(t)\dot X(t) + \tfrac12 m\omega^2X^2(t) - \tfrac12 m\dot X(t)^2$$

 0022-2488/84/040786-05$02.50

(see Skagerstam,[8,9] Griffin, and Kan[6]); here, also, one must have the weak damping assumption

$$\omega^2 > \gamma^2/4. \tag{1.8}$$

Kostin[7] constructs a solution of (1.1) for $V=0$ of the form $\psi(x,t) = e^{i\Theta(t)}\exp((i/\hbar)[\alpha(t)x + \beta(t)])$ which unfortunately is not in L^2. If ϕ is a solution of the argument equation with $\int_{-\infty}^{\infty} |\phi|^2 dx = 1$, then via a function $\Theta(t)$ with $\dot{\Theta}(t) = \gamma W(\phi)$, one gets a solution of (1.2) with

$$\psi(x,t) = e^{i\Theta(t)}\phi(x,t) \tag{1.9}$$

(and vice versa; see Kostin[7]), so it suffices to consider the argument equation instead of (1.1) or (1.2).

In this paper we consider first (Sec. II) Gaussian solutions of (1.3) for the harmonic oscillator potential $V(x) = \omega^2 x^2$ of the form

$$\psi(x,t) = e^{i(\alpha(t)x^2 + \beta(t))} = e^{i(\alpha_1(t)x^2 + \beta_1(t))}\, e^{-\alpha_2(t)x^2 - \beta^2(t)}, \tag{1.10}$$

with real functions α_j, β_j; we show that nontrivial solutions of that form exist for all time if $0 \leqslant \omega^2 \leqslant 1/16$ (for $\gamma = 1$). For Eq. (1.1), with $V(x) = \frac{1}{2}m\omega^2 x^2$ and arbitrary γ, this last condition is the same as $\omega^2 \leqslant \frac{1}{4}\gamma^2$; therefore solutions of the form (1.10) are, in some sense, complementary to those Gaussian solutions of the shape (1.6) or (1.7) with condition (1.8).

In Sec. III, we show that the argument equation (1.3) (with $V=0$; damped free particle) can have no nontrivial solitary wave solutions in L^2, i.e., there are no nontrivial solutions to (1.3) of the form

$$\psi(x,t) = \psi(x + ct),$$

with $c \in \mathbb{R}$, $\psi \in L^2(\mathbb{R}) \cap C^2(\mathbb{R})$. Especially there is no L^2-solution of (1.1), (1.2), or (1.3), which is time-independent (and $\neq 0$), but there are stationary solutions of (1.1) of the form $\psi(x,t) = e^{i\lambda t}\psi_0(x)$.

We close with some remarks on open problems, generalizations, and further properties of the discussed equations (Sec. IV).

II. GAUSSIAN SOLUTIONS

In this section, we consider the harmonic oscillator argument equation

$$i\psi_t = -\partial_x^2\psi + \omega^2 x^2 \psi + \gamma \arg \psi \psi, \tag{2.1}$$

where we set $\gamma = 1$ for simplicity. We prove the following theorem.

Theorem 1: The argument equation (2.1) has uniquely determined solutions of the form

$$\psi(x,t) = e^{i(\alpha_1(t)x^2 + \beta_1(t))}\, e^{-\alpha_2(t)x^2 - \beta_2(t)}$$

with

$$\int_{-\infty}^{\infty} |\psi(x,t)|^2\, dx = 1, \quad \alpha_2(t) > 0$$

and

$$\neq \text{const}(\forall\, t \in [0,\infty))$$

if

$$0 \leqslant \omega^2 \leqslant \tfrac{1}{16}\gamma^2. \tag{2.2}$$

The real functions $\alpha_1, \alpha_2, \beta_1, \beta_2$ are in $C^\infty[0,\infty)$.

Remark: The case $\alpha_2 = \text{const}$, $\alpha_1 = 0$ corresponds to the special solution of the form (1.7) for $n=0$. For Eq. (1.3) with $V = \frac{1}{2}m\omega^2 x^2$, condition (1.1) reads as $\omega^2 \leqslant \gamma^2/4$. If we go back from (1.3) to (1.2), we get solutions of (1.2) in the form

$$\psi(x,t) = e^{i\theta(t)} e^{i(\alpha_1(t)x^2 + \beta_1(t))}\, e^{-\alpha_2(t)x^2 - \beta_2(t)}$$

with

$$\Theta(t) = \gamma\left\{\int_0^t \alpha_1(s)\int_{-\infty}^{\infty} x^2 e^{-2\alpha_2(s)x^2 - 2\beta_2(s)}dx\, ds + \int_0^t \beta_1(s)ds\right\};$$

here, also,

$$\int_{-\infty}^{\infty} |\psi(x,t)|^2\, dx = 1.$$

Proof of Theorem 1: Observe $\gamma = 1$. If we make the "ansatz" $\psi(x,t) = \exp(i\sigma(x,t))$, $\sigma(x,t) = \alpha(t)x^2 + \beta(t)$, $\alpha = \alpha_1 + i\alpha_2$, $\beta = \beta_1 + i\beta_2$; α_j, β_j real, we get, from (10),

$$\sigma_t = i\sigma_{xx} - \sigma_x^2 - \omega^2 x^2 - \operatorname{Re}\sigma$$

and, therefore,

$$(\dot{\alpha} + \operatorname{Re}\alpha + 4\alpha^2 + \omega^2)x^2 + (\dot{\beta} + \operatorname{Re}\beta - 2i\alpha) = 0$$

(where the dot means $\partial/\partial t$). If we set the coefficients of x^2 and 1 in the last equation equal to zero and take real and imaginary parts, we get

$$\dot{\alpha}_1 + \alpha_1 + 4(\alpha_1^2 - \alpha_2^2) + \omega^2 = 0, \tag{2.3}$$

$$\dot{\alpha}_2 + 8\alpha_1\alpha_2 = 0, \tag{2.4}$$

$$\dot{\beta}_1 + \beta_1 + 2\alpha_2 = 0, \tag{2.5}$$

$$\dot{\beta}_2 - 2\alpha_1 = 0. \tag{2.6}$$

Evidently, every solution to the system (2.3)–(2.6) gives a desired solution of (2.1); obviously, (2.5) and (2.6) are solvable for all $t \geqslant 0$ if α_1, α_2 are known. By (2.4) we have

$$\alpha_2(t) = c\exp\left(-8\int_0^t \alpha_1(s)ds\right), \tag{2.7}$$

with $c = \alpha_2(0)$; from (2.7) we have $\alpha_2(t) > 0$, $\alpha_2(t) \neq \text{const}$ if c is chosen to be >0, α_1 is known, and $\neq 0$. If we insert (2.7) into (2.3) it follows that (2.3) and (2.4) is equivalent to the nonlinear integral equation

$$\begin{aligned}\alpha_1(t) &= \alpha_1(0)\exp\left(-\int_0^t [4\alpha_1(s)+1]ds\right) \\ &\quad + \int_0^t \exp\left(-\int_s^t [4\alpha_1(\tau)+1]d\tau\right) \\ &\quad \times \left\{4c^2\exp\left(-16\int_0^s \alpha_1(\tau)d\tau\right) - \omega^2\right\} ds\end{aligned} \tag{2.8}$$

or

$$\begin{aligned}\alpha_1(t) &= \alpha_1(0)\exp\left(-\int_0^t [4\alpha_1(s)+1]ds\right) \\ &\quad + 4c^2\exp\left(-\int_0^t [4\alpha_1(s)+1]ds\right) \\ &\quad \times \int_0^t \exp\left(-\int_0^s [12\alpha_1(\tau)-1]d\tau\right) ds \\ &\quad - \omega^2\int_0^t \exp\left(-\int_s^t [4\alpha_1(\tau)+1]d\tau\right) ds.\end{aligned} \tag{2.9}$$

Thus it is enough to show that (2.9) has a global solution $\alpha_1 \in C[0,\infty)$ [which is $\neq 0$ if $\alpha_1(0) \neq 0$]; from this we then have $\alpha_1, \alpha_2, \beta_1, \beta_2 \in C^\infty[0,\infty)$. Now it is clear that the right-hand side of (2.9) defines a contraction in $C[0,T]$ if T is small enough; thus (2.9) has a unique local solution in a small time interval by the contraction mapping principle, and the solution, wherever it exists, can be constructed as the uniform limit of the iteration sequence $(\tilde{\alpha}_n)$ defined by

$$\tilde{\alpha}_{n+1}(t) = \alpha_1(0)\exp\Big(-\int_0^t [4\tilde{\alpha}_n(s)+1]ds\Big) + 4c^2\exp\Big(-\int_0^t [4\tilde{\alpha}_n(s)+1]ds\Big) \times \int_0^t \exp\Big(-\int_0^s [12\tilde{\alpha}_n(\tau)-1]d\tau\Big)ds - \omega^2\int_0^t \exp\Big(-\int_s^t [4\tilde{\alpha}_n(\tau)+1]d\tau\Big)ds,$$

$$\tilde{\alpha}_0(t) = \alpha_1(0)\exp\Big(-\int_0^t [4\alpha_1(s)+1]ds\Big) = \frac{\alpha_1(0)}{4\alpha_1(0)+1}(1-e^{-t})$$

[if we choose $\alpha_1(0)>0$].

Lemma 1: Under the assumptions of Theorem 1 we have for all $t\geqslant 0$,

$$-\omega^* \leqslant \tilde{\alpha}_n(t) \leqslant c_1 e^{\alpha^* t}, \tag{2.10}$$

where

$$\omega^* = \tfrac{1}{8}(1-\sqrt{1-16\omega^2}), \quad \alpha^* = 16\omega^*,$$

$$c_1 = \alpha_1(0) + 4c^2/(12\omega^*+1).$$

Proof of Lemma 1 by induction: We have $0\leqslant\tilde{\alpha}_0(t)\leqslant\alpha_1(0)/(4\alpha_1(0)+1)\leqslant\alpha_1(0)\leqslant c_1e^{\alpha^*t}$. If (2.10) is valid for $\tilde{\alpha}_j$, $j=0,\dots,n$, we have (observe $\omega^*<\frac{1}{4}$)

$$\tilde{\alpha}_{n+1}(t) \leqslant \alpha_1(0)e^{-(1-4\omega^*)t} + \frac{4c^2}{12\omega^*+1}e^{(4\omega^*-1)t}[e^{(12\omega^*+1)t}-1] \leqslant \Big(\alpha_1(0)+\frac{4c^2}{12\omega^*+1}\Big)e^{16\omega^*t} = c_1e^{\alpha^*t},$$

$$\tilde{\alpha}_{n+1}(t) \geqslant -\omega^2\int_0^t \exp\Big(-\int_s^t [4\tilde{\alpha}_n(\tau)+1]d\tau\Big)ds \geqslant -\frac{\omega^2}{1-4\omega^*}e^{-(1-4\omega^*)t}[e^{(1-4\omega^*)t}-1] \geqslant -\frac{\omega^2}{1-4\omega^*} = -\omega^*,$$

since ω^* is a solution of $z^2-\frac{1}{4}z+\frac{1}{4}\omega^2=0$.

By Lemma 1 all local solutions of (2.9) or (2.3)–(2.6) can be extended globally in time to solutions on $[0,\infty)$.

To get $\int|\psi(x,t)|^2dx=1$, it suffices to consider the case $t=0$, since it is well known (see Sec. IV) that $\int|\psi(x,t)|^2dx=\text{const}$ for every H^2-solution of (1.3) [and of (1.1) and (1.2) also]. But for $t=0$, we have

$$\int_{-\infty}^{\infty}|\psi(x,0)|^2dx = e^{-2\beta_2(0)}\int_{-\infty}^{\infty}e^{-\alpha_2(0)x^2}dx = 1$$

if $\beta_2(0)$ is chosen appropriately.

III. NONEXISTENCE OF SOLITARY WAVE SOLUTIONS

The main result in this section is the following theorem:

Theorem 2: The Schrödinger–Langevin equation (1.3) has no nontrivial solitary wave type solutions of the form

$$\psi(x,t) = r(x+ct)\exp\{i\phi(x+dt)\}, \tag{3.1}$$

where

$$c^2+d^2>0,\ r\geqslant 0,\ r\in C^2(\mathbb{R})\cap L^2(\mathbb{R}),\ \phi\in C^2(\mathbb{R}),\ r,\phi \text{ real}$$

and

$V\in C^1(\mathbb{R})$.

Remark: Naturally for solutions of type (3.1), only the case $V=\text{const}$ is of interest; we have treated the nonconstant case only for completeness. At the end of this section we give some remarks on the stationary case $c=d=0$.

Before we start the proof of the theorem, we need some preparations. In Eq. (1.3) we consider only the case $\gamma=1$ for simplicity. Substituting (3.1) into (1.3), we arrive at

$$cr' + 2r'\phi' + \phi''r = 0, \tag{3.2}$$

$$d\phi'r - r'' + r(V(x)+\phi'^2+\phi) = 0, \tag{3.3}$$

and if we set $r=e^u$ (for $r>0$), we get

$$cu' + 2u'\phi' + \phi'' = 0, \tag{3.4}$$

$$d\phi' - u'' - u'^2 + V(x) + \phi'^2 + \phi = 0. \tag{3.5}$$

From (3.2)–(3.5), we can conclude that r, u, $\phi\in C^4(\mathbb{R})$ (under the assumptions of the theorem).

Lemma 2: Let $f(x)\geqslant 0$ for $x\geqslant a$ and $\lim_{x\to\infty}f(x)=\infty$. Then for any solution of

$$r''=f(x)r,\quad r\geqslant 0,\quad r\in C^2(\mathbb{R})\cap L^2(\mathbb{R}),$$

we have

(i) r is a bounded decreasing function for $x\geqslant a$,

(ii) $\lim_{x\to\infty}r(x)=0$.

Proof: For $x\geqslant a$ we have $r''(x)\geqslant 0$, hence $r(x)\geqslant r(y)+r'(y)(x-y)$ for all $x\geqslant y\geqslant a$. Because of $r\in L^2(\mathbb{R})$, this implies $r'(y)\leqslant 0$ for all $y\geqslant a$; thus we have (i). Now (ii) is trivial, as we know from (i) that $\lim_{x\to\infty}r(x)$ exists.

Remark: Actually we have $\lim_{|x|\to\infty}r(x)=0$ for any $r\in L^2(\mathbb{R})\cap C^1(\mathbb{R})$; see Benjamin, Bona, and Mahony.[11]

Lemma 3: If r is in $C^2(\mathbb{R})\cap L^2(\mathbb{R})$ and $r''=(ax+b)r$ for all x, then $r=0$ ($a,b=\text{const}$).

Proof: For $a=0$, the statement is trivial. For $a\neq 0$, every solution of $r''=(ax+b)r$ can be represented as

$$r(x) = c_1\text{Ai}((ax+b)/a^{2/3}) + c_2\text{Bi}((ax+b)/a^{2/3}) \tag{3.6}$$

with the Airy functions Ai and Bi (see Abramowitz–Stegun[12]). We may assume $a>0$ (otherwise we substitute $x\to -x$). Then by Lemma 2 we have that r is bounded for $x\geqslant x_0=-b/a$. Since Ai(z) is bounded then for $c_2\neq 0$, we would have that Bi(z) is bounded for big $z\geqslant 0$, a contradiction. Hence we must have $c_2=0$. If $c_1\neq 0$ (3.6) implies that Ai(z) is an L^2-function, which is not the case since we have[12]

$$\text{Ai}(z) = \pi^{-1/2}|z|^{-1/4}\sin(\zeta+\pi/4) + O(|z|^{-7/4})$$

for

$$z\leqslant 0,\quad |z|\to\infty,$$

where

$$\zeta = \tfrac{2}{3}|z|^{3/2}.$$

Hence Ai$(z)\in L^2(\mathbb{R})$ would imply that

$$\int_{z_0}^{\infty} \frac{\sin^2(\frac{2}{3}|z|^{3/2} + \frac{1}{4}\pi)}{|z|^{1/2}} dz$$

would converge. But this is not the case since, by Hölder's inequality,

$$\int_{z_0}^{\infty} \frac{\sin^2(\frac{2}{3}z^{3/2} + \frac{1}{4}\pi)}{z^{1/2}} dz$$
$$\geqslant c_1 \int_{z_1}^{\infty} \frac{\sin^2 x}{x^{3/2}} dx \geqslant c_1 \int_{z_1}^{\infty} \frac{|\sin x|}{x} dx \times \Big(\int_{z_1}^{\infty} \frac{dx}{x^{4/3}} = \infty\Big)^{-1}.$$

Lemma 4: If $r \in L^2(\mathbb{R})$, $r(x) > 0$ for all $x \in \mathbb{R}$, then we have $1/r \notin L^2(\mathbb{R})$ and

$$\int_0^x \frac{ds}{r^2(s)} \geqslant c_1 x^2, \quad \int_{-x}^0 \frac{ds}{r^2(s)} \geqslant c_1 x^2$$

for all $x \in \mathbb{R}$, $x \geqslant 0$, where c_1 is some constant > 0.

Proof: We have, by Hölder's inequality,

$$b - a = \int_a^b dx \leqslant \Big(\int_a^b r^2(x) dx\Big)^{1/2} \times \Big(\int_a^b \frac{dx}{r^2(x)}\Big)^{1/2} \leqslant c_2 \Big(\int_a^b \frac{dx}{r^2(x)}\Big)^{1/2}$$

and, therefore,

$$\int_a^b \frac{dx}{r^2(x)} \geqslant \Big(\frac{b-a}{c_2}\Big)^2,$$

from which the assertion follows.

Proof of Theorem 2:

(i) At first we may assume that $r(x) > 0$ for all $x \in \mathbb{R}$, for if $r(t_0) = 0$, we must have $r'(t_0) = 0$ since $r(x) \geqslant 0$. But then, by the uniqueness of solutions of equations of the form

$$r'' = f(x) r$$

[with $f(x) = d\phi' + \phi'^2 + \phi + V$ from Eq. (3.3) where we have set $t = 0$], we get $r = 0$.

(ii) Assume $d = 0$. Then (3.3) reads as

$$r'' = (V(x) + \phi'^2 + \phi) r. \tag{3.7}$$

If we differentiate (3.6) with respect to x and t, we find $r''' = (V(x) + \phi'^2 + \phi) r' + r\, d/dx(V(x) + \phi'^2 + \phi)$ and $cr''' = c(V(x) + \phi'^2 + \phi) r'$, hence $r(x) d/dx(V(x) + \phi'^2 + \phi) = 0$ and $V(x) + \phi'^2 + \phi = \text{const} = b$. By (3.7) this means $r'' = br$ and, hence, $r \in L^2$ only if $r = 0$.

(iii) Now let $d \neq 0$.

Lemma 5: There exist constants a, b, $\in \mathbb{R}$ such that

$$V(x) = a(1 - c/d) x + b,$$

and for $c \neq d$ we have

$$r'' = (\alpha x + \beta) r$$

with some constants $\alpha, \beta \in \mathbb{R}$.

Proof of the Lemma: If we differentiate (3.5) with respect to x and t and compare the results, we get

$$V'(x) = (1 - c/d)[u''' + 2u'u''];\tag{3.8}$$

thus for $c = d$ the result follows. If $c \neq d$, by differentiating (3.7) with respect to t one gets $u'''' + 2u''^2 + 2u'u''' = 0$, and by integrating with respect to $z = x + ct$, this gives

$$u''' + 2u'u'' = \alpha \tag{3.9}$$

and

$$u'' + u'^2 = \alpha z + \beta \tag{3.10}$$

with some constants α, β. From (3.8) and (3.9) we get the first, and by (3.10) and $r = e^u$, the second statement of the lemma.

By Lemmas 3 and 5, we see that the assertion of Theorem 2 is true for $c \neq d$. Now let $c = d \neq 0$. Then by Lemma 5, we must have $V(x) = \text{const} = b$, and using the transformation $\phi \to \phi - b$, we may assume $V(x) \equiv 0$.

Solving (3.2) for ϕ', one arrives at

$$\phi'(z) = A/r^2(z) - c/2 \quad (A \in \mathbb{R});$$

thus (3.3) reads as

$$r'' = f(z) r, \; f(z) = \frac{A^2}{r^4(z)} - \frac{c^2}{4} - \Big(\frac{c}{2}\Big) z + \int_0^z \frac{A}{r^2(s)} ds + b. \tag{3.11}$$

Now if $A = 0$, the result follows by Lemma 3. If $A > 0$, we have by Lemma 4 $\lim_{z\to\infty} f(z) = \infty$; hence by Lemma 2, $\lim_{z\to\infty} r(z) = 0$. But using (3.11) and Lemma 4, it follows that

$$\lim_{z\to\infty} r''(z) \geqslant \lim_{z\to\infty} r(z) \frac{A^2}{r^4(z)} = \infty,$$

from which we must conclude that $\lim_{z\to\infty} r(z) = \infty$, a contradiction. Along the same lines we get a contradiction in the case $A < 0$ by the substitution $z \to -z$.

Remark: In the stationary case ($c = d = 0$), we have $\phi'(z) = A/r^2(z)$ and $r'' = (V(x) + \phi'^2 + \phi) r$. If we assume an additional growth condition on the potential $V(x)$, namely,

$$\lim_{|x|\to\infty} \{V(x) + \alpha |x|^2\} = \infty \tag{3.12}$$

for every $\alpha > 0$, then the arguments used above lead to $A = 0$. Hence the stationary solutions of the argument equation (1.3) may be characterized as those solutions of

$$r'' = (V(x) + d) r \tag{3.13}$$

which lie in $L^2(\mathbb{R})$ (here d is a constant). The solvability of (3.12) in $L^2(\mathbb{R})$ naturally depends on the potential $V(x)$. For example, if $V(x) = ax + b$, Lemma 3 shows the nonexistence of L^2-solutions, whereas for $V(x) = x^2$, we have L^2-solutions of the argument equation, for instance $\psi(x) = \exp(-x^2/2 - i)$.

IV. CONCLUDING REMARKS

We close with some hints on open problems and some further remarks and properties of the equations discussed.

(a) The main open problem in connection with Eqs. (1.1), (1.2), or (1.3) is a reasonable global existence result for the Cauchy problem. The difficulty arises from the fact that the function $\psi \to \arg \psi$ is discontinuous if we consider it in the complex plane as a single-valued function. Nevertheless, equations of the argument type have some interesting monotonicity properties. Thus if we consider (1.3) with negative γ (choose $\gamma = -1$), and write the equation in the form

$$\psi_t + T\psi = 0,\ T\psi = i(-\partial_x^2\psi + V\psi - \arg\psi\psi),$$
$$(D(T) = \{\psi\in L^2(\mathbb{R}) | -\partial_x^2\psi + V\psi\in L^2(\mathbb{R})\}),$$

the operator T gives rise to a monotone operator $S = T + \lambda I$ in the sense of

$$\mathrm{Re}(S\psi - S\phi, \psi - \phi)\geqslant 0$$

$(\forall\phi, \psi\in D(T))$ if $\lambda > 1$ and $V=$ real. But evidently S is not maximal monotone, since T is not demicontinuous.

The monotonicity proof does not work for $\gamma > 0$, but if one cuts off the function arg z in a small sector with angle $\epsilon > 0$ around the positive real axis in such a way that the resulting function $\arg_\epsilon(z)$ is continuous for $z\neq 0$, $\arg_\epsilon(z)z$ is continuous on $\mathbb{C}$, and $\lim_{\epsilon\to 0}\arg_\epsilon(z) = \arg z$ (for all $z\in\mathbb{C}\setminus\{0\}$), then one can show along the same lines as above that the approximative operator

$$S_\epsilon = T_\epsilon + \lambda_\epsilon I$$

[with an appropriately chosen $\lambda_\epsilon > 0$ and $T_\epsilon\psi = i(-\partial_x^2\psi + V\psi + \arg_\epsilon(\psi)\psi)$, $V=$ real] is maximal monotone and coercitive. Thus the initial-value problem

$$i\psi_t = -\partial_x^2\psi + V\psi + \arg_\epsilon(\psi)\psi,$$
$$\psi(x,0) = \psi_0(x)$$

has a unique solution $\psi_\epsilon(x,t)$ in the sense of monotone operator theory (see H. Brézis[13]). These monotonicity properties will be treated more rigorously in a forthcoming paper.

(b) The following conservation laws for (1.3) (with $\gamma = 1$) are known in the case $V(x) = \omega^2x^2$:

$$\int_{-\infty}^{\infty} |\psi(x,t)|^2dx = \text{const}, \tag{4.1}$$

$$\frac{d}{dt}\int_{-\infty}^{\infty}\{|\psi_x(x,t)|^2 + \omega^2x^2|\psi(x,t)|^2\}dx = -2\int_{-\infty}^{\infty}\phi_x^2r^2\,dx, \tag{4.2}$$

where $\phi = \arg\psi$, $r = |\psi|$. This can be proved formally (and justified rigorously) by multiplying (1.3) with $\bar\psi$, $-\bar\psi_{xx}$, $x^2\bar\psi$, respectively, taking complex conjugates and adding up; after some partial integrations one arrives at (4.1) and (4.2).

(c) Mathematically it is not known whether any of the known special solutions of the equations discussed are stable in any sense.

(d) Although there are no L^2-solutions to (1.3) of solitary wave type, (1.3) has generalized solitary wave type solutions, namely, for instance, if we set $c = -c_0$, $c_0 > 0$, $\lambda = c_0^{1/3}$, and

$$r(z) = \begin{cases} N\cdot\mathrm{Ai}(\lambda z - a_1), & \text{for } z\geqslant 0,\\ 0, & \text{for } z\leqslant 0,\end{cases}$$
$$\phi(z) = (c_0/2)z + c_1,$$

where $c_1 = c_0^{2/3}a_1 + c_0^2/4$, and $-a_1$ is the first negative zero of Ai(z). Then $\psi(x,t) = r(x-c_0t)e^{i\phi(x-c_0t)}$ (where $N>0$ may be chosen to satisfy $\int|\psi|^2dx = 1$) is a generalized solitary wave solution which we may call pseudosolitary-wave; we have $\psi\in L^2(\mathbb{R})\cap C(\mathbb{R})$; ψ satisfies the argument equation (with $V=0$, $\gamma=1$) on $\mathbb{R}^2$ except on the line where $x = c_0t$; actually it is a distribution solution of

$$i\psi_t = -\partial_x^2\psi + \arg\psi\psi + \mu\delta$$

($\delta =$ Dirac distribution), where μ characterizes the jump of r' at $z = x - c_0t = 0$.

[1] J. Messer, Acta Phys. Austriaca **50**, 75–91 (1979).
[2] R. W. Hasse, J. Math. Phys. **16**, 2005–2011 (1975).
[3] H. Dekker, Phys. Rep. **80**, 1–112 (1981).
[4] I. R. Svinin, Theor. Math. Phys. (USSR) **22**, 67–75 (1975).
[5] I. R. Svinin, Theor. Math. Phys. (USSR) **27**, 478–483 (1976).
[6] J. J. Griffin and K. -K. Kan, Rev. Mod. Phys. **48**, 467–477 (1976).
[7] M. D. Kostin, J. Chem. Phys. **57**, 3589–3591 (1972).
[8] B. -S. K. Skagerstam, Phys. Lett. B **58**, 21–25 (1975).
[9] B. -S. K. Skagerstam, J. Math. Phys. **18**, 308–311 (1977).
[10] J. D. Immele, K. -K. Kan, and J. J. Griffin, Nucl. Phys. A **241**, 47–60 (1975).
[11] T. B. Benjamin, J. L. Bona, and J. J. Mahony, Philos. Trans. R. Soc. London, Ser. A **272**, 47–78 (1972).
[12] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[13] H. Brézis, *Opérateurs maximaux monotones*. Math. Studies Vol. 5 (North–Holland, Amsterdam, 1973).

Symmetry reduction for nonlinear relativistically invariant equations[a)]

A. M. Grundland,[b)] J. Harnad, and P. Winternitz
Centre de recherche de mathématiques appliquées, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 22 April 1983; accepted for publication 17 June 1983)

Symmetry reduction is studied for the relativistically invariant scalar partial differential equation $H(\Box u,(\nabla u)^2,u) = 0$ in $(n+1)$-dimensional Minkowski space $M(n,1)$. The introduction of k symmetry variables $\xi_1, \dots, \xi_k$ as invariants of a subgroup G of the Poincaré group $P(n,1)$, having generic orbits of codimension $k \leqslant n$ in $M(n,1)$, reduces the equation to a PDE in k variables. All codimension-1 symmetry variables in $M(n,1)$ (n arbitrary), reducing the equation studied to an ODE are found, as well as all codimension-2 and -3 variables for the low-dimensional cases $n = 2,3$. The type of equation studied includes many cases of physical interest, in particular nonlinear Klein–Gordon equations (such as the sine–Gordon equation) and Hamilton–Jacobi equations.

PACS numbers: 02.30.Jr, 02.20. + b, 11.30.Cp

1. INTRODUCTION

The purpose of this paper is a systematic study of symmetry reduction for a relativistically invariant first- or second-order nonlinear partial differential equation in $(n+1)$-dimensional Minkowski space $M(n,1)$, involving one scalar field. Such equations are of the general form

$$H(\Box u, (\nabla u)^2, u) = 0, \tag{1.1}$$

where H is some sufficiently smooth given function of the indicated variables and the dependent variable $u = u(x_0, x_1, \dots, x_n)$ is a scalar function of a point $\tilde{x}$ in $M(n,1)$. The Laplace–Beltrami (or d'Alembert) and the gradient square operators are respectively

$$\begin{aligned}\Box u &= u_{x_0x_0} - u_{x_1x_1} - \cdots - u_{x_nx_n},\\ (\nabla u)^2 &= (u_{x_0})^2 - (u_{x_1})^2 - \cdots - (u_{x_n})^2.\end{aligned} \tag{1.2}$$

By reduction we mean the introduction of new independent variables

$$\xi_1(\tilde{x}), \dots, \xi_k(\tilde{x}), \quad 1 \leqslant k \leqslant n \tag{1.3}$$

such that the Ansatz

$$u(\tilde{x}) = u(\xi_1, \dots, \xi_k) \tag{1.4}$$

reduces Eq. (1.1) to a k-dimensional partial differential equation involving the variables ξ_i only, or in particular, for $k = 1$, to an ordinary differential equation.

The main result of this paper is that the variables $\xi_i(\tilde{x})$ can be systematically generated as "symmetry variables," i.e., invariants of subgroups G of the symmetry group of the equation, in this case, the Poincaré group $P(n,1)$ in $n+1$ dimensions. More specifically, subgroups G having generic orbits of codimension k in $M(n,1)$, will have k functionally independent invariants $\xi_1, \dots, \xi_k$, depending on the coordinates of a point $\tilde{x} \in M(n,1)$. These invariants can serve as the coordinates (1.3), (1.4) and will provide the desired dimensional reduction.

Equation (1.1) is sufficiently general to encompass many cases of considerable physical interest. For example, nonlinear Klein–Gordon equations

$$\Box u = F(u,(\nabla u)^2) \tag{1.5}$$

occur in many physical problems with a variety of nonlinearities on the right hand side. As obvious examples, let us mention the sine–Gordon equation with $F = \sin u$ (arising in studies of Josephson junctions, self-induced transparency, in relation to the Thirring model and nonlinear σ model in elementary particle theory and elsewhere), the Liouville equation with $F = e^u$, various field theories in which F is a polynomial in u, and many others.[1–6] The relativistically invariant Hamilton–Jacobi equation

$$(\nabla u)^2 + V(u) = E, \tag{1.6}$$

figuring in relativistic classical mechanics and general relativity theory is also of this form. [A potential $V(u)$ depending on the action u but not directly on the space-time coordinates arises in the eikonal approximation of the nonlinear Klein–Gordon equations.]

The motivation for reducing an equation to one in fewer dimensions is quite obvious: Ordinary differential equations are usually easier to solve than partial differential equations, lower-dimensional PDE's are easier to treat than higher-dimensional ones. The reduction to ODE's is of particular interest since these can very often be explicitly integrated in terms of known functions, or at least their singularity structure can be investigated using well-known methods. In particular, it is possible to determine whether these ODE's are of the Painlevé type,[7–10] i.e., whether all their critical points are fixed (independent of the initial conditions). This is of interest both from the viewpoint of solving these ODE's and from that of the "Painlevé conjecture,"[3,11] namely, that a PDE is integrable by inverse scattering techniques only if all ODE's that it reduces to are of the Painlevé type.

The solutions obtained by symmetry reduction are of interest in their own right, and they can in some cases also be used to generate more general solutions of the original PDE, e.g., as input functions in Bäcklund transformations.[2–6]

The idea of symmetry reduction is a very old one and goes back at least to Lie. (It is useful to recall that the theory

---

[a)] Supported in part by the Natural Sciences and Engineering Research Council of Canada and by an FCAC Grant of the "Gouvernement du Québec."

[b)] On leave from the Department of Physics, Institute of Geophysics, Warsaw University, Warsaw, Poland.

of Lie groups was originally developed in connection with the theory of differential equations.[12–14]) More recently, this subject has been treated in several books[15–17] and is a standard technique for obtaining solutions of nonlinear equations, e.g., the sine–Gordon equation,[18] the nonlinear Schrödinger equation,[19,20] the Yang–Mills equations,[21–23] and, of course, the Einstein equations of general relativity.[24–26]

Our contribution is that we make a systematic use of the subgroup structure of the invariance group of the equation to generate, in some sense, all symmetry variables providing the desired reduction. In the case of Eq. (1.1) the symmetry group is the Poincaré group $P(n,1)$ (the inhomogeneous Lorentz group) in $n+1$ dimensions. (This is true for arbitrary functions H, but in special cases the symmetry group may be larger.)

In Sec. 2, we formulate the problem and show how reduction of dimensions is obtained by introducing the appropriate symmetry variables $\xi_1, \ldots, \xi_k$. In Sec. 3 we find the invariants of all subgroups of $P(n,1)$ having orbits of codimension 1 and hence reducing Eq. (1.1) to an ODE. Section 4 is devoted to the physically interesting cases of $(2+1)$- and $(3+1)$-dimensional Minkowski space, and we obtain all possible symmetry reductions to lower dimensional PDE's (as well as ODE's). The obtained ODE's are investigated for special cases of interest in Sec. 5, where we also discuss some of the lower order PDE's. The conclusions are summarized in Sec. 6.

2. FORMULATION OF THE PROBLEM OF SYMMETRY REDUCTION

Let us consider Eq. (1.1),

$$H(\Box u, (\nabla u)^2, u) = 0,$$

with $\Box$ and $(\nabla\cdot)^2$ as in (1.2). This equation is invariant under the Poincaré group $P(n,1)$ (the inhomogeneous Lorentz group), i.e., if $u(x)$ is a solution of (1.1), then so is

$$[T(g)u](x) = u(x') = u(\Lambda x + a), \tag{2.1}$$

where $g = (\Lambda, a) \in P(n,1)$ is a Poincaré transformation,

$$a \in \mathbb{R}^{(n+1)}, \quad \Lambda \in \mathbb{R}^{(n+1)\times(n+1)},$$
$$\Lambda I_{1,n} \Lambda^T = I_{1,n}, \quad I_{1,n} = \begin{pmatrix} 1 & \\ & -I_n \end{pmatrix} \tag{2.2}$$

(the superscript T denotes matrix transposition).

A basis for the Lie algebra $p(n,1)$ of the Lie group $P(n,1)$ is given by the infinitesimal Lorentz transformations $M_{\mu\nu}$ and infinitesimal translations P_μ, satisfying the commutation relations

$$\begin{aligned}
&[M_{\mu\nu}, M_{\rho\sigma}] = g_{\mu\rho} M_{\nu\sigma} - g_{\nu\rho} M_{\mu\sigma} + g_{\mu\sigma} M_{\rho\nu} - g_{\nu\sigma} M_{\rho\mu},\\
&[M_{\mu\nu}, P_\sigma] = -g_{\nu\sigma} P_\mu + g_{\mu\sigma} P_\nu, \qquad (2.3)\\
&[P_\mu, P_\nu] = 0,\\
&\mu,\nu = 0,1,\ldots,n, \quad g_{00} = 1, \quad g_{aa} = -1, \quad a = 1,\ldots,n,\\
&\quad g_{\mu\nu} = 0 \quad \text{for } \mu \neq \nu.
\end{aligned}$$

When acting on scalar functions $u(\tilde{x})$, $\tilde{x} \in M(n,1)$, the infinitesimal operators (2.3) can be represented by the following differential operators:

$$\begin{aligned}
&M_{ab} = x_a \partial_b - x_b \partial_a \quad \text{(rotations)},\\
&M_{0a} = -x_0 \partial_a - x_a \partial_0 \quad \text{(Lorentz boosts)},\\
&P_\mu = \partial_\mu \quad \text{(translations)}, \qquad (2.4)\\
&a,b = 1,\ldots,n, \quad \mu = 0,1,\ldots,n, \quad \partial_\mu \equiv \frac{\partial}{\partial x^\mu}.
\end{aligned}$$

In particular, we shall call the operators

$$X_a = M_{0a} - M_{1a} = -(x_0 + x_1)\partial_a - x_a(\partial_0 - \partial_1), \quad a = 2,\ldots,n, \tag{2.5}$$

"light cone translations," and they play an important role.

Let us now turn to the question of reduction for Eq. (1.1). We introduce the variables $\xi_1(\tilde{x}), \ldots, \xi_k(\tilde{x})$, $1 \leqslant k \leqslant n-1$, as in (1.3). The conditions imposed on $\xi_i(\tilde{x})$ are expressed in the following lemma.

Lemma 1: The Ansatz $u(\tilde{x}) = u(\xi_1, \ldots, \xi_k)$ will reduce the equation $H(\Box u, (\nabla u)^2, u) = 0$ to a PDE in the $1 \leqslant k \leqslant n$ variables $\xi_i(x_0, x_1, \ldots, x_n)$ if and only if these variables satisfy partial differential equations of the form

$$\begin{aligned}
&\Box \xi_a = \alpha_a(\xi_1, \ldots, \xi_k),\\
&(\nabla\xi_a, \nabla\xi_b) = \beta_{ab}(\xi_1, \ldots, \xi_k),
\end{aligned} \quad 1 \leqslant a,b \leqslant k, \tag{2.6}$$

where α_a and β_{ab} are arbitrary functions of the indicated variables. In other words, the functions $\Box\xi_a$ and $(\nabla\xi_a, \nabla\xi_b)$ are constants on level sets of $\{\xi_1, \ldots, \xi_k\}$.

Proof: Substitute $u(\xi_1, \ldots, \xi_k)$ into (1.1). We have

$$\begin{aligned}
&\Box u = \sum_{a,b=1}^{k} u_{\xi_a\xi_b}(\nabla\xi_a, \nabla\xi_b) + \sum_{a=1}^{k} u_{\xi_a}\Box\xi_a,\\
&(\nabla u)^2 = \sum_{a,b=1}^{k} u_{\xi_a} u_{\xi_b}(\nabla\xi_a, \nabla\xi_b).
\end{aligned} \tag{2.7}$$

In (2.7) the derivatives u_{ξ_a} and $u_{\xi_a\xi_b}$ depend on $\xi_1, \ldots, \xi_k$ only; hence $\Box u$ and $(\nabla u)^2$ will depend on these variables alone if and only if (2.6) is satisfied.

Comments: (1) If $\xi_1, \ldots, \xi_k$ satisfy (2.6), then $\xi'_a = \phi_a(\xi_1, \ldots, \xi_k)$ with $\det(\partial\phi_a/\partial\xi_b) \neq 0$ satisfy the same type of equations (with different α_a and β_{ab}). The functions ξ_a are arbitrary sufficiently smooth functions, and this arbitrariness can be used to simplify Eq. (2.6).

(2) If $k = 1$, Eq. (1.1) is reduced to an ODE. In this case (2.6) can be simplified by an appropriate choice of ϕ in $\xi = \phi(\xi_1)$ to the form

$$\Box\xi = \alpha(\xi), \quad (\nabla\xi)^2 = \kappa, \quad \kappa = 0, 1, \text{ or } -1. \tag{2.8}$$

The problem of reduction of the number of independent variables for Eq. (1.1) has thus been reduced to the problem of generating solutions of Eq. (2.6) [or in particular (2.8)].

The method of symmetry reduction for Eq. (1.1) generates functions $\{\xi_i(x)\}$ satisfying Eq. (2.6) in a simple and straightforward manner. Consider a subgroup of the Poincaré group $G \subset P(n,1)$ having generic orbits of codimension k in Minkowski space $M(n,1)$. Such a subgroup will have k independent invariants in $M(n,1)$, i.e., k functionally independent functions of $\tilde{x} = (x_0, x_1, \ldots, x_n)$ satisfying

$$\xi_a(\tilde{x}) = \xi_a(g\tilde{x}), \quad \text{for } g \in G, \ a = 1, \ldots, k. \tag{2.9}$$

These invariants can be obtained as solutions of a system of first-order linear partial-differential equations. To obtain these equations, consider the Lie algebra L of the Lie group G. Let $\{X_1, \ldots, X_m\}$ be a set of operators providing a basis of L. Each X_i is thus some linear combination of the operators (2.4). The condition that some function $\psi(x_0, x_1, \ldots, x_n)$ be invariant under the action of G is equivalent to the requirement that $\psi(x_0, x_1, \ldots, x_n)$ satisfy

$$X_i\psi(x_0, x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, m. \tag{2.10}$$

The general solution of (2.10) will be an arbitrary function of k functionally independent solutions of (2.10),

$$\psi = \phi(\xi_1, \xi_2, \ldots, \xi_k), \tag{2.11}$$

where k is the codimension of the generic orbits of G in $M(n,1)$.

We can now present a theorem on symmetry reduction.

Theorem 1: Let $\xi_a(x_0, x_1, \ldots, x_n), a = 1, \ldots, k$, be a set of k functionally independent invariants of a subgroup G of the Poincaré group $P(n,1)$, having generic orbits of codimension k, i.e., k functionally independent solutions of Eq. (2.10) for $X_i \in L \equiv LG$. The Ansatz $u = u(\xi_1, \ldots, \xi_k)$ will then reduce the studied Eq. (1.1) to a partial differential equation in $\xi_1, \ldots, \xi_k$ (i.e., an ODE if $k = 1$).

Proof: We must prove that the variables ξ_i, $i = 1, \ldots, k$, obtained as solutions of (2.10), will satisfy Eqs. (2.6) of Lemma 1. The operators $\Box$ and $(\nabla\cdot,\nabla\cdot)$ are invariant under $P(n,1)$ [and, in particular, under $G \subset P(n,1)$], hence

$$\begin{aligned} X\Box\xi_a &= \Box(X\xi_a) = 0, \\ X(\nabla\xi_a,\nabla\xi_b) &= (\nabla X\xi_a,\nabla\xi_b) + (\nabla\xi_a,\nabla X\xi_b) \end{aligned} \tag{2.12}$$

since $X\xi_a = 0$ for all $X \in LG$. The group G, by assumption, has generic orbits of codimension k in $M(n,1)$. The general solution of (2.10) is a function of $\xi_1, \ldots, \xi_k$; hence (2.12) implies that $\Box\xi$ and $(\nabla\xi_a,\nabla\xi_b)$ are functions of $\xi_1, \ldots, \xi_k$, as required by (2.6).

Definition: We shall call variables ξ_a, $a = 1, \ldots, k$ that are invariants of a subgroup $G \subset P(n,1)$ with generic orbits of codimension k in $M(n,1)$ "*codimension-k symmetry variables.*"

We should stress that we do not claim that the introduction of codimension k symmetry variables is the only way to reduce (1.1) to a k-dimensional PDE. A more general formulation of symmetry reduction involving transformations of both the dependent and independent variables could be used. Besides this, we shall give examples below of "degenerate codimension k symmetry variables" that reduce (1.1) to a PDE in k-1 dimensions. There is, however, a systematic way of performing symmetry reduction in this limited sense for Eq. (1.1). We simply let G run through all subgroups of $P(n,1)$ that have generic orbits of the required dimension, calculate their invariants ξ_a, and use the Ansatz $u = u(\xi_1, \ldots, \xi_k)$ to reduce the equation.

For low-dimensional cases all subgroups of the Poincaré group are known [for $P(2,1)$ and $P(3,1)$].[27–30] A complete subgroup classification for $n \geqslant 4$ is a formidable task. Luckily this is not necessary since the relation between groups and their orbits is not one-to-one. What is needed are the maximal subgroups G_M that have a given type of orbits, i.e., a given set of invariants $\{\xi_1, \ldots, \xi_k\}$.

The most interesting case of symmetry reduction is reduction to an ODE, involving orbits of codimension 1. In Sec. 3 we shall solve this problem completely for general $M(n,1)$ Minkowski spaces; i.e., we find all codimension-1 symmetry variables.

3. SYMMETRY REDUCTION TO ORDINARY DIFFERENTIAL EQUATIONS

In this section we shall derive a complete list of all codimension 1 symmetry variables in Minkowski space $M(n,1)$ for arbitrary n. As a by-product we also find all codimension-1 symmetry variables for the Euclidean space $E(n)$. In more technical terms, we derive a complete list of differentiable invariants determining hypersurface foliations in $E(n)$ and $M(n,1)$, whose leaves are the generic orbits of an isometry subgroup. We are concerned with identifying the generic orbits of these subgroups, or equivalently their differentiable invariants. In the process we identify, for each set of such invariants, the maximal subgroup of the isometry group that leaves the set invariant. The results are of local nature which allows us to work completely on the level of Lie algebras, rather than Lie groups.

The following notational conventions will be used. Euclidean n space $E(n)$ will be identified with $\mathbb{R}^n$ with typical element $\mathbf{x} = (x_1, \ldots, x_n)$ and metric $(\mathbf{x},\mathbf{y}) = \Sigma_{i=1}^n x_i y_i$. The corresponding Euclidean isometry group will be denoted

$$P(n,0) = \mathrm{O}(n) \circledS T_n,$$

signifying the semidirect product of the orthogonal group $\mathrm{O}(n)$ with the translation group T_n. The corresponding Lie algebra is denoted $p(n,0) = \mathrm{o}(n) \oplus t_n$.

Similarly, $(n+1)$-dimensional Minkowski space $M(n,1)$ is identified with $\mathbb{R}^{n+1}$ with typical element $\tilde{x} = (x_0, \mathbf{x})$, $x_0 \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^n$, and Minkowski metric $(\tilde{x},\tilde{y}) = x_0 y_0 - \mathbf{xy}$. The isometry group is the Poincaré group $P(n,1) = \mathrm{O}(n,1) \circledS T_{n,1}$ with Lie algebra $p(n,1) \oplus t_{n,1}$; a standard basis for $p(n,1)$ is given by (2.3), (2.4).

The principal results of this section are summarized in the following two theorems, concerning codimension-1 symmetry variables in Euclidean and Minkowski spaces, respectively.

Theorem 2 (Euclidean case): Let $G \subset P(n,0)$ be a Lie group with generic orbits of dimension $n-1$ (hypersurfaces) in $E(n)$. These orbits are the level sets of an invariant of the type

$$\xi_{(V_k,\mathbf{a})}(\mathbf{x}) = \|P_{V_k}(\mathbf{x} - \mathbf{a})\|^2, \tag{3.1}$$

where P_{V_k} denotes orthogonal projection onto a fixed k-dimensional subspace V_k $(1 \leqslant k \leqslant n)$ and $\mathbf{a} \in \mathbb{R}^n$ denotes a translated origin. The orbits are hyperplanes for $k = 1$, spherical cylinders for $2 \leqslant k \leqslant n-1$ and spheres for $k = n$.

Comment: We can choose an orthonormal basis $(\mathbf{A}_1, \ldots, \mathbf{A}_k)$, $(\mathbf{A}_i, \mathbf{A}_k) = \delta_{ik}$ for the space V_k. The invariant (3.1) can then be written as

TABLE I. Codimension 1 symmetry variables for the Euclidean space $E(n)$.

No.	Standards variable ξ_s	General variable ξ	Invariance group G	$(\nabla\xi)^2$	$\Box\xi$
1	x_1	$(\mathbf{x}-\mathbf{a}, \mathbf{A}), \mathbf{A}^2 = 1$	$P(n-1,0)$	1	0
2	$(x_1^2+\cdots+x_k^2)^{1/2}$ $2\leqslant k\leqslant n-1$	$[(\mathbf{x}-\mathbf{a}, \mathbf{A}_1)^2+\cdots+(\mathbf{x}-\mathbf{a}, \mathbf{A}_k)^2]^{1/2}$	$O(k)\otimes P(n-k,0)$	1	$(k-1)/\xi$
3	$(x_1^2+\cdots+x_n^2)^{1/2}$	$[(\mathbf{x}-\mathbf{a}, \mathbf{A}_1)^2+\cdots+(\mathbf{x}-\mathbf{a}, \mathbf{A}_n)^2]^{1/2}$	$O(n)$	1	$(n-1)\xi$

$$\xi_k(\mathbf{x}) = \left[\sum_{a=1}^{k}(\mathbf{x}-\mathbf{a},\mathbf{A}_a)^2\right]^{1/2}, \quad (\mathbf{A}_a,\mathbf{A}_b) = \delta_{ab},$$
$$a,b = 1,\ldots,k, \quad 1\leqslant k\leqslant n. \tag{3.1'}$$

By a Euclidean transformation we can simplify the variable (3.1') to a "standard" form, namely,

$$\xi^s_k(\mathbf{x}) = (x_1^2 + \ldots + x_k^2)^{1/2}, \quad k = 1,\ldots,n. \tag{3.2}$$

Each ξ^s_k represents a $P(n,0)$ conjugacy class of codimension-1 symmetry variables.

We summarize the results on codimension-1 symmetry variables for the Euclidean space $E(n)$ in Table I. We list the standard variables $\xi_s = \xi^s_k$, the general variable $\xi = \xi_k$, the maximal group $G\subset P(n,0)$, leaving ξ invariant and also the values of

$$\Box\xi = \left(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}\right)\xi = \frac{k-1}{\xi}, \quad 1\leqslant k\leqslant n,$$
$$(\nabla\xi)^2 = \left(\frac{\partial\xi}{\partial x_1}\right)^2 + \cdots + \left(\frac{\partial\xi}{\partial x_n}\right)^2 = 1. \tag{3.3}$$

The Euclidean invariant equation

$$H(\Box u,(\nabla u)^2,u) = 0 \tag{3.4}$$

reduces to the ODE

$$H(u_{\xi\xi} + [(k-1)/\xi\,]U_\xi,u_\xi^2,u) = 0, \quad 1\leqslant k\leqslant n. \tag{3.5}$$

The proof of Theorem 2 will be given below. The results for Minkowski space $M(n,1)$ are less obvious.

Theorem 3 (Minkowski space): Let $G\subset P(n,1)$ be a Lie group with generic orbits of dimension n (hypersurfaces) in $M(n,1)$. These orbits are then the level sets of one of the following types of invariants:

(i) $\xi_{V_k,\tilde a}(\tilde x) = \|P_{V_k}(\tilde x - \tilde a)\|$,

where P_{V_k} denotes the orthogonal projection onto a fixed k-dimensional subspace V_k $(1\leqslant k\leqslant n+1)$ in which the induced metric is of maximal rank and $\tilde a\in\mathbb{R}^{n+1}$ is a translated origin;

(ii) $\xi^0_{\tilde N,\mathbf{a}}(\tilde x) = (\tilde N,\tilde x - \tilde a)$;

(iii) $\xi_{\tilde A,\tilde N,\tilde a}(\tilde x) = (\tilde A,\tilde x - \tilde a) + (\tilde N,\tilde x - \tilde a)^2$;

(iv) $\xi^L_{(\tilde A,\tilde N,\tilde a)}(\tilde x) = (\tilde A,\tilde x - \tilde a) + a\ln(\tilde N,\tilde x - \tilde a), \quad a\neq 0.$

In (i), ... ,(iv), we have $\tilde N, \tilde A,\tilde a\in\mathbb{R}^{n+1}$, $a\in\mathbb{R}$,

$$(\tilde N,\tilde N) = 0, \quad (\tilde A,\tilde A) = -1, \quad (\tilde N,\tilde A) = 0.$$

Comments: (1) Case (i) is the analog of the Euclidean invariants of Theorem 2 and includes hyperboloids, hyperbolic and spherical cylinders, and hyperplanes. Case (ii) is a hyperplane determined by an isotropic (lightlike) vector $\tilde N$. Cases (iii) and (iv) have no Euclidean analog.

(2) We can choose an orthonormal basis for V_k to make the invariant (i) more explicit. Each class of invariants is invariant under the Poincaré group $P(n,1)$ so we can again use "standard" invariants ξ^s to represent each conjugacy class. In this case the ξ^s can be chosen as follows:

$$r_k = (x_1^2 + \cdots + x_{k+1}^2)^{1/2} \quad \text{or}$$
$$\tau_k = (x_0^2 - x_1^2 - \cdots - x_k^2)^{1/2}, \tag{3.6a}$$
$$\eta = x_0 + x_1, \tag{3.6b}$$
$$\zeta = x_2 + \tfrac14(x_0 + x_1)^2, \tag{3.6c}$$
$$\sigma = x_2 + a\ln(x_0 + x_1), \quad a\neq 0. \tag{3.6d}$$

The results on codimension-1 symmetry variables in $M(n,1)$ are summarized in Table II. In column 4 we give the Lie algebra of the Lie group $G\subset P(n,1)$, leaving the corresponding standard variable ξ_s invariant. The Poincaré-invariant Eq. (1.1) reduces to the ODE

$$H(\kappa[u_{\xi\xi} + (k/\xi)u_\xi],\kappa u_\xi^2,u) = 0,$$
$$\kappa = 0,\pm 1, \; k = 0,1,\ldots,n, \tag{3.7}$$

with k and κ given in Table II for each case.

To prove these theorems, a number of preparatory results will first be established. The translation algebra t may be identified with the space $E(n)$ or $M(n,1)$ itself. Under this identification, the adjoint action of $O(n)$ [resp. $O(n,1)$] is equivariant to the adjoint action on t. We shall denote the subalgebra corresponding to a subspace $E\subset E(n)$ [or $E\subset M(n,1)$] as $t^E\subset t$ and speak of the Euclidean or Minkowski metric on t and the induced bilinear form on t^E. Let L denote the Lie algebra of $G\subset P(n,0)$ or $G\subset P(n,1)$ and denote by

$$t_L \equiv L\cap t < L \tag{3.8}$$

the ideal of "pure" translations. The following lemma allows us to reduce the problem of determining invariants to the case where t_L is totally isotropic (i.e., the induced bilinear form vanishes); that is, for $E(n)$, $t_L = 0$ and, for $M(n,1)$, $\dim t_L = 0$ or 1. To avoid repetition, when results are equally applicable to $E(n)$ and $M(n,1)$, we shall denote both spaces by M.

Lemma 2: There exists an orthogonal projection $P_E{:}M\to E$ and a corresponding projected algebra L_p contained in the infinitesimal isometry algebra for the projected metric such that the L-invariant functions on M are those of the form $\phi\circ P_E$, where ϕ is a L_p-invariant function on E and t_{L_p} is totally isotropic.

TABLE II. Codimension 1 symmetry variables for the Minkowski space $M(n, 1)$.

No.	Standard variable ξ_s	General variable ξ	Invariance algebra L	$(\nabla\xi)^2$	$\Box\xi$
1	x_0	$(\tilde{A},\tilde{x}),\quad \tilde{A}^2=1$	$M_{ik}, P_i,\quad i,k=1,\dots,n$	1	0
2	x_1	$(\tilde{A},\tilde{x}),\quad \tilde{A}^2=-1$	$M_{\alpha\beta}, P_\alpha,\quad \alpha,\beta=2,\dots,n,0$	-1	0
3	x_0+x_1	$(\tilde{A},\tilde{x}),\quad \tilde{A}^2=0$	$M_{ab}, M_{0a}-M_{1a}, P_a, P_0-P_1,\quad a,b=2,\dots,n$	0	0
4	$x_2+\mu\ln(x_0+x_1)$	$(\tilde{A},\tilde{x})+\mu\ln(\tilde{B},\tilde{x}),\quad \mu\neq 0$ $\tilde{A}^2=-1,\quad (\tilde{A},\tilde{B})=0,\quad \tilde{B}^2=0$	$M_{ab}, M_{0a}-M_{1a}, P_a, M_{01}+\mu P_2, P_0-P_1,\quad a,b=3,\dots,n$	-1	0
5	$x_2+\frac{1}{4}(x_0+x_1)^2$	$(\tilde{A},\tilde{x})+(\tilde{B},\tilde{x})^2$	$M_{ab}, M_{0a}-M_{1a}, P_a, M_{02}-M_{12}+P_0+P_1, P_0-P_1,\quad a,b=3,\dots,n$	-1	0
6_k	$(x_1^2+\cdots+x_{k+1}^2)^{1/2}$ $k=2,\dots,n-1$	$\left[\sum_{a=1}^{k}(\tilde{x}-\tilde{B},\tilde{A}_a)^2\right]^{1/2}$ $(\tilde{A}_a,\tilde{A}_b)=-\delta_{ab}$, $a,b=1,\dots,k+1$	$M_{\alpha\beta}, P_\alpha, M_{ab},\quad \alpha,\beta=k+2,\dots,n,0,\quad a,b=1,\dots,k+1$	-1	$-k/\xi$
7_k	$(x_0^2-x_1^2-\cdots-x_k^2)^{1/2}$	$\left[\sum_{\alpha=0}^{k}g_{\alpha\alpha}(\tilde{x}-\tilde{B},\tilde{A}_\alpha)^2\right]^{1/2}$ $(\tilde{A}_\alpha,\tilde{A}_\beta)=g_{\alpha\beta}$, $\alpha,\beta=0,1,\dots,k$	$M_{ab}, P_a, M_{\alpha\beta},\quad a,b=k+1,\dots,n,\quad \alpha,\beta=0,1,\dots,k$	1	k/ξ

Proof: Suppose t_L is not totally isotropic. Decompose it into a direct sum:

$$t_L = t_N + t^F,$$

where t_N is the null space of the induced bilinear form on t_L and t^F is an arbitrarily chosen complement corresponding to a subspace $F \subset M$. Define $E_1 \equiv F^\perp$ and denote orthogonal projection relative to F by

$$P_{E_1}: M \to E_1.$$

For each element $X \in L$, regarded as a vector field on M, we define a projection X_p onto E by

$$X_p(q) \equiv P_{E_1} X(q), \quad q \in E_1,$$

where the value $X(q)$ at q is identified with an element of the space itself. Then X_p is an infinitesimal isometry for the induced metric on E_1. Any L-invariant function on M is in particular F-invariant and hence of the form $\phi \circ P_E$, where ϕ is a function on E_1. Since it is X-invariant, we also have

$$X(\phi \circ P_{E_1})_q = P_{E_1} X(q) \phi \,|_q = 0.$$

Hence ϕ is $P_{E_1} X$ invariant for all $X \in L$, and therefore L_P invariant where $L_P \equiv \{P_{E_1} X, X \in L\}$ is the algebra generated by all $P_{E_1} X(q)$. Conversely, given such an L_P-invariant function ϕ, $\phi \circ P_{E_1}$ is annihilated by all $X(q)$, for $q \in E_1 \subset M$. Since an arbitrary point $p \in M$ may be obtained by a translation T parallel to F, translating $X(p)$ to a point q in E_1 gives another vector $\tilde{X}(q) = \mathrm{Ad}\, TX(q)$ which annihilates $\phi \circ P_{E_1}$.

Now, if the projected algebra L_p has vanishing translational part t_{L_p}, then $E = E_1$ and the result is established. If not, the procedure is repeated with respect to (L_p, E_1) and so on until it terminates.

Lemma 2 reduces the problem of finding all codimension-1 symmetry variables in $M(n,1)$ to that of finding the invariants of all subgroups G of $P(k,1)$ in $M(k,1)$ $(k = 0,1, \dots ,n)$, involving either no free translations at all or a one-dimensional subgroup of lightlike translations.

The remaining lemmas deal with the case $t_L = 0$, which always holds in the Euclidean case and may hold in the Minkowski case. The other possibility, with $\dim\, t_L = 1$ for the Minkowski case will be treated separately. We define the projection $\pi: L \to \mathrm{o}(n)$ or $\mathrm{o}(n,1)$ by the value of the first component in the semidirect sum decomposition of $p(n,0)$ or $p(n,1)$. Then π is a Lie algebra homomorphism and its image defines a subalgebra

$$L_0 \equiv \mathrm{Im}\, \pi \subset \mathrm{o}(n) \ \text{ or } \ \mathrm{o}(n,1). \tag{3.9}$$

Under the assumption $t_L = \ker\ \pi = 0$, the map is 1–1 and hence its inverse on L_0

$$\pi^{-1}: L_0 \to L_1$$

determines a linear map

$$\tau: L \to t$$

such that

$$\pi^{-1}(X) = X + \tau(X), \quad X \in L_0,$$
$$\tau[X,Y] = [\tau(X),Y] + [X,\tau(Y)]. \tag{3.10}$$

Such a map is referred to in representation theory as a cocycle. Any two algebras $L, \tilde{L}$ related by conjugation

$$L = e^{Y}\tilde{L}e^{-Y}, \quad Y \in t,$$

have the same image under

$$\pi(L) = \pi(\tilde{L}) = L_0$$

and have their cocyles related by

$$\tilde{\tau}(X) = \tau(X) + [Y,X], \quad X \in L_0. \tag{3.11}$$

Any two cocyles $\tau,\tilde{\tau}$ so related are cohomologous, the additive group of equivalence classes $[\tau]$ under the relation (3.11) forming the first cohomology group

$$H^1(L_0,t) = \{[\tau]\}.$$

The coboundaries are the trivial cocycles τ_Y of the form

$$\tau_Y(X) = [Y,X], \tag{3.12}$$

which are cohomologous to zero.

We may decompose L_0 according to Levi's theorem into a semidirect sum

$$L_0 = S \oplus R, \tag{3.13}$$

where S is semisimple and R is the radical (maximal solvable ideal). A particular case of Whitehead's theorem on Lie algebra cohomology implies[31]:

Lemma 3 (Whitehead): The restriction $\tau|_S$ is cohomologous to zero.

A general result concerning the structure of Lie algebras[32] implies that the commutator $[L_0,R]$ contains only nilpotent matrices, from which follows:

Lemma 4: If $L_0 \subset o(n)$, then $L_0 = S \oplus A$ (direct sum). (ii) If $L_0 \subset o(n,1)$, then either $L_0 = S \oplus A$, A abelian or R contains k $(k \geqslant 1)$ elements conjugate to the subalgebra $\{M_{0a} - M_{1a}\}$, $a = 2, \ldots, k+1$ of "light-cone translations."

The first assertion follows from the above result since $O(n)$, being compact, contains no nilpotent elements and the second from the fact that any abelian subalgebra of $o(n,1)$ consisting of nilpotent elements is conjugate to $\{M_{0i} - M_{1i}\}$, $i = 2, \ldots, k+1$.[33]

We are now in a position to state and prove our principal lemma, which, together with Lemma 2 implies Theorem 2 and reduces Theorem 3 to the case where either $\dim t_L = 1$ or L_0 contains $k \geqslant 1$ light-cone translations.

Lemma 5: If $t_L = 0$ and either $L_0 \subset o(n)$ or L_0 contains no light-cone translations, there either exists an orthogonal direct sum decomposition

$$M = E \oplus E^{\perp}$$

such that $\|P_E(x-a)\|^2$ is an invariant of L, or L is abelian.

Proof: By Lemma 4, we have an orthogonal direct sum decomposition

$$L_0 = S \oplus A$$

with S semisimple and A abelian. Since $S \subset o(n)$ or $o(n,1)$, there exists an orthogonal direct sum decomposition of M into irreducible subspaces under S. Let E denote the sum of those subspaces upon which S acts nontrivially, and $E^{\perp}$ the orthogonal complement upon which S vanishes. Thus $t^{E^{\perp}} \subset t$ consists of those elements which commute with all elements in S. By Lemma 3, we can assume the restriction of τ to S vanishes:

$$\tau|_S = 0$$

since conjugation of the algebra L simply corresponds to a motion of the entire space M. Hence by the cocycle condition (3.10) and the commutativity of S and A we have

$$[X,\tau(a)] = \tau[X,a] = 0, \quad \forall X \in S, \ a \in A \tag{3.14}$$

and therefore

$$\operatorname{Im} \tau \subset t^{E^{\perp}}. \tag{3.15}$$

It also follows from the commutativity of S and A and the Jacobi identity that

$$[X,[\tau(a),Y]] = 0, \quad \forall X \in S, \ a \in A, \ Y \in t^{E^{\perp}} \tag{3.16}$$

and hence $E^{\perp}$ and E are invariant under A. Consequently, L is of the form

$$L = S \oplus \{a + \tau(a), \ a \in A\}, \tag{3.17}$$

where

$$\begin{aligned} &S{:}E \subset E, \quad S|_{E^{\perp}} = 0, \\ &a{:}E \subset E, \quad a{:}E^{\perp} \subset E^{\perp}, \\ &\tau(a)|_E = 0, \quad \tau(a){:}E^{\perp} \subset E^{\perp}. \end{aligned} \tag{3.18}$$

Since each component $S,a,\tau(a)$ separately preserves the orthogonal decomposition $E \oplus E^{\perp}$ and the restriction of L to E involves no translations, the length squared of vectors in E is invariant under L, and hence so is the projected length squared $\|P_E(x)\|^2$. Since the elimination of coboundaries in $\tau|_S$ involved a translation, the corresponding invariant is generally of the form $\|P_E(x-a)\|^2$.

In the case where $E = 0$ this provides no invariant, but this means S acts trivially on all of M and hence vanishes, and therefore L is abelian.

Theorem 2 follows from this result for $\dim M \geqslant 3$, since any abelian subgroup of $O(n)$ has dimension $\leqslant n/2$ and hence cannot have codimension 1 orbits. Similarly, abelian subgroups of $O(n,1)$ with no nilpotent elements have dimension $\leqslant (n+1)/2$. For $\dim M = 2$, the theorems are verified by explicitly considering all one-dimensional subalgebras.

To complete the proof of Theorem 3, we must examine separately the two remaining cases.

Case 1: $t_L = 0$, $\dim N = k \geqslant 1$, where $N = \{M_{0a} - M_{1a}\}$, $a = 2, \ldots, k+1$, is the set of nilpotent elements in the radical of L_0 (up to a redefinition of L by conjugation). The general form of elements in $\pi^{-1}(N)$ is

$$\begin{aligned} \pi^{-1}(M_{0a} - M_{1a}) \equiv X_a &= M_{0a} - M_{1a} + \tau(M_{0a} - M_{1a}) \\ &\equiv M_{0a} - M_{1a} + \sum_{b=2}^{k+1} \alpha_{ab} P_b \\ &\quad + \sum_{i=k+2}^{n} \beta_{ai} P_i \\ &\quad + A_a(P_0 + P_1) + B_a(P_0 - P_1). \end{aligned} \tag{3.19}$$

Since

$$[X_a, P_a] = P_0 - P_1 \tag{3.20}$$

we may take $B_a = 0$ up to a coboundary. Since N is abelian, we must, for $k \geqslant 2$, in view of the commutation relations

$$[X_a, X_b] = 0, \tag{3.21}$$

have $A_a = 0$. Since $\pi^{-1}(N)$ must be an ideal in L, L must be contained within the normalizer $\operatorname{nor}(\pi^{-1}(N))$ in $p(n,1)$.

For the case $k=1$, we either have $A_2=0$ or, up to conjugation by elements of the form $\exp[\lambda_i(M_{0i}-M_{1i})]$ and rotations, we have

$$X_2 = M_{02} - M_{12} + \alpha_2 P_2 + A(P_0+P_1).$$

The term $\alpha_2 P_2$ is a coboundary and may be set equal to zero, and a further conjugation by a dilatation $\exp \alpha M_{01}$ may be used to normalize $A=1$ so that

$$X_2 = M_{02} - M_{12} + P_0 + P_1. \tag{3.22}$$

The normalizer of X_2 is $\mathrm{nor}\{X_2\} = \{X_2, P_0 - P_1, M_{ij}, P_i\}$, $i,j = 3,\ldots,n$, and hence $x_2 + \frac{1}{4}(x_0+x_1)^2$ is an invariant.

Returning to the general case with $A_a = 0$ (and $k\geqslant 1$), we have

$$X_a = M_{0a} - M_{1a} + \sum_{b=2}^{k+1} \alpha_{ab} P_b + \sum_{i=k+2}^{n} \beta_{ai} P_i, \tag{3.23}$$

where, up to a coboundary,

$$\sum_{a=2}^{k+1} \alpha_{aa} = 0. \tag{3.24}$$

Henceforth, we use the summation convention and, unless stated otherwise, indicate by $\{a,b,c\}$ and $\{i,j,k\}$ indices with range

$$2 \leqslant a,b,c \leqslant k+1, \quad k+2 \leqslant i,j,k \leqslant n. \tag{3.25}$$

An arbitrary element of the normalizer has the form

$$\begin{aligned} Y = \lambda M_{01} &+ B_{ij} M_{ij} + C_{ab} M_{ab} + \lambda_i (M_{0i} - M_{1i}) \\ &+ B_i P_i + C_a P_a + r_a X_a \\ &+ r(P_0 - P_1) + s(P_0 + P_1), \end{aligned} \tag{3.26}$$

where

$$C_a = \beta_{ai} \lambda_i, \tag{3.27a}$$

$$\alpha_{ab} C_{bc} - C_{ab} \alpha_{bc} = \tfrac{1}{2} \lambda \alpha_{ac} - s \delta_{ac}, \tag{3.27b}$$

$$\beta_{aj} B_{ji} - C_{ac} \beta_{ci} = \tfrac{1}{2} \lambda \beta_{ai}. \tag{3.27c}$$

It follows from Eqs. (3.27b), (3.27c), and (3.24) that

$$s = 0$$

and either (i) $\lambda = 0$ for all Y in the normalizer or (ii) $\alpha_{ab} = 0$ and $\beta_{ai} = 0$.

In the first case, $(x^0 + x^1)$ is an invariant. In the second case, we have

$$X_a = M_{0a} - M_{1a} \tag{3.28}$$

and, by (3.27a), $C_a = 0$ and hence the entire normalizer consists of elements of the form

$$\begin{aligned} Y = \lambda M_{01} &+ B_{ij} M_{ij} + C_{ab} M_{ab} + \lambda_i (M_{0i} - M_{1i}) \\ &+ B_i P_i + r(P_0 - P_1) \end{aligned} \tag{3.29}$$

plus any linear combinations of X_a's. Invariant functions under all X_a are of the form

$$\phi = \phi\left(x^0 + x^1, (x_0^2 - x_1^2 - \cdots - x_{k+1}^2)^{1/2}, x_{k+2}, \ldots, x_n\right). \tag{3.30}$$

Considering actions upon functions of the last $n-k-1$ variables only

$$f = f(x_{k+2}, \ldots, x_n), \tag{3.31}$$

the elements $P_0 - P_1$, M_{01}, M_{ab}, and X_a act trivially (i.e., annihilate f), and the elements $M_{0i} - M_{1i}$ are identical to P_i multiplied by $-(x_0 + x_1)$ regarded as a parameter. We therefore need only consider the action of terms like

$$\widetilde{Y} = B_{ij} M_{ij} + \widetilde{B}_i P_i \tag{3.32}$$

[where $\widetilde{B}_i = B_i - (x^0 + x^1)\lambda_i$], which are contained in the Euclidean group of the reduced space $(x_{k+2}, \ldots, x_n)$. There are, by assumption, no pure translations P_i or pure light-cone translations $M_{0i} - M_{1i}$. Hence we may apply the same argument as in Lemma 5 to conclude that either there exists an orthogonal splitting $E \oplus E^{\perp}$ such that the terms $\widetilde{B}_i P_i$ act trivially on the E component while the remaining terms, which are rotations, preserve the orthogonal decomposition, or the terms $\{\widetilde{Y}\}$ form an abelian algebra. In the first case, we have an invariant $\|P_E(x)\|^2$. In the latter, if $n-k-1 \geqslant 3$, we have at least two invariants. If $n-k-1 = 2$, there can be only one term $\widetilde{Y}$ [since it is isomorphic to a subalgebra of $O(n-k-1)$], and hence in the original space there are at least two invariants. If $n-k-1 = 1$ or 0, there is no $\widetilde{Y}$ and hence the only terms in the normalizer are $\{X_a\}$ plus one element of the form (3.24) with $\lambda \neq 0$. By conjugation with group elements of the form $\exp[\mu_i(M_{0i} - M_{1i})]$ and $\exp[a(P_0 - P_1)]$, the terms $\lambda_i(M_{0i} - M_{1i})$ and $r(P_0 - P_1)$ may be transformed to zero and the remaining part leaves $(x_0^2 - x_1^2 - \cdots - x_{k+1}^2)$ invariant. This completes the analysis for Case 1.

Case 2: $\dim t_L = 1$. We may, up to conjugation, assume that $t_L = \{P_0 - P_1\}$. Since this must be an ideal, L is contained in the normalizer: $\mathrm{nor}\{P_0 - P_1\} = \{M_{01}, M_{0a} - M_{1a}, M_{ab}, P_a, P_0 - P_1, P_0 + P_1\}$, $a = 2, \ldots, n$. Denote, as before, projection onto the first component of $p(n,1) = o(n,1) \oplus t$ by

$$\pi: L \to o(n,1) \quad \text{and} \quad L_0 \equiv \operatorname{Im} \pi.$$

Let $L' \subset L$ be the subspace consisting of elements with translational part of the form $\beta_a P_a + \beta(P_0 + P_1)$. This will not generally be a subalgebra, but may be identified with the quotient $L/\{P_0 - P_1\}$. Since $\ker \pi = \{P_0 - P_1\}$, the restriction $\pi|_L$ is invertible on L_0 and again defines a linear map $\tau: L' \to t$ by

$$\pi^{-1}: X \to X + \tau(X) \in L', \quad X \in L_0,$$

such that

$$\tau[X,Y] = [\tau(X), Y] + [X, \tau(Y)], \quad \mathrm{mod}\ \{P_0 - P_1\}.$$

This is not precisely a cocycle in the sense previously defined, but may be treated similarly since conjugation by an element $Y \in t$ has the effect

$$\tau(X) \to \tau(X) + [Y, X].$$

We now consider again the ideal $N \subset L_0$ consisting of nilpotent elements in the radical, which again may be identified with light-cone translations:

$$N = \{M_{0a} - M_{1a}\}_{a=2,\ldots,k+1}.$$

If $N = 0$, the same arguments as in the proof of Lemma 4 apply, since $\{P_0 - P_1\}$, being an invariant subspace under the action of L_0, must be in the component $t^{E^{\perp}} \subset t$ upon which the semisimple part S acts trivially. The case when the semisimple part vanishes again leads to two or more invariants because of the maximal dimension of abelian subgroups of $O(n,1)$ whose radical has no nilpotent elements when $n\geqslant 3$. The case $n=2$ is treated separately in the following section. For dim $N=k\geqslant 1$, the argument follows similar lines to case 1.

Again, the subcase $k=1$ must be considered separately. Up to conjugation, there exist three possibilities, namely:

(a) $\pi^{-1}(N)=\{M_{02}-M_{12}+P_0+P_1\}$,

(b) $\pi^{-1}(N)=\{M_{02}-M_{12}+P_3\}$,

(c) $\pi^{-1}(N)=\{M_{02}-M_{12}\}$.

Since $\pi^{-1}(N)+\{P_0-P_1\}$ must be an ideal in L, we have

$$L\subset \operatorname{nor}\{\pi^{-1}(N),P_0-P_1\}.$$

For case (a),

$$\operatorname{nor}\{\pi^{-1}(N),P_0-P_1\}=\{M_{ij},P_i,P_2,X,P_0-P_1\},$$

where $X=M_{02}-M_{12}+P_0+P_1$, $3\leqslant i,j\leqslant n$. Invariant functions under $\{X,P_0-P_1\}$ have the form

$$\phi=\phi\,(x_2+\tfrac{1}{4}(x_0+x_1)^2,x_3,\ldots,x_n).$$

By the Euclidean argument, if there are any elements involving $\{M_{ij},P_i\}$ they either leave invariant a function of the type $\|P_E(x)\|^2$, where E is a subspace of $(x_3,\ldots,x_n)$, or they form an abelian algebra. If $n>4$, this latter gives rise to more than one invariant. If $n=4$, we can have an element

$$M_{34}+aP_3+bP_4,$$

which up to a conjugation is just M_{34}, and hence $(x_3^2+x_4^2)$ is an invariant. For $n=3$ there can be no rotations and hence no translation, since we have assumed $\{P_0-P_1\}$ is the only pure translation. Therefore, x_3 is an invariant.

For case (b),

$$\operatorname{nor}\{\pi^{-1}(N),P_0-P_1\}$$
$$=\{M_{ij},M_{0i}-M_{1i},M_{03}-M_{13},P_i,X,P_2,P_3,P_0-P_1\},$$

where $4\leqslant i,j\leqslant n$ and

$$X=M_{02}-M_{12}+P_3,$$

and therefore (x^0+x^1) is an invariant.

In case (c), we may assume without loss of generality that P_2 is also present, since invariants of $\{\pi^{-1}(N),P_0-P_1\}$ are also annihilated by P_2. Hence, applying Lemma 1, we may by orthogonal projection reduce the problem to one in fewer dimensions.

If $k\geqslant 2$ and $\pi^{-1}(N)=\{M_{0a}-M_{1a}\}_{a=2,\ldots,k+1}$, we may similarly add to L, without loss of generality, P_a and thereby reduce the dimension again by orthogonal projection. As in Case 1, Eqs. (3.27a) and (3.27b) determining the normalizer are still valid and imply, for

$$X_a=M_{0a}-M_{1a}+\alpha_{ab}P_b+\beta_{ai}P_i,$$

with either $\alpha_{ab}\neq 0$ or $\beta_{ai}\neq 0$, that (x^0+x^1) is an invariant.

4. SYMMETRY REDUCTION IN $M(2,1)$ AND $M(3,1)$

In the previous section we have given a list of all codimension-1 symmetry variables in $E(n)$ and $M(n,1)$. To complete the proof of Theorem 3, we must show that it is valid for the low-dimensional case of $M(2,1)$ [the results being trivial for $M(1,1)$]. In addition, we obtain all codimension 2 symmetry variables in $M(2,1)$ and codimension-1, -2, and -3 symmetry variables in $M(3,1)$. We also discuss a new phenomenon, namely "degenerate codimension k variables" that reduce Eq. (1.1) to a $(k-1)$-dimensional PDE. In particular, degenerate codimension-2 variables provide additional reductions to ODE's.

A. Symmetry variables in $M(2,1)$

The basis (2.3) for $p(2,1)$ consists of six operators

$$\{M_{01},M_{02},M_{23},P_0,P_1,P_2\}. \tag{4.1}$$

The subalgebras of $p(2,1)$ are known.[30] In keeping with the general considerations of Sec. 3, we organize the subalgebras according to the number of free translations they contain.

i. Three translations $\{P_0,P_1,P_2\}$

The action of $\{P_0,P_1,P_2\}$ on $M(2,1)$ is transitive; hence there is no invariant: $P_\mu\phi\,(x_0,x_1,x_2)=0$ for $\mu=0,1,2$ implies $\phi=$ const.

ii. Two translations

The space spanned by the two translations can have signature $(+\,+)$, $(+\,-)$, or $(+0)$ (degenerate case):

$$(+\,+)\quad \{P_1,P_2\},\quad \xi_s=x_0, \tag{4.2a}$$

$$(+\,-)\quad \{P_0,P_2\},\quad \xi_s=x_1, \tag{4.2b}$$

$$(+0)\quad \{P_0-P_1,P_2\},\quad \xi_s=x_0+x_1. \tag{4.2c}$$

The invariant can in general be written as

$$\xi=A_\mu x^\mu=A_0x_0-A_1x_1-A_2x_2$$

with $A^2=1$, -1, or 0, respectively. The maximal subalgebras of $p(2,1)$ having these invariants are $\{M_{12},P_1,P_2\}$, $\{M_{02},P_0,P_2\}$, and $\{M_{02}-M_{12},P_2,P_0-P_1\}$, respectively.

iii. One translation

(a) Signature $(+)$: P_2: The equation $P_2\phi\,(x_0,x_1,x_2)$ $=\partial_2\phi\,(x_0,x_1,x_2)=0$ implies $\phi=\phi\,(x_0,x_1)$ (4.3)

and (x_0,x_1) are codimension-2 symmetry variables. To reduce the codimension to 1, we must look into the normalizer of P_2 in $p(2,1)$:

$$\operatorname{nor}P_2=\{M_{01},P_0,P_1,P_2\}. \tag{4.4}$$

The algebra L by assumption contains no further translations, so the only element we can add to P_2 is

$$M=M_{01}+a_\mu P_\mu.$$

Using conjugacy by P_0 and P_1, we reduce L in this case to

$$L=\{M_{01},P_2\}. \tag{4.5}$$

The conditions $M_{01}\phi=0$, $P_2\phi=0$ imply $\phi=\phi\,(\xi)$,

$$\xi=(x_0^2-x_1^2)^{1/2}. \tag{4.6}$$

(b) Signature $(-)$: P_0: The equation $P_0\phi\,(x_0,x_1,x_2)=0$ implies

$$\phi=\phi\,(x_1,x_2), \tag{4.7}$$

and we find that (x_1,x_2) are codimension-2 symmetry variables. We have

$$\operatorname{nor}P_0=\{M_{12},P_1,P_2,P_0\}, \tag{4.8}$$

and the only (up to conjugacy) subalgebra of nor P_0 containing P_0 as the only translation is

$$\{M_{12},P_0\}. \tag{4.9}$$

The only independent invariant of (4.9) is

$$\xi=(x_1^2+x_2^2)^{1/2}. \tag{4.10}$$

(c) Signature (0): P_0-P_1: Equation $(P_0-P_1)\phi=0$ provides the result

$$\phi=\phi(x_0+x_1,x_2), \tag{4.11}$$

and the codimension-2 invariants (x_0+x_1,x_2). We have

$$\text{nor}(P_0-P_1)=\{M_{01},M_{02}-M_{12},P_0+P_1,P_2,P_0-P_1\}. \tag{4.12}$$

Up to conjugacy under $P(2,1)$ nor (P_0-P_1) has five different types of subalgebras containing (P_0-P_1), as the only translation. Let us consider each of them separately

$$(\mathrm{c}_1)\quad \{X_i\}=\{M_{01}+aP_2,M_{02}-M_{12},P_0-P_1\},\quad a\in\mathbb{R}. \tag{4.13}$$

The corresponding Lie group acts transitively on $M(2,1)$ and correspondingly $X_i\phi=0$ implies $\phi=\text{const}$.

$$(\mathrm{c}_2)\quad \{M_{01},P_0-P_1\}. \tag{4.14}$$

The invariant of (4.14) is x_2.

$$(\mathrm{c}_3)\quad \{M_{01}+aP_2,P_0-P_1\},\quad a\neq 0. \tag{4.15}$$

The equations

$$\begin{aligned}&(M_{01}+aP_2)\phi=[-(x_0\partial_1+x_1\partial_0)+a\partial_2]\phi=0,\\ &(P_0-P_1)\phi=(\partial_0-\partial_1)\phi=0\end{aligned} \tag{4.16}$$

imply $\phi=\phi(\xi)$,

$$\xi=x_2+a\ln(x_0+x_1) \tag{4.17}$$

providing us with the invariant (3.6d).

$$(\mathrm{c}_4)\quad \{M_{02}-M_{12},P_0-P_1\}. \tag{4.18}$$

This algebra yields an invariant that is already known, namely $\eta=x_0+x_1$.

$$(\mathrm{c}_5)\quad \{M_{02}-M_{12}+P_0+P_1,P_0-P_1\}. \tag{4.19}$$

The equations

$$\begin{aligned}&(M_{02}-M_{12}+P_0+P_1)\phi\\ &\quad=[-(x_0+x_1)\partial_2-x_2(\partial_0-\partial_1)+\partial_0+\partial_1]\phi=0,\end{aligned} \tag{4.20}$$

$$(P_0-P_1)\phi=(\partial_0-\partial_1)\phi=0$$

imply $\phi=\phi(\xi)$, where

$$\xi=x_2+\tfrac{1}{4}(x_0+x_1)^2, \tag{4.21}$$

i.e., we obtain the invariant (3.6c).

iv. No free translations

Among the subalgebras of $p(2,1)$ with no free translations, only two have generic orbits of codimension 1, namely, the o(2,1) algebra $\{M_{01},M_{02},M_{12}\}$ and its subalgebra $\{M_{01},M_{02}-M_{12}\}$. Both have the same orbits and hence lead to the same codimension-1 symmetry variable, namely,

$$\xi=(x_0^2-x_1^2-x_2^2)^{1/2}. \tag{4.22}$$

We see that the complete list of "standard" codimension-1 symmetry variables in $M(2,1)$ consists of

$$\begin{aligned}&x_0,\quad x_1,\quad x_0+x_1,\quad (x_0^2-x_1^2)^{1/2},\quad (x_1^2+x_2^2)^{1/2},\\ &\quad(x_0^2-x_1^2-x_2^2)^{1/2},\\ &x_2+a\ln(x_0+x_1),\quad x_2+\tfrac{1}{4}(x_0+x_1)^2,\end{aligned} \tag{4.23}$$

in complete agreement with Table II and Theorem 3 (the proof of which is hereby completed).

Let us continue with the study of codimension-2 symmetry variables in $M(2,1)$. It remains to discuss one-dimensional subalgebras of o(2,1), possibly extended by translations. Six cases occur, and we list them together with their invariants:

$$M_{01}:\quad \rho=(x_0^2-x_1^2)^{1/2},\quad x_2; \tag{4.24}$$

$$M_{12}:\quad r=(x_1^2+x_2^2)^{1/2},\quad x_0; \tag{4.25}$$

$$M_{02}-M_{12}:\sigma=(x_0^2-x_1^2-x_2^2)^{1/2},\quad x_0+x_1; \tag{4.26}$$

$$\begin{aligned}&M_{01}+aP_2:\rho=(x_0^2-x_1^2)^{1/2},\\ &\xi=x_2+a\ln(x_0+x_1),\quad a\neq 0;\end{aligned} \tag{4.27}$$

$$\begin{aligned}&M_{12}+aP_0:\quad r=(x_1^2+x_2^2)^{1/2},\\ &\phi=x_0+a\arcsin[x_1/(x_1^2+x_2^2)^{1/2}],\quad a\neq 0;\end{aligned} \tag{4.28}$$

$$\begin{aligned}&M_{02}-M_{12}+P_0+P_1:\quad \alpha=x_2+\tfrac{1}{4}(x_0+x_1)^2,\\ &\beta=x_0-x_1+(x_0+x_1)x_2+\tfrac{1}{6}(x_0+x_1)^3.\end{aligned} \tag{4.29}$$

To summarize: (4.3), (4.7), (4.11), and (4.24)–(4.29) provide a complete list of representatives of all $P(2,1)$ conjugacy classes of codimension-2 symmetry variables in $M(2,1)$. The reduced two-dimensional PDE's are obtained by substituting the obtained variables into (2.7), where (ξ_1,ξ_2) is identified with one of the obtained pairs of variables and $n=2$. Finally, the expressions (2.7) are substituted back in Eq. (1.1). The results are summarized in Table III.

Each pair of "standard" variables (ξ_1,ξ_2) listed in column 2 of Table III represents a $P(2,1)$ conjugacy class of variables. These are obtained by translating the origin o to an arbitrary point $\widetilde{B}\in M(2,1)$ and Lorentz-transforming the standard triad $(\tilde{e}_0,\tilde{e}_1,\tilde{e}_2)$ to the general triad $(\widetilde{A}_0,\widetilde{A}_1,\widetilde{A}_2)$ satisfying $(\widetilde{A}_\mu\widetilde{A}_\nu)=g_{\mu\nu}$. For instance, the pair (4.28) represents the class

$$\begin{aligned}&\xi_1=[(\tilde{x}-\widetilde{B},\widetilde{A}_1)^2+(\tilde{x}-\widetilde{B},\widetilde{A}_2)^2]^{1/2},\\ &(\widetilde{A}_i,\widetilde{A}_j)=-\delta_{ij},\quad i,j=1,2,\\ &\xi_2=(\widetilde{A}_0,\tilde{x}-\widetilde{B})\\ &\quad+a\arcsin\{(\widetilde{A}_1,\tilde{x}-\widetilde{B})/[(\tilde{x}-\widetilde{B},\widetilde{A}_1)^2+(\tilde{x}-\widetilde{B},\widetilde{A}_2)^2]^{1/2}\}.\end{aligned} \tag{4.30}$$

The reduced equations will be discussed in Sec. 5. Notice that in Table III the last two pairs (ξ_1,ξ_2) each include a variable that never occurs on its own as a codimension-1 invariant.

B. Symmetry variables in $M(3,1)$

The procedure for $M(3,1)$ is exactly the same as for $M(2,1)$. We make use of the classification of all subalgebras of $p(3,1)$ (see Refs. 27–29), identify those with orbits of codimension 1, 2, and 3, respectively, and calculate their invariants. The codimension-1 symmetry variables are those giv-

TABLE III. Codimension 2 symmetry variables in $M(2, 1)$.

No.	Standard variables ξ_1, ξ_2	Algebra	$(\nabla\xi_1)^2$	$(\nabla\xi_2)^2$	$(\nabla\xi_1, \nabla\xi_2)$	$\Box\xi_1$	$\Box\xi_2$
1	x_1, x_2	P_0	-1	-1	0	0	0
2	x_0, x_1	P_2	1	-1	0	0	0
3	$x_0 + x_1, x_2$	$P_0 - P_1$	0	-1	0	0	0
4	$x_2, (x_0^2 - x_1^2)^{1/2}$	M_{01}	-1	1	0	0	$1/\xi_2$
5	$x_0, (x_1^2 + x_2^2)^{1/2}$	M_{12}	1	-1	0	0	$-1/\xi_2$
6	$x_0 + x_1, (x_0^2 - x_1^2 - x_2^2)^{1/2}$	$M_{02} - M_{12}$	0	1	ξ_1/ξ_2	0	$2/\xi_2$
7	$x_2 + a\ln(x_0 + x_1), (x_0^2 - x_1^2)^{1/2}, \quad a\neq 0$	$M_{01} + aP_2$	-1	1	a/ξ_2	0	$1/\xi_2$
8	$x_2 + a\arcsin[x_1/(x_1^2 + x_2^2)^{1/2}], (x_1^2 + x_2^2)^{1/2}$	$M_{12} + aP_0$	$1 - a^2/\xi_2^2$	-1	0	0	$-1/\xi_2$
9	$x_2 + \frac{1}{4}(x_0 + x_1)^2, x_0 - x_1 + (x_0 + x_1)x_2 + \frac{1}{6}(x_0 + x_1)^3$	$M_{02} - M_{12} + P_0 + P_1$	-1	$4\xi_1$	0	0	0

en in Table II (once we particularize to $n = 3$), as stated in Theorem 3. We shall just give one example of a subalgebra $L\subset p(3,1)$ having generic orbits of codimension 2 and then summarize the results on all codimension-2 and -3 symmetry variables in tables.

Consider the subalgebra $\{aM_{23} + M_{01}, P_0 - P_1\}$, $a\neq 0$. The equation

$$(P_0 - P_1)\phi(x_0, x_1, x_2, x_3) = \left(\frac{\partial}{\partial x_0} - \frac{\partial}{\partial x_1}\right)\phi = 0 \quad (4.31)$$

implies $\phi = \phi(x_0 + x_1, x_2, x_3)$. The equation

$$\begin{aligned}(aM_{23} + M_{01})\phi(\eta, x_2, x_3) &= \left[a\left(x_2\frac{\partial}{\partial x_3} - x_3\frac{\partial}{\partial x_2}\right)\right. \\ &\quad \left. - \left(x_0\frac{\partial}{\partial x_1} + x_1\frac{\partial}{\partial x_0}\right)\right]\phi = 0, \\ \eta &\equiv x_0 + x_1,\end{aligned} \quad (4.32)$$

is solved by completing (4.32) to a complete differential, i.e., requiring

$$\frac{dx_3}{ax_2} = -\frac{dx_2}{ax_3} = \frac{d\eta}{-\eta}. \quad (4.33)$$

This in turn is integrated to yield

$$\begin{aligned}\xi_1 &= \arcsin[x_2/(x_2^2 + x_3^2)^{1/2}] + a\ln(x_0 + x_1), \\ \xi_2 &= (x_2^2 + x_3^2)^{1/2},\end{aligned} \quad (4.34)$$

i.e., the general solution of (4.31) and (4.32) is $\phi(\xi_1, \xi_2)$, with ξ_1 and ξ_2 as in (4.34).

In order to obtain the reduced equation, we calculate

$$\begin{aligned}&(\nabla\xi_1)^2 = -(1/\xi_2^2), \quad (\nabla\xi_2)^2 = -1, \quad (\nabla\xi_1, \nabla\xi_2) = 0, \\ &\Box\xi_1 = 0, \quad \Box\xi_2 = -1/\xi_2.\end{aligned} \quad (4.35)$$

Equation (1.1) hence reduces to

$$H\left(-\frac{1}{\xi_2^2}u_{\xi_1\xi_1} - u_{\xi_2\xi_2} - \frac{1}{\xi_2}u_{\xi_2}, -\frac{1}{\xi_2^2}u_{\xi_1}^2 - u_{\xi_2}^2, u\right) = 0. \quad (4.36)$$

A complete list of representatives of $P(3,1)$ classes of codimension-2 and codimension-3 symmetry variables in $M(3,1)$ are given in Tables IV and V, respectively.

Notice that a certain structure emerges. In Table IV entries 1–9 are identical with those in Table III. The reason for this is that the presence of a spacelike translation (say P_3) in the invariance algebra L immediately reduces the $M(3,1)$ problem to an $M(2,1)$ one. Entries 10 and 11 are codimension-2 symmetry variables in $E(3)$ (as is entry 1); the $M(3,1)$ problem is reduced to an $E(3)$ one by the presence of P_0 in the algebra L. Entries 12 and 13 correspond to the introduction of codimension-1 symmetry variables in each component of the tensor product $M(1,1)\otimes E(2)$. Similarly, 14,15,16 (as well as 1, ... ,5) can be interpreted as codimension-1 variables in $S(1)\times M(2,1)$, where $S(1)$ is a spacelike line. The entry 17 (as well as 2,3) corresponds to codimension-1 variables in $T(1)\times E(3)$, where $T(1)$ is a timelike line. The entries 18,19,20 (as well as 1, 3, 13, 15, and 16) are actually codimension-2 variables in $L(0++)$, a three-dimensional space with degenerate metric annihilated by a lightlike translation, say $(P_0 - P_1)$. The only new codimension-2 variables that do not involve a direct product decomposition of $M(3,1)$ are given by entries 21, 22, and 23. Notice, however, that variables that do not occur as codimension-1 symmetry variables in $M(3,1)$ or codimension-2 variables in $E(3)$ or $M(2,1)$ figure in entries 18, 19, 20, and 22. [The variables in 18–20 would figure in an analysis of codimension-2 variables in $L(0++)$.]

Table V contains no new types of variables with respect to Table IV, but the triplets of codimension-3 variables are by no means obvious ones.

C. Degenerate symmetry variables

We have seen that the introduction of the codimension-k symmetry variables $(\xi_1, \ldots, \xi_k)$, as invariants of a subgroup $G\subset P(n,1)$ with orbits of codimension k, assures that Eqs. (2.6) of Lemma 1 are satisfied and that hence Eq. (1.1) is reduced to a PDE in k variables (in particular an ODE if $k = 1$).

On the quotient space the invariants $\{\xi_i\}$ define a local coordinate system. The matrix of scalar products of their gradients

$$\alpha_{ij}(\xi) = (\nabla\xi_i, \nabla\xi_j),$$

together with their Laplacian

$$\beta_j(\xi) = \Box\xi_j,$$

determines the reduced operators $\Box u$ and $(\nabla u)^2$. If there exists a kernel (null space) for the linear system (α_{ij}, β_j) which is integrable, in the sense that it may be associated with a coordinate (i.e., a hypersurface foliation), the reduced equation will involve no derivatives with respect to that coordinate. A

TABLE IV. Codimension 2 symmetry variables in $M(3, 1)$.

No.	Variables ξ_1, ξ_2	Algebra	$(\nabla\xi_1)^2$	$(\nabla\xi_2)^2$	$(\nabla\xi_1, \nabla\xi_2)$	$\Box\xi_1$	$\Box\xi_2$
1	x_1, x_2	M_{03}, P_0, P_3	-1	-1	0	0	0
2	x_0, x_1	M_{23}, P_2, P_3	1	-1	0	0	0
3	$x_0 + x_1, x_2$	$M_{03} - M_{13}, P_0 - P_1, P_3$	0	-1	0	0	0
4	$x_2, (x_0^2 - x_1^2)^{1/2}$	M_{01}, P_3	-1	1	0	0	$1/\xi_2$
5	$x_0, (x_1^2 + x_2^2)^{1/2}$	M_{12}, P_3	1	-1	0	0	$-1/\xi_2$
6	$x_0 + x_1, (x_0^2 - x_1^2 - x_2^2)^{1/2}$	$M_{02} - M_{12}, P_3$	0	1	ξ_1/ξ_2	0	$2/\xi_2$
7	$x_2 + a\ln(x_0 + x_1), (x_0^2 - x_1^2)^{1/2}$	$M_{01} + aP_2, P_3$	-1	1	a/ξ_2	0	$1/\xi_2$
8	$x_0 + a\arcsin[x_1/(x_1^2 + x_2^2)^{1/2}], (x_1^2 + x_2^2)^{1/2}$	$M_{12} + aP_0, P_3$	1	-1	0	0	$-1/\xi_2$
9	$x_2 + \frac{1}{4}(x_0 + x_1)^2, x_0 - x_1 + (x_0 + x_1)x_2 + \frac{1}{6}(x_0 + x_1)^3$	$M_{02} - M_{12} + P_0 + P_1, P_3$	-1	$4\xi_1$	0	0	0
10	$x_3, (x_1^2 + x_2^2)^{1/2}$	M_{12}, P_0	-1	-1	0	0	$-1/\xi_2$
11	$x_3 + a\arcsin[x_1/(x_1^2 + x_2^2)^{1/2}], (x_1^2 + x_2^2)^{1/2}$	$M_{12} + aP_3, P_0$	$-1 - a^2/\xi_2^2$	-1	0	0	$-1/\xi_2$
12	$(x_0^2 - x_1^2)^{1/2}, (x_2^2 + x_3^2)^{1/2}$	M_{01}, M_{23}	1	-1	0	$1/\xi_1$	$-1/\xi_2$
13	$x_0 + x_1, (x_2^2 + x_3^2)^{1/2}$	$M_{23}, P_0 - P_1$	0	-1	0	0	$-1/\xi_2$
14	$x_3, (x_0^2 - x_1^2 - x_2^2)^{1/2}$	M_{01}, M_{02}, M_{12}	-1	1	0	0	$2/\xi_2$
15	$x_3, x_2 + \frac{1}{4}(x_0 + x_1)^2$	$M_{02} - M_{12} + P_0 + P_1, P_0 - P_1$	-1	-1	0	0	0
16	$x_3, x_2 + a\ln(x_0 + x_1)$	$M_{01} + aP_2, P_0 - P_1$	-1	-1	0	0	0
17	$x_0, (x_1^2 + x_2^2 + x_3^2)^{1/2}$	M_{12}, M_{23}, M_{31}	1	-1	0	0	$-2/\xi_2$
18	$x_0 + x_1, x_2 + \epsilon(x_0 + x_1)x_3$	$M_{02} - M_{12} + \epsilon P_3, P_0 - P_1$	0	$-1 - \xi_1^2$	0	0	0
19	$\epsilon(x_0 + x_1) + \arcsin[x_2/(x_2^2 + x_3^2)^{1/2}], (x_2^2 + x_3^2)^{1/2}$	$M_{23} - (\epsilon/2)(P_0 + P_1), P_0 - P_1$	$-4/\xi_2^2$	-1	0	0	$-1/\xi_2$
20	$a\ln(x_0 + x_1) + \arcsin[x_2/(x_2^2 + x_3^2)^{1/2}], (x_2^2 + x_3^2)^{1/2}$	$aM_{23} + M_{01}, P_0 - P_1$	$-1/\xi_2^2$	-1	0	0	$-1/\xi_2$
21	$x_0 + x_1, (x_0^2 - x_1^2 - x_2^2 - x_3^2)^{1/2}$	$M_{23}, M_{02} - M_{12}, M_{03} - M_{13}$	0	1	ξ_1/ξ_2	0	$3/\xi_2$
22	$x_0 + x_1, x_3^2(x_0 + x_1) + (1 - x_0 - x_1)(x_0^2 - x_1^2 - x_2^2)$	$M_{02} - M_{12}, M_{03} - M_{13} + P_3$	0	$-4\xi_2(2\xi_1 + 1)$	$-2\xi_1(\xi_1 + 1)$	0	$-6(2\xi_1 + 1)$
23	$x_2 + a\ln(x_0 + x_1), (x_0^2 - x_1^2 - x_3^2)^{1/2}$	$M_{01} + aP_2, M_{03} - M_{13}$	-1	1	a/ξ_2	0	$2/\xi_2$

TABLE V. Codimension 3 symmetry variables in $M(3, 1)$.

No.	Standard variables ξ_1, ξ_2, ξ_3	Algebra	$(\nabla\xi_1)^2$	$(\nabla\xi_2)^2$	$(\nabla\xi_3)^2$	$(\nabla\xi_1, \nabla\xi_2)$	$(\nabla\xi_2, \nabla\xi_3)$	$(\nabla\xi_3, \nabla\xi_1)$	$\Box\xi_1$	$\Box\xi_2$	$\Box\xi_3$
1	x_1, x_2, x_3	P_0	-1	-1	-1	0	0	0	0	0	0
2	x_0, x_1, x_2	P_3	1	-1	-1	0	0	0	0	0	0
3	$x_0 + x_1, x_2, x_3$	$P_0 - P_1$	0	-1	-1	0	0	0	0	0	0
4	$x_0, x_3, (x_1^2 + x_2^2)^{1/2}$	M_{12}	1	-1	-1	0	0	0	0	0	$-1/\xi_3$
5	$x_1, x_2, (x_0^2 - x_3^2)^{1/2}$	M_{03}	-1	-1	1	0	0	0	0	0	$1/\xi_3$
6	$x_0 + x_1, x_3, (x_0^2 - x_1^2 - x_2^2)^{1/2}$	$M_{02} - M_{12}$	0	-1	1	0	0	ξ_1/ξ_3	0	0	$2/\xi_3$
7	$a \ln(x_0 + x_1) + \arcsin[x_2/(x_2^2 + x_3^2)^{1/2}], (x_2^2 + x_3^2), (x_0^2 - x_1^2)^{1/2}$	$aM_{23} + M_{01}$	$-1/\xi_2^2$	-1	1	0	0	$1/\xi_3$	0	$-1/\xi_2$	$1/\xi_3$
8	$x_0 + x_1, \epsilon(x_0 - x_1) + \arcsin[x_2/(x_2^2 + x_3^2)^{1/2}], (x_2^2 + x_3^2)^{1/2}$	$M_{23} + (\epsilon/2)(P_0 - P_1)$	0	$-1/\xi_3^2$	-1	2ϵ	0	0	0	0	$-1/\xi_3$
9	$x_3, x_0 + a \arcsin[x_1/(x_1^2 + x_2^2)^{1/2}], (x_1^2 + x_2^2)^{1/2}$	$M_{12} - aP_0$	-1	$1 - a^2/\xi_3^2$	-1	0	0	0	0	0	$-1/\xi_3$
10	$x_0, x_3 + a \arcsin[x_1/(x_1^2 + x_2^2)^{1/2}], (x_1^2 + x_2^2)^{1/2}$	$M_{12} - aP_3$	1	$-1 - a^2/\xi_3^2$	-1	0	0	0	0	0	$-1/\xi_3$
11	$x_3, x_2 + a \ln(x_0 + x_1), (x_0^2 - x_1^2)^{1/2}$	$M_{01} + aP_2$	-1	-1	1	0	a/ξ_3	0	0	0	$1/\xi_3$
12	$x_3, x_2 + \frac{1}{4}(x_0 + x_1)^2, x_0 - x_1 + (x_0 + x_1)x_2 + \frac{1}{6}(x_0 + x_1)^2$	$M_{02} - M_{12} + P_0 + P_1$	-1	-1	$4\xi_2$	0	0	0	0	0	0
13	$x_0 + x_1, x_2 + \epsilon(x_0 + x_1)x_3, (x_0^2 - x_1^2 - x_2^2)^{1/2}$	$M_{02} - M_{12} + \epsilon P_3$	0	$-1 - \xi_1^2$	1	0	ξ_2/ξ_3	ξ_1/ξ_3	0	0	$2/\xi_3$

necessary condition for this to occur is that the induced metric on the generic orbits be degenerate.

We plan to return to this phenomenon in the future in greater detail. Here we shall just discuss the consequences of such a degeneracy in the low-dimensional cases of $M(2,1)$ and $M(3,1)$.

For a codimension $k = 1$ variable ξ, Eq. (2.8) are always satisfied. In the degenerate case, namely, $\eta = x_0 + x_1$ [or more generally $\eta = (\widetilde{A}, \tilde{x}), \widetilde{A}^2 = 0$] we have a stronger condition, namely,

$$\Box\eta = 0, \quad (\nabla\eta)^2 = 0. \tag{4.37}$$

Equation (1.1) thus reduces to a functional equation (involving no derivatives)

$$H(0,0,u(\eta)) = 0.$$

This usually gives only trivial results, e.g., for the sine–Gordon equation,

$$\Box u = \sin u, \tag{4.38}$$

the ansatz $u = u(\eta)$, $\eta = x_0 + x_1$ yields $\sin u = 0$, i.e., $u(\eta) = m\pi (m \in \mathbb{Z})$.

For higher codimensions, degenerate symmetry variables lead to much more interesting results.

Let us first consider the codimension $k = 2$ case. We denote $\xi_1 = \eta$ and $\xi_2 = \xi$. The derivatives $u_\eta, u_{\eta\eta}, u_{\eta\xi}$, drop out of the reduced equation if

$$(\nabla\eta)^2 = 0, \quad (\nabla\eta, \nabla\xi) = 0, \quad \Box\eta = 0. \tag{4.39}$$

Instead of a PDE in two variables we obtain an ODE in ξ only. The variable η either drops out altogether [if $(\nabla\xi)^2$ and $\Box\xi$ depend on ξ only] or remains as a parameter.

This suggests the possibility of a new type of reduction to ODE's, namely, the introduction of "degenerate codimension-2 symmetry variables." To do this, we look for a new variable

$$\tau = \phi(\eta, \xi) \tag{4.40}$$

satisfying

$$(\nabla\tau)^2 = \phi_\xi^2 (\nabla\xi)^2 = c_1, \quad \Box\tau = \phi_{\xi\xi}(\nabla\xi)^2 + \phi_\xi \Box\xi = c_2, \tag{4.41}$$

where c_1 and c_2 are constants. If such a τ exists, then according to Lemma 1 it will reduce Eq. (1.1) to an ODE in τ alone. This phenomenon can also be studied directly in terms of symmetry reduction, applied, however, to Eq. (2.6) rather than (1.1) directly.

Without going into details, let us just find all degenerate codimension-2 variables in $M(2,1)$ and $M(3,1)$. Looking at Table III, we see that conditions (4.39) are satisfied in one case only, namely, the variables $\eta = x_0 + x_1$ and $\xi = x_2$ of case no. 3. Equations (4.41) in this case reduce to

$$(\nabla\tau)^2 = -\phi_\xi^2 = c_1, \quad \Box\tau = -\phi_{\xi\xi} = c_2. \tag{4.42}$$

The solution of (4.42) is

$$\tau = x_2 + \psi(x_0 + x_1) \tag{4.43}$$

(we have put $c_1 = -1$ and obtain $c_2 = 0$), where ψ is an arbitrary function. Putting $u = u(\tau)$, we reduce (1.1) to the ODE

$$H(-u_{\tau\tau}, -u_\tau^2, u) = 0. \tag{4.44}$$

Now let us turn to Table IV, listing the codimension-2 symmetry variables in $M(3,1)$. Conditions (4.39) are satisfied by the variables nos. 3, 13, and 18. Case no. 3 is identical with the one discussed above, and we obtain the same variable (4.43). For no. 13 we have $\eta = x_0 + x_1$, $\xi = (x_2^2 + x_3^2)^{1/2}$, and (4.41) reduces to

$$(\nabla\tau)^2 = -\phi^2 = -1, \quad \Box\tau = -\phi_{\xi\xi} - (1/\xi)\phi_\xi = c_2. \tag{4.45}$$

The first equation implies $\tau = \xi + \psi(\eta)$, but then $\Box\tau = -1/\xi$ and the second of Eqs. (4.45) cannot be satisfied.

In case no. 18 we have $\eta = x_0 + x_1$, $\xi = x_2 + \epsilon(x_0 + x_1)x_3$, and Eqs. (4.41) yield

$$(\nabla\tau)^2 = -\phi_\xi^2(1+\eta^2) = -1,$$
$$\Box\tau = -\phi_{\xi\xi}(1+\eta^2) = c_2. \tag{4.46}$$

The solution is

$$\tau = [x_2 + \epsilon(x_0+x_1)x_3]/[1+(x_0+x_1)^2]^{1/2} + \psi(x_0+x_1), \tag{4.47}$$

where ψ is again an arbitrary function and $c_2 = 0$. The ansatz $u = u(\tau)$ reduces Eq. (1.1) to (4.44).

The variable (4.43) will reduce (1.1) to an ODE in $M(n,1)$ for any $n \geqslant 2$, and the variable (4.47), for any $n \geqslant 3$. Further degenerate codimension-2 symmetry variables will exist for $n \geqslant 4$.

Finally, let us consider "degenerate codimension-3 symmetry variables" that reduce (1.1) to a PDE in two dimensions (rather than 3). In this case we must require that one of the three codimension-3 invariants, say $\xi_1 = \eta$, drop out of the reduced equation, i.e.,

$$(\nabla\eta)^2 = 0, \quad \Box\eta = 0, \quad (\nabla\eta, \nabla\xi_i) = 0, \quad i,k = 2,3. \tag{4.48}$$

Looking at Table V for $M(3,1)$, we see that conditions (4.48) are satisfied in case 3 only. Equations (4.48) in this case allow us to introduce new variables, involving arbitrary functions of the "lightlike" variable η, that will reduce (1.1) to a PDE in two variables. Without going into a systematic study here, we give two examples of such variables. The obvious generalization of (4.43) is

$$\alpha = x_2 + f(x_0+x_1), \quad \beta = x_3 + g(x_0+x_1). \tag{4.49}$$

Equation (1.1) in these variables reduces to

$$H(-u_{\alpha\alpha} - u_{\beta\beta}, -u_\alpha^2 - u_\beta^2, u) = 0. \tag{4.50}$$

Similarly, if we put

$$\xi = x_2 \cos\phi(x_0+x_1) + x_3 \sin\phi(x_0+x_1), \tag{4.51}$$
$$\rho = (x_2^2 + x_3^2)^{1/2},$$

we find that (1.1) reduces to

$$H(-u_{\xi\xi} - u_{\rho\rho} - (\xi/\rho)u_{\xi\rho} - (1/\rho)u_\rho,$$
$$-u_\xi^2 - u_\rho^2 - 2(\xi/\rho)u_\xi u_\rho, u) = 0. \tag{4.52}$$

Equivalently, instead of (4.51) we can introduce

$$\xi = \arcsin[x_2/(x_2^2+x_3^2)^{1/2}] + \phi(x_0+x_1), \tag{4.51'}$$
$$\rho = (x_2^2 + x_3^2)^{1/2},$$

reducing (1.1) to

$$H(-(1/\rho^2)u_{\xi\xi} - u_{\rho\rho} - (1/\rho)u_\rho, -(1/\rho^2)u_\xi^2 - u_\rho^2, u) = 0. \tag{4.52'}$$

5. DISCUSSION OF THE REDUCED EQUATIONS

A. The ordinary differential equations

We have seen in the preceding sections that symmetry reduction of Eq.(1.1) via codimension-1 or degenerate codimension-2 variables leads to a single type of ODE, namely, Eq. (3.7):

$$H(\kappa[u_{\xi\xi} + (k/\xi)u_\xi], \kappa u^2, u) = 0.$$

Let us discuss some special cases of this equation. To be more specific, let us consider the nonlinear Klein–Gordon equation (1.5), with the right-hand side depending on u alone:

$$\Box u = F(u). \tag{5.1}$$

Puting $u = u(\xi)$, where ξ is one of the variables in Table II, or (4.43) (for $n \geqslant 2$), or (4.47) (for $n \geqslant 3$), we obtain

$$u_{\xi\xi} + (k/\xi)u_\xi = \epsilon F(u), \quad \epsilon = \pm 1, \quad k = 0,1,\ldots,n \tag{5.2}$$

(we exclude the degenerate variable $\xi = x_0 + x_1$ for which $\kappa = 0$). For $k = 0$ Eq. (5.2) reduces to quadratures:

$$\xi + \xi_0 = \int \frac{du}{\sqrt{2\epsilon\int_{u_0}^u F(u')\,du'}}, \tag{5.3}$$

where ξ_0 and u_0 are constants and ξ is one of the variables

$$x_0,\ x_1,\ x_2 + \phi(x_0+x_1),$$
$$[x_2 + \epsilon(x_0+x_1)x_3]/[1+(x_0+x_1)^2]^{1/2} + \phi(x_0+x_1), \tag{5.4}$$

[or a variable obtained by applying a Lorentz transformation to one of the variables (5.4)].

For $k \geqslant 1$ a direct integration is impossible. In each particular case, we can investigate properties of solutions, in particular, their singularity structure. Let us, for example, consider the sine–Gordon equation (4.38) in $M(n,1)$ and reduce it to

$$u_{\xi\xi} + (k/\xi)u_\xi = \epsilon \sin u, \quad k = 1,2,\ldots,n, \tag{5.5}$$

where

$$\xi = (x_1^2 + \cdots + x_{k+1}^2)^{1/2}, \quad \epsilon = -1,$$

or (5.6)

$$\xi = (x_0^2 - x_1^2 - \cdots - x_k^2)^{1/2}, \quad \epsilon = +1.$$

We can now determine whether (5.5) has the "Painlevé property,"[7–11] i.e., whether it has any moving singularities other than poles. To do this, we follow a procedure due to Ablowitz *et al.*[11] First, we transform (5.5) into a form that is algebraic in the dependent variable and its derivatives by putting

$$u = 2i \ln y \tag{5.7}$$

so that (5.5) reduces to

$$y\ddot{y} = \dot{y}^2 - (k/\xi)y\dot{y} + \tfrac{1}{4}\epsilon(y^4 - 1). \tag{5.8}$$

We now look for a solution in the form

$$y=\sum_{r=0}^{\infty} a_r(\xi-\xi_0)^{r-\alpha}, \tag{5.9}$$

where a_r are constants. The necessary condition for (5.8) to be of Painlevé type is that the general solution have the form (5.9), where α is a positive integer. This means that (5.9) should involve two arbitrary constants: ξ_0 (the position of the pole) and one of the coefficients a_r in the expansion (related to the residue at the pole). Substituting (5.9) into (5.8) and comparing different powers of $(\xi-\xi_0)$, we find

$$\alpha=1 \tag{5.10}$$

(from a leading order analysis) and obtain a recursion relation of the form

$$(r+1)(r-2)a_r=F(a_0,a_1,\ldots,a_{r-1}). \tag{5.11}$$

At $r=2$ we obtain a "resonance,"[11] i.e., (5.11) is either satisfied identically and a_2 is arbitrary, or we obtain a contradiction if

$$F(a_0,a_1)\neq 0. \tag{5.12}$$

In this second case, Eq. (5.8) is not of the Painlevé type. Evaluating the first few terms in (5.9) using (5.8), we find

$$a_0=2\sqrt{\epsilon},\quad a_1=-k\sqrt{\epsilon}/\xi,$$
$$0\cdot a_2=k(k-1). \tag{5.13}$$

Thus, Eq. (5.8) is of Painlevé type only for $k=0$, or $k=1$. For $k=0$ we already know the solution, since in this case (5.5) is

$$u_{\xi\xi}=\epsilon\sin u,$$

i.e., the exact pendulum equation[34] and (5.3) reduces either to soliton like solutions of the sine–Gordon equation, or to periodic or quasiperiodic solutions, depending on the choice of integration constants u_0,ξ_0 in (5.3) for $F(u)=\sin u$.[18,35]

For $k=1$, (5.8) is an equation satisfied by one of the Painlevé transcendants, namely, a special case of P_{III}.[7–10] Its properties have recently been studied.[36,37]

Let us stress that the above analysis provides new solutions for the sine-Gordon equation in $n+1$ dimensions. Thus for $k=0$ we obtain

$$u=4\arctan \alpha e^{\epsilon_0\xi}-\tfrac{1}{2}(1-\epsilon)\pi,\quad \epsilon_0=\pm 1,$$
$$u=2\arccos[\mathrm{dn}(\xi+\alpha,m)]+\tfrac{1}{2}(1+\epsilon)\pi,\quad 0<m<1, \tag{5.14}$$
$$u=2\arccos\left[\mathrm{cn}\left(\frac{\xi+\alpha}{m},m\right)\right]+\tfrac{1}{2}(1+\epsilon)\pi,\quad 0<m<1,$$
$$\alpha=\text{const},\quad \epsilon=\pm 1.$$

These solutions are known for $n=1$ and $\xi=(x-vt)/\sqrt{1-v^2}$. The novelty is in the fact that ξ can be any of the variables (5.4).

The Painlevé transcendant $P_{\mathrm{III}}(\xi)$ is a solution of the sine-Gordon equation for ξ as in (5.6), $k=1$.

For $k\geqslant 2$, (5.8) is not of Painlevé type and will hence have moving critical points. Following a known procedure,[11] we search for a solution of the form

$$y=2\sqrt{\epsilon}(\xi-\xi_0)^{-1}+a_1+a_2(\xi-\xi_0)$$
$$+b_2(\xi-\xi_0)\ln(\xi-\xi_0)+\sum_{r=2}^{\infty}a_{r+1}(\xi-\xi_0)^r+\cdots \tag{5.15}$$

in the neighborhood of a singularity ξ_0 (the dots stand for terms involving further logarithms). Substituting (5.15) into Eq. (5.8), we find

$$a_1=-k\sqrt{\epsilon}/\xi_0,$$
$$b_2=2k(k-1)/3\sqrt{\epsilon}\xi_0^2 \tag{5.16}$$

with ξ_0 and a_2 arbitrary and a_r expressible in terms of ξ_0, a_2, and k for $r\geqslant 3$. Thus, for $k\geqslant 2$ the general solution has a logarithmic branchpoint at $\xi=\xi_0$ and is hence not of the Painlevé type.

According to the Painlevé conjecture,[3] this would imply that the sine–Gordon equation in $M(n,1)$ is not integrable for $n\geqslant 2$ [since, e.g., for $n=2$, k in (5.5) can be equal to 0, 1, or 2].

Equation (5.2) for other functions $F(u)$ can be treated in a completely analogous way.

B. Partial differential equations in two variables

For definiteness, let us restrict ourselves to the nonlinear Klein–Gordon equation

$$\Box u=F(u,(\nabla u)^2). \tag{5.17}$$

Let us first look at the codimension-2 variables in $M(2,1)$, i.e., at Table III.

The degenerate case no. 3 in Table III leads to an ODE. Case no. 1 is an elliptic PDE in canonical form. The entries nos. 2, 4, ... ,7 are all directly recognizable as hyperbolic equations on the entire (ξ_1,ξ_2) surface. Equations 2, 4, and 5 are already in canonical form. Number 6 can be simplified by putting

$$\alpha=\xi_1=x_0+x_1,$$
$$\beta=\xi_1/\xi_2^2=(x_0+x_1)/(x_0^2-x_1^2-x_2^2). \tag{5.18}$$

We then obtain the reduced equation

$$u_{\alpha\beta}+(1/\alpha)u_\beta$$
$$=-(1/2\beta^2)F(u,-2\beta^2u_\alpha u_\beta). \tag{5.19}$$

Number 7 is hyperbolic; to reduce it to canonical form, we would have to solve transcendental equations for the characteristics.

For no. 8 the discriminant of the reduced equation is $\Delta(\xi_1,\xi_2)=4(1-a^2/\xi_2^2)$. The equation

$$(1-a^2/\xi_2^2)u_{\xi_1\xi_1}-u_{\xi_2\xi_2}-(1/\xi_2)u_{\xi_2}$$
$$=F(u,(1-a^2/\xi_2^2)u_{\xi_1}^2-u_{\xi_2}^2) \tag{5.20}$$

is hyperbolic for $\xi_2^2>a^2$, elliptic for $\xi_2^2<a^2$, and parabolic on the circle $\xi_2^2=a^2$.

For no. 9 the discriminant is $\Delta(\xi_1,\xi_2)=16\xi_1$. Thus, the equation

$$-u_{\xi_1\xi_1}+4\xi_1u_{\xi_2\xi_2}=F(u,-u_{\xi_1}^2+4\xi_1u_{\xi_2}^2) \tag{5.21}$$

is hyperbolic for $\xi_1>0$, elliptic for $\xi_1<0$, and parabolic for $\xi_1=0$.

In $M(3,1)$ we have altogether 23 types of codimension-2 symmetry variables, listed in Table IV. Of these three, namely, nos. 3, 13, and 18 lead to ODE's. Seven systems, namely, 1, 10, 11, 15, 16, 19, and 20 lead to elliptic equations. Notice that systems 1, 10, and 11 correspond to a reduction of $M(3,1)$ to $E(3)$. Systems 15 and 16 are both of the form (4.49); 19 and 20, of the form (4.51′). These degenerate codimension-3 variables also lead to elliptic equations [see (4.50) and (4.52′)].

Hyperbolic type equations are obtained for system no. 2, which really "lives" in an $M(1,1)$ space, systems 4, 5, 6, 7, living in $M(2,1)$, 12, 14, and 17, living in the direct product of a Euclidean and Minkowskian space, and also for systems 21, 22, and 23, living in all of $M(3,1)$.

The only two systems in $M(3,1)$ that lead to equations that are hyperbolic in part of the (ξ_1,ξ_2) plane, elliptic in another and parabolic on the boundary, are systems 8 and 9, for which $\{\xi_1,\xi_2\}$ lies entirely in an $M(2,1)$ subspace.

C. Partial differential equations in three variables

Let us again consider the nonlinear Klein–Gordon equation (5.17), this time in $M(3,1)$, and look through the codimension-3 symmetry variables of Table V.

Coordinates no. 3 lead to an elliptic PDE in two variables only; all the other ones to three-dimensional equations.

Among the reduced three-dimensional equations only one is elliptic, namely, no. 1. Systems, 2, 4, 5, 6, 7, 8, 10, and 11 correspond to hyperbolic equations. For the system 9 the reduced equation is elliptic for $\xi_3^2 < 1$, hyperbolic for $\xi_3^2 > 1$, and elliptic on the surface $\xi_3^2 = 1$. Similarly, for system 12 the reduced equation is parabolic on the surface $\xi_2 = 0$, elliptic for $\xi_2 < 0$, and hyperbolic for $\xi_2 > 0$. Equation 13 is parabolic for $\xi_1 = 0$ and hyperbolic for $\xi_1 \neq 0$.

6. CONCLUSIONS AND OUTLOOK

The results of this article can be briefly summarized as follows:

(1) The equation $H(\Box u,(\nabla u)^2,u) = 0$ in Minkowski space $M(n,1)$ can be reduced to a differential equation in k independent variables $\xi_1, \ldots, \xi_k$ $(1 \leqslant k \leqslant n)$, where $\xi_1, \ldots, \xi_k$ are k functionally independent invariants of a group $G \subset P(n,1)$ having generic orbits of codimension k in $M(n,1)$.

(2) All codimension 1 symmetry variables reducing (1.1) to an ODE are listed in Table II for $M(n,1)$, n arbitrary.

(3) All codimension-2 and -3 symmetry variables, reducing (1.1) to a two- or three-dimensional PDE, respectively, are listed in Tables III, IV, and V for $n = 2$ and 3.

(4) All degenerate codimension-2 symmetry variables in $M(2,1)$ and $M(3,1)$ were found; they reduce (1.1) to an ODE and involve arbitrary functions of a null coordinate $\phi(x_0 + x_1)$ [Eqs. (4.43) and (4.47)].

(5) The ODE's obtained by symmetry reduction are always of the form

$$H(\kappa(u_{\xi\xi} + (k/\xi)u_\xi),\kappa(u_\xi)^2,u) = 0,$$
$$\kappa = \pm 1,0, \quad k = 0,1, \ldots ,n.$$

The entire analysis of Sec. 2 is valid for arbitrary Riemannian or pseudo-Riemannian manifolds, with an appropriate redefinition of the invariant operators $\Box$ and $(\nabla\cdot)^2$:

$$\Box u = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x_k} g^{ik} \sqrt{g} \frac{\partial}{\partial x_i} u,$$
$$(\nabla u)^2 = g^{ik} \frac{\partial u}{\partial x_i} \frac{\partial u}{\partial x_k},$$

where g^{ik} is the metric tensor and $g = |\det g_{ik}|$. It is thus possible to extend the analysis of this paper to nonlinear equations invariant under the isometry groups of such Riemannian or pseudo-Riemannian spaces.

We plan to apply the method of reducing nonlinear PDE's to lower-dimensional equations, developed in this article, to other types of spaces and equations. The question of degenerate symmetry variables and the corresponding reductions involving arbitrary functions will receive a more detailed treatment in the near future.

ACKNOWLEDGMENTS

The authors are indebted to H. Zassenhaus for a very helpful discussion.

[1] A. Barone, F. Esposito, C. I. Magee, and A. C. Scott, Nuovo Cimento **1**, 227 (1971).
[2] A. C. Scott, F. Y. F. Chu, and D. W. McLaughlin, Proc. IEEE **61**, 1443 (1973).
[3] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[4] R. L. Anderson and N. H. Ibragimov, *Lie–Bäcklund Transformations in Applications* (SIAM, Philadelphia, 1979).
[5] G. L. Lamb, Jr., *Elements of Soliton Theory* (Wiley, New York, 1980).
[6] V. E. Zakharov, S. B. Manakov, S. P. Novikov, and L. P. Pitaevskiĭ, *Teoriya Solitonov* (*Theory of Solitons*) (Nauka, Moscow, 1980).
[7] P. Painlevé, Acta Math. **25**, 1 (1902).
[8] B. Gambier, Acta Math. **33**, 1 (1910).
[9] E. Hille, *Ordinary Differential Equations in the Complex Domain* (Wiley, New York, 1976).
[10] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[11] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715, 1006 (1980).
[12] (a) S. Lie, Math. Ann. **25**, 71 (1885); (b) S. Lie, *Vorlesungen über Differentialgleichungen mit bekannten infinitesimalen Transformationen* (Teubner, Leipzig, 1891) (reprinted by Chelsea, New York, 1967).
[13] S. Lie and G. Sheffers, *Vorlesungen über Continuierliche Gruppen mit geometrischen und anderen Anwendungen* (Teubner, Leipzig, 1893) (reprinted by Chelsea, New York, 1967).
[14] M. E. Vessiot, Ann. Sci. Ec. Norm. Sup. **10**, 53 (1983).
[15] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations* (Springer-Verlag, New York, 1974).
[16] W. F. Ames, *Nonlinear Partial Differential Equations in Engineering* (Academic, New York, 1965 and 1972), Vols. 1 and 2.
[17] (a) L. V. Ovsyannikov, *Gruppovoĭ analiz differentsial'nykh uravneniĭ* (*Group Theoretical Analysis of Differential Equations*) (Nauka, Moscow, 1978); (b) N. Kh. Ibragimov, *Gruppovye svoĭstva nekotorykh differentsial'nykh uravneniĭ* (*Group Theoretical Properties of Some Differential Equations*) (Nauka, Novosibirsk, 1967).
[18] A. M. Grundland, J. Harnad, and P. Winternitz, KINAM Rev. Fis. **4**, 333 (1982).
[19] C. P. Boyer, R. T. Sharp, and P. Winternitz, J. Math. Phys. **17**, 1439 (1976).
[20] J. Harnad and P. Winternitz, J. Math. Phys. **23**, 517 (1982).
[21] J. Harnad and L. Vinet, Phys. Lett. B **6**, 589 (1978).
[22] (a) J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **21**, 2719 (1980); (b) J. Harnad, S. Shnider, and L. Vinet, Pitman Research Notes Math. **32**, 219 (1979); (c) J. Harnad, S. Shnider, and L. Vinet, J. Math. Phys. **20**, 931 (1979).
[23] J. Harnad, S. Shnider, and J. Tafel, Lett. Math. Phys. **4**, 107 (1980).
[24] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[25]F. J. Ernst, Phys. Rev. **167**, 1175 (1968).
[26]W. M. Kinnersley, in *Lecture Notes in Physics*, **135** (Springer–Verlag, Berlin, 1980), p. 432.
[27]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1597 (1975).
[28]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 1439 (1976).
[29]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **16**, 1615 (1975).
[30]J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, Can. J. Phys. **54**, 986 (1976).
[31]N. Jacobson, *Lie Algebras* (Dover, New York, 1962).
[32]H. Zassenhaus, *Lie Groups, Lie Algebras and Representation Theory* (Les Presses de l'Université de Montréal, Montréal, 1981).
[33]J. Patera, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **15**, 1932 (1974).
[34]P. Appell and E. Lacour, *Principes de la théorie des fonctions elliptiques et applications* (Gauthier-Villars, Paris, 1922).
[35]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[36]B. M. McCoy, C. A. Tracy, and T. T. Wu, J. Math. Phys. **18**, 1058 (1977).
[37]M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978).

Representation of solutions to Helmholtz's equation

A. G. Ramm
Mathematics Department, Kansas State University, Manhattan, Kansas 66506

(Received 14 December 1982; accepted for publication 15 April 1983)

It is proved that any potential of a single layer v is identically equal to a potential of a double layer w in the bounded domain, $\mathscr{D}$, and a necessary and sufficient condition for $v \equiv w$ in $\Omega = \mathbb{R}^3 \setminus \mathscr{D}$, the exterior domain, is given.

PACS numbers: 02.30.Jr

I. INTRODUCTION

Let $\mathscr{D} \subset \mathbb{R}^3$ be a bounded domain with a smooth closed boundary Γ (an obstacle) and $\Omega = \mathbb{R}^3 \setminus \mathscr{D}$ be the exterior domain. Let

$$v(\sigma) = \int_\Gamma g(x,t)\sigma\, dt, \quad w(\mu) = \int_\Gamma \frac{\partial g(x,t)}{\partial n_t}\mu(t)dt,$$
$$g = \frac{\exp(ik|x-y|)}{4\pi|x-y|}, k>0; \tag{1}$$

n_t is the exterior unit normal to Γ at point t. The following questions are discussed in this paper: given a potential $v(\sigma)$, can one find a potential $w(\mu)$ such that $v \equiv w$ in $\mathscr{D}$? $v \equiv w$ in Ω?

Since the potentials solve Helmholtz's equation in $\mathscr{D}$ and in Ω, the above question is connected with the representation of solutions to Helmholtz's equation. Some basic properties of the potentials and some results on the representation of solutions to Helmholtz's equation are given in Ref. 1. These questions are also of interest in the singularity and eigenmode expansions methods.[2] Below, $N(A) = \{f: Af = 0\}$ denotes the nullity of a linear operator A, $(u,v) = \int_\Gamma u\bar{v}\, ds, Q\sigma = \int_\Gamma g(s,t)\sigma(t)dt, H^q = W^{2,q}(\Gamma)$ is the Sobolev space.

II. RESULTS

Theorem 1: For any $w(\mu)$ $[v(\sigma)]$, there exists a $v(\sigma)$ $[w(\mu)]$ such that $v(\sigma) \equiv w(\mu)$ in $\mathscr{D}$. The $v(\sigma)$ $[w(\mu)]$ is uniquely defined.

Theorem 2: A necessary and sufficient condition for a $v(\sigma)$ to be identically equal to a $w(\mu)$ in Ω is $(v, \nu_j) = 0, 1 \leqslant j \leqslant r'$, where $\{\nu_j\}$ forms a basis of $N(I+A)$, $Af \equiv \int_\Gamma (\partial g(s,t)/\partial n_s) f(t)dt$. A necessary and sufficient condition for a $w(\mu)$ to be identically equal to a $v(\sigma)$ in Ω is $(\mu, \sigma_j) = 0, 1 \leqslant j \leqslant r$, where $\{\sigma_j\}$ forms a basis of $N(I-A)$.

Corollary 1: If a problem $(\nabla^2 + k^2)u = 0$ in $\mathscr{D}$, $k \geqslant 0$, $u^+ = f$ is solvable, then the solution can be represented as $u = v(\sigma)$ and also as $u = w(\mu)$.

III. PROOFS

A. Auxiliary results

Let us denote by $v^\pm$, $\partial v^\pm/\partial n$ the limit values on Γ from the interior $(+)$ and exterior $(-)$ of a function and its normal derivative. It is well known[1] that

$$\frac{\partial v^\pm}{\partial n} = \frac{A\sigma \pm \sigma}{2}, \quad w^\pm = \frac{A'\mu \mp \mu}{2}, \tag{2}$$

where the operator A is defined in Theorem 2, and $A'\mu = \int_\Gamma (\partial g(s,t)/\partial n_t)\mu(t)dt$. Note that $A^* = \overline{A}'$, where the star denotes the adjoint in operator $H = L^2(\Gamma)$ and the bar denotes complex conjugation.

Lemma 1: If $(\nabla^2 + k^2)u = 0$ in Ω, $u = 0$, or $\partial u^-/\partial n = 0$ on Γ, $k > 0$, and $|x|(\partial u/\partial|x| - iku) \to 0$ as $|x| \to \infty$, then $u \equiv 0$.

Lemma 2: Let the problem

$$(\nabla^2 + k^2)u = 0 \quad \text{in} \quad \mathscr{D}, \quad \frac{\partial u^+}{\partial n} = 0 \tag{3}$$

have r' linearly independent solutions u_j, $1 \leqslant j \leqslant r'$. Then the equation

$$A'\mu + \mu = 0 \tag{4}$$

has precisely r' linearly independent solutions $\mu_j = u_j^+$, $1 \leqslant j \leqslant r'$, $u_j = w(\mu_j)$. Let the problem

$$(\nabla^2 + k^2)u = 0 \text{ in } \mathscr{D}, \quad u^+ = 0 \tag{5}$$

have r linearly independent solutions ϕ_j, $1 \leqslant j \leqslant r$. Then the equation

$$A\sigma - \sigma = 0 \tag{6}$$

has precisely r linearly independent solutions $\sigma_j = \partial \phi_j^+/\partial n$, $\phi_j = v(\sigma_j)$.

Lemma 3: If (5) has only the trivial solution, then $Q: H^q \to H^{q+1}$ is an isomorphism.

Lemma 4: If (5) has r linearly independent solutions, then equation $Q\sigma = f$ is solvable iff $(f, \bar{\sigma}_j) = 0$, $1 \leqslant j \leqslant r$.

Lemmas 1 and 2 are proved, e.g., in Ref. 1. Lemmas 3 and 4 are proved in the Appendix.

B. Proof of Theorem 1

(i) Assume first that (*) problem (5) has only the trivial (zero) solution. If $v(\sigma) = w(\mu)$ in $\mathscr{D}$, then (2) implies that (**) $A'\mu - \mu = 2v(\sigma)$. If $w(\mu)$ is given, then the above equation is an equation for σ. This equation is uniquely solvable (by Lemma 3) because of (*). If σ is its solution, then the corresponding $v(\sigma)$ satisfies (***) $v^+(\sigma) = w^+(\mu)$ and, again by (*), $v(\sigma) \equiv w(\mu)$ in $\mathscr{D}$. If $v(\sigma)$ is given, then (**) is an equation for μ. If this equation is solvable, then the corresponding $w(\mu)$ satisfies (***), and, by (*), $v(\sigma) \equiv w(\mu)$ in $\mathscr{D}$. It remains to be proved that (**) is solvable for μ. By Fredholm's alternative, it is so if $(v(\sigma), \nu_j) = 0$, where ν_j are all linearly independent solutions to the equation $\overline{A}\nu - \nu = 0$. [Notice that $\overline{A} = (A')^*$]. By Lemma 2, the functions $\nu_j = \bar{\sigma}_j = \overline{\partial \phi_j^+/\partial n}$,

 0022-2488/84/040807-03$02.50

$v^{\pm}(\sigma_j) = 0 = v^{\pm}(\bar{v}_j)$. Therefore $(v^+(\sigma),v_j) = (\sigma, v^+(\bar{v}_j)) = 0$. (Or $\int_\Gamma v^+(\sigma)(\partial\phi_j^+/\partial n)ds = \int_\Gamma (\partial v^+/\partial n)\phi_j^+ \, ds = 0$). Thus (**) is solvable. In the proof of the solvability of (**), the assumption (*) was not used.

(ii) Assume now that (5) has $r > 0$ linearly independent solutions. If $v(\sigma)$ is given then $w(\mu)$, where μ solves (**), satisfies (***). Therefore $w(\mu) = v(\sigma) + \Sigma_{j=1}^r c_j\phi_j(x)$ in $\mathscr{D}$, where $c_j = \text{const}$. By Lemma 2, Eq. (6) has precisely r linearly independent solutions. Thus the equations $A^*\bar{\eta} - \bar{\eta} = 0$,

$$A'\eta - \eta = 0 \tag{7}$$

have precisely r linearly independent solutions. Let η_j, $1 \leqslant j \leqslant r$, be r such solutions of (7), $w(\eta_j)$ be the corresponding potentials. Then $w(\eta_j) = \phi_j$ are r linearly independent solutions to (5). Therefore $v(\sigma) = w(\mu - \Sigma_{j=1}^r c_j\eta_j)$. Thus if $v(\sigma)$ is given, one can find a $\mu' = \mu - \Sigma_{j=1}^r c_j\eta_j$ such that $w(\mu') \equiv v(\sigma)$ in $\mathscr{D}$. This μ is uniquely defined by the requirement $w(\mu') \equiv v(\sigma)$ in $\mathscr{D}$.

Consider now the case when $w(\mu)$ is given and $v(\sigma)$ is to be found such that $w(\mu) \equiv v(\sigma)$ in $\mathscr{D}$. In this case (**) is an equation for σ of the type $Q\sigma = f \equiv (A'\mu - \mu)/2$. By Lemma 4, this equation is solvable iff $(f,\bar{\sigma}_j) = 0$, where σ_j, $1 \leqslant j \leqslant r$ solve (6). One has $(A'\mu - \mu, \bar{\sigma}_j) = (\mu, \overline{(A - I)\sigma_j}) = 0$. Thus (**) is solvable for σ. As above, the requirement $w(\mu) \equiv v(\sigma)$ in $\mathscr{D}$ defines σ uniquely. Theorem 1 is proved.

C. Proof of Theorem 2

(i) Assume that (5) has only the trivial solution. Given $w(\mu)$, one can find the unique σ from the equation

$$w^-(\mu) = (A'\mu + \mu)/2 = v^-(\sigma). \tag{8}$$

Because of Lemma 1, $w(\mu) \equiv v(\sigma)$ in Ω. Suppose now that $v(\sigma)$ is given. Then μ is to be found from (8). If this equation is solvable, then as above, $w(\mu) \equiv v(\sigma)$ in Ω. Equation (8) is solvable for μ iff $(v^-(\sigma),v_j) = 0, 1 \leqslant j \leqslant r$, where $\bar{A}v_j + v_j = 0$. Notice that $A\bar{v}_j + \bar{v}_j = 0$, $v(\bar{v}_j)$ solve (3). If (3) has only the trivial solution, then (8) is uniquely solvable for μ, and $v(\sigma) \equiv w(\mu)$ in Ω. If (3) has r' linearly independent solutions, then (4) has r' linearly independent solutions, and, by Fredholm's alternative, the equation $\bar{A}v + v = 0$ has r' linearly independent solutions: v_j, $v(v_j)$ solve (3), $v(v_j) = w(\mu_j)$ in $\mathscr{D}$. Therefore

$$\begin{aligned}(v^-(\sigma),v_j) &= (\sigma, v^-(\bar{v}_j)) = (\sigma, w^+(\mu_j)) \\ &= (\sigma, w^+(\mu_j)) = (\sigma, \overline{\tfrac{1}{2}(A' - I)\mu_j}) \\ &= -(\sigma, \bar{\mu}_j).\end{aligned}$$

Thus $(\sigma,\bar{\mu}_j) = 0$, $1 \leqslant j \leqslant r'$ is a necessary and sufficient condition for a potential $v(\sigma)$ to be identically equal to $w(\mu)$ in Ω in the case when (3) has r' linearly independent solutions.

(ii) Assume that (5) has r linearly independent solutions. Given $w(\mu)$ one has to find σ from (8). By Lemma 4, (8) is solvable for σ iff $(A'\mu + \mu, \bar{\sigma}_j) = 0$, $1 \leqslant j \leqslant r$, i.e. $(\mu, \overline{(A + I)\sigma_j}) = 0$, or $(\mu,\bar{\sigma}_j) = 0$, $1 \leqslant j \leqslant r$, because $\sigma_j = A\sigma_j$. If $v(\sigma)$ is given, the analysis does not depend on the assumption about the number of the linearly independent solutions of (5) and is given above in part (i). In the case when problem (5) [(3)] has a nontrivial solution and $w(\mu)$ [$v(\sigma)$] is given, the density $\sigma(\mu)$ is not uniquely defined. But the difference between two densities σ and $\tilde{\sigma}$ (μ and $\tilde{\mu}$) generates $v \equiv 0$ in Ω ($\mu \equiv 0$ in Ω) because of Lemma 1.

Proof of Corollary 1: If the Dirichlet problem in Corollary 1 is solvable, then its solution by Green's formula is of the form $u(x) = v(\sigma) + w(\mu)$. By Theorem 1, $v(\sigma)$ [$w(\mu)$] can be substituted by $w(\tilde{\mu})$ [$v(\tilde{\sigma})$]. Thus $u = w(\mu + \tilde{\mu}) = v(\sigma + \tilde{\sigma})$.

APPENDIX

A. Proof of Lemma 3

If problem (5) has only the trivial solution then $N(Q) = \{0\}$ and Q is injective. The fact that $Q:H^q \to H^{q+1}$ is bounded follows from the known results about the smoothness of solutions to Helmholtz's equation. Indeed, if $\sigma \epsilon H^q$, then

$$\frac{\partial v^+}{\partial n} \epsilon H^q, \quad \nabla v \epsilon H^{q+1/2}(D), \quad v \epsilon H^{q+3/2}(D), \quad v^{\pm} \epsilon H^{q+1}.$$

Here, the well-known trace theorem is used: $u \epsilon H^q(D) \Rightarrow u^+ \epsilon H^{q-1/2}(\Gamma)$. It remains to be proved that Range $Q \equiv R(Q) = H^{q+1}$. Take any $f \epsilon H^q$. Solve the problem

$$(\nabla^2 + k^2)v = 0 \quad \text{in} \quad \Omega,$$
$$\frac{\partial u^-}{\partial n} = f, \quad |x|\left(\frac{\partial v}{\partial |x|} - ikv\right) \to 0, \quad |x| \to \infty. \tag{A1}$$

The solution exists, is unique, and can be found as $v(\sigma)$ because problem (5) [and therefore Eq. (6)] has only the trivial solution, and (*) $A\sigma - \sigma = 2f$. Since A is a smoothing operator, σ has the same smoothness as f, i.e., $\sigma \epsilon H^q$. Therefore $Q\sigma = v^- \epsilon H^{q+1}$. It is now easy to show that $R(Q) = H^{q+1}$. Take any $h \epsilon H^{q+1}$ and solve the exterior Dirichlet problem with the data $u^- = h$. Calculate $\partial u^-/\partial n \equiv f \epsilon H^q$. Find (the unique) σ from (*). By the uniqueness theorem (Lemma 1), $v(\sigma) = u$ in Ω, $Q\sigma = v^- = u^- = h$. Thus $Q:H^q \to H^{q+1}$ is a linear, injective, surjective, and continuous mapping. From the Banach theorem (about inverse operator) it follows that $Q^{-1}:H^{q+1} \to H^q$ is continuous. Lemma 3 is proved.

B. Proof of Lemma 4

If $Q\sigma = f$, $Q\sigma_j = 0$, then $(f,\bar{\sigma}_j) = (Q\sigma,\bar{\sigma}_j) = (\sigma, \overline{Q\sigma_j}) = 0$. This proves the necessity. The operator $Q:H^q \to H^{q+1}$ is a Fredholm operator and ind $Q = 0$, ind $Q \equiv$ index $Q \equiv \dim N(Q) - \dim N(Q^*) = \dim N(Q) - \operatorname{codim} R(Q)$. Therefore there are $r = \dim N(Q)$ necessary and sufficient conditions of the type $(f,h_j)_{q+1} = 0$, $1 \leqslant j \leqslant r$, where $(u,v)_q$ is the inner product in H^q. Suppose that the r conditions $(f,b_j)_0 = 0$ are necessary for $f \epsilon R(Q)$. Since $H^q \subset H^0$, $q > 0$, and $||\cdot||_0 \leqslant ||\cdot||_q$, one has $(f,b_j)_0 = (f,a_j)_q$. Let $L_r = \operatorname{span}(h_1,...,h_r)$, $M_r = \operatorname{span}(a_1,...,a_r)$. Then $L_r = M_r$. Indeed, suppose there exists $a_m \notin L_r$. Then the number of the necessary conditions for f to belong to $R(Q)$ would be at least $r + 1$, namely, $(f,h_j)_q = 0$, $1 \leqslant j \leqslant r$, and $(f,a_m)_q = 0$. This contradiction proves that $L_r = M_r$. Thus the conditions $(f,\bar{\sigma}_j) = 0$, $1 \leqslant j \leqslant r$, where $\{\sigma_j\}_{j=1}^r$ form a basis of $N(Q)$, are necessary and sufficient for $f \epsilon R(Q)$.

C. Solution of the scattering problem via an indetermined equation

Consider the problem

$$(\nabla^2 + k^2)u = 0 \quad \text{in} \quad \Omega, \quad k > 0,$$
$$u^- = f, \quad |x|(\partial u/\partial|x| - iku) \to 0, \quad |x| \to \infty. \qquad \text{(A2)}$$

The existence and uniqueness of the solution to this problem were studied extensively and a complete analysis is given in Ref. 1. It follows from the arguments in Sec. III that if (5) has only the trivial solution, then (A2) is (uniquely) solvable by a potential of the single layer $u = v(\sigma)$, while if (3) has only the trivial solution then (A2) is (uniquely) solvable by a potential of the double layer $u = w(\mu)$. Since it is known[1] that (A2) is (uniquely) solvable for any $k \geqslant 0$, one sees from Green's formula that the representation (*)$u = v(\sigma) + w(\mu)$ holds for all $k \geqslant 0$. The aim of this Section is to give a short proof of the existence of the solution to (A2) following Ref. 3, p. 98. The idea in Ref. 3 is very elegant. Let us look for a solution to (A2) of the form (*). Then

$$Q\sigma + (A'\mu + \mu)/2 = f, \quad \mu + A'\mu + 2Q\sigma = 2f. \quad \text{(A3)}$$

This is an equation in $H = L^2(\Gamma)$ for the two unknown functions μ and σ.

Lemma A 1[3]: Let A and B be compact operators on a Hilbert space H, and $N(I + A^*) \cap N(B^*) = \{0\}$. Then the equation $(I + A)\mu + B\sigma = f$ is solvable for any $f \epsilon H$.

Let us postpone a proof of this lemma and show that Eq. (A3) is solvable. In our case, $B \to 2Q$, $A \to A'$. Thus one should check that $N(I + A'^*) \cap N(Q^*) = \{0\}$. Assume that $\sigma \epsilon N(I + A'^*)$, $\sigma \neq 0$, and consider $v(\sigma)$. By assumption, $\bar{v}^-(\sigma) = 0$. Thus $\bar{v}(\sigma) \equiv 0$ in Ω. On the other hand, $A'^* = \bar{A}$ and $\sigma + \bar{A}\sigma = 0$ implies that $\partial \bar{v}^+/\partial n = 0$. Since $(\nabla^2 + k^2)\bar{v} = 0$ in D, $\bar{v}^+ = \partial \bar{v}^+/\partial n = 0$, one concludes that $\bar{v} \equiv 0$ in D. Thus from (2) it follows that $\sigma = 0$. From Lemma A1 it now follows that Eq. (A3) is solvable for any $f \epsilon H$.

Proof of Lemma A 1: Let $T \equiv I + A$. Since A is compact, $R(T)$ is closed and $\dim N(T^*) = \operatorname{codim} R(T) < \infty$. It is clear that a necessary and sufficient condition for the dense solvability of the equation $T\mu + B\sigma = f$ is $N(T^*) \cap N(B^*) = \{0\}$. The dense solvability means the solvability for all f in a dense set of H. It remains to be proved that this condition is sufficient for everywhere solvability (solvability for any $f \epsilon H$). Let P be a projection onto $R(T)$ in H. Then

$$T\mu + PB\sigma = Pf \qquad \text{(A4)}$$
$$(I-P)B\sigma = (I-P)f, \qquad \text{(A5)}$$

If (A5) is solvable then (A4) and

$$T\mu + B\sigma = f \qquad \text{(A6)}$$

are solvable. Indeed, let σ solve (A5) and let $T\mu = P(f - B\sigma)$. Since the right-hand side belongs to $R(T)$, this equation has a solution μ. The pair (μ, σ) solves (A6):
$T\mu + B\sigma = Pf - PB\sigma + B\sigma = Pf + (I-P)f = f$. Let us show that (A5) is everywhere solvable if $N(T^*) \cap N(B^*) = \{0\}$. This condition implies dense solvability of (A6) and therefore dense solvability of the equivalent to (A6), system (A4), (A5); in particular, dense solvability of (A5). But (A5) is an equation in the finite-dimensional space $(I - P)H$. If such an equation is densely solvable, then it is everywhere solvable. But then, as was shown above, Eqs. (A4) and (A6) are everywhere solvable. Lemma A1 is proved.

Remark. All we used in the proof are the following assumptions: (i) $R(T)$ is closed, and (ii) $\dim N(T^*) < \infty$.

[1] V. Kupradze, *Randwertaufgaben der Schwingungs theorie und integralgleichungen* (VEB Verlag, Berlin, 1956).
[2] A. G. Ramm, *Theory and applications of some new classes of integral equations* (Springer-Verlag, New York, 1980); "Mathematical foundations of the singularity and eigenmode expansion methods," J. Math. Anal. Appl. **86**, 562–591 (1982).
[3] S. G. Krein, *Linear equation in a Banach space* (Fizmatgiz, Moscow, 1971) (in Russian).

On the solution of the two-dimensional Helmholtz equation

G. S. Singh and L. S. Kothari
Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India

(Received 17 March 1983; accepted for publication 6 May 1983)

The solution of the two-dimensional Helmholtz equation has been obtained in the annular region having eccentric circular boundaries. It is shown that in the limit of eccentricity zero, the solution reduces to that corresponding to the case of concentric circular boundaries.

PACS numbers: 02.30.Jr

I. INTRODUCTION

The two-dimensional Helmholtz equation appears in many physical problems, like acoustical waveguides, vibration of membranes, diffusion of neutrons in thermal nuclear reactors under a one-group approximation, etc. This paper deals with the solution of this equation in an annular region enclosed between a circular outer boundary and an eccentric circular inner boundary. This would apply to the case of a cylindrical reactor with an eccentric cylindrical control rod, vibrations of a membrane having eccentric circular boundaries, and waveguide problems with eccentric-annular cross sections.

The approximate solution for the control rod problem in reactor physics has been obtained[1] for the special case when the eccentricity is large and the cross section of the control rod is small as compared to the eccentricity. The numerical solution of the waveguide problems has been obtained using various techniques[2] which require a large size digital computer. Recently the vibrations of membranes for the aforesaid geometry has been discussed by Nagaya[3] by using the Fourier expansion method. We follow essentially Nagaya's approach and obtain a solution of the two-dimensional Helmholtz equation in a mathematically compact and computationally simple form.

II. SOLUTION

We consider the region bounded by the circles of radii a and b centered at O and O', respectively, as shown in Fig. 1. Let the centers of the two circles be separated by a distance d. Let the polar coordinates of any point P in the annular region be (r,θ) and (r',θ') with respect to O and O'.

The two-dimensional Helmholtz equation in the polar coordinates referred to the origin at O is

$$\left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+k^2\right]\psi(r,\theta)=0. \tag{1}$$

Here k is the unknown parameter to be determined along with the function $\psi(r,\theta)$.

The symmetric part of the solution of Eq. (1) can be written in the form

$$\psi_s(r,\theta)=\sum_{m=0}^{\infty}\epsilon_m\left[A_mJ_m(kr)+B_mY_m(kr)\right]\cos m\theta, \tag{2}$$

where A_m and B_m are the constants of integration, $J_m(kr)$ and $Y_m(kr)$ are the Bessel functions of the first and the second kind, respectively, and

$$\epsilon_m=2-\delta_{m,0} \tag{3}$$

with $\delta_{m,0}$ as the Kronecker delta function.

The application of the boundary condition

$$\psi_s(a,\theta)=0 \tag{4}$$

for all values of θ gives

$$\psi_s(r,\theta)=\sum_{m=0}^{\infty}\epsilon_mA_m\left[J_m(kr)-F_m(ka)Y_m(kr)\right]\cos m\theta, \tag{5}$$

where we have written

$$F_m(ka)=J_m(ka)/Y_m(ka). \tag{6}$$

The variables r and θ can be expressed in terms of r' and θ' by simple coordinate transformations. However, the use of the boundary condition

$$\psi_s(r,\theta)\big|_{r'=b}=0 \tag{7}$$

for all values of θ' in Eq. (5) does not lead to any appropriate equation for determining k. Hence we first expand[4] the function $\psi_s(r,\theta)$ into a cosine Fourier series in θ':

$$\psi_s(r,\theta)=\frac{1}{2}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\epsilon_n\epsilon_mA_mQ_{mn}\cos n\theta', \tag{8}$$

where the Fourier coefficients are given by

$$Q_{mn}(r')=\frac{2}{\pi}\int_0^{\pi}\left[J_m(kr)-F_m(ka)Y_m(kr)\right]\times\cos m\theta\cos n\theta'\,d\theta'. \tag{9}$$

Now the use of the boundary condition (7) in Eq. (8) yields simultaneous equations in A_m's and the nontrivial solution of these equations requires that

$$\det|Q_{mn}(r'=b)|=0. \tag{10}$$

This is the basic equation[3] whose solution would give

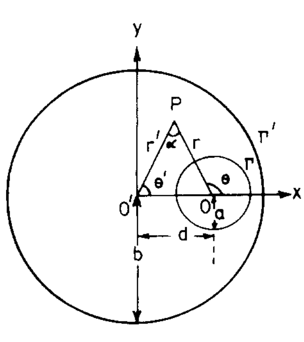


FIG. 1. The annular region enclosed between the circular boundaries Γ and Γ'.

various values of k so the corresponding A_m's could then be determined. We find that in order to determine Q_{mn} defined by Eq. (9) as such one has to proceed numerically since r and θ are functions of r' and θ'. Here we show that by using the addition theorem of the Bessel functions one gets an analytical form for Q_{mn}.

We have to evaluate Q_{mn} for $r' = b$. Moreover, in the problems of interest we will always have the condition that the radius of the external circle is greater than the eccentricity, i.e., $b > d$. Now since $\theta = \alpha + \theta'$ (from triangle POO' in Fig. 1), we make use of the expressions[5]

$$Z_m(kr) \begin{smallmatrix}\cos\\ \sin\end{smallmatrix} m\alpha = \sum_{p=-\infty}^{\infty} Z_{m+p}(kb) J_p(kd) \begin{smallmatrix}\cos\\ \sin\end{smallmatrix} p\theta', \quad (11)$$

where $Z_m(x)$ may be $J_m(x)$ or $Y_m(x)$.

Combining Eqs.(9) and (11), we obtain

$$Q_{mn} = \frac{2}{\pi} \sum_{p=-\infty}^{\infty} [J_{m+p}(kb) - F_m(ka) Y_{m+p}(kb)] \times J_p(kd) \int_0^\pi \cos(m+p)\theta' \cos n\theta' \, d\theta'. \quad (12)$$

The evaluation of the integration appearing in Eq. (12) finally yields

$$Q_{mn} = [J_n(kb) - F_m(ka) Y_n(kb)] \times [J_{n-m}(kd) + (-)^m J_{n+m}(kd)], \quad (13)$$

with m and n taking the integral values from 0 to ∞.

Now let us consider the antisymmetric part of the solution of Eq. (1) given by

$$\psi_a(r,\theta) = 2 \sum_{m=0}^{\infty} [C_m J_m(kr) + D_m Y_m(kr)] \sin m\theta, \quad (14)$$

where C_m and D_m are the constants of integration.

Proceeding in a way exactly similar to the symmetric case but considering the sine Fourier series expansion in θ', the counterpart of Eq. (13) is obtained in the form

$$Q_{mn} = [J_n(kb) - F_m(ka) Y_n(kb)] \times [J_{n-m}(kd) + (-)^{m+1} J_{n+m}(kd)], \quad (15)$$

with m and n taking the integral values from 1 to ∞.

III. DISCUSSION

It can be observed that the expressions given by Eqs. (13) and (15) differ in the ranges of m and n. It apparently seems that the signs appearing before $J_{n+m}(kd)$ in the second square brackets are also different. But if we consider the full matrices in two cases, then the signs under consideration in the expressions for the corresponding elements of two matrices become the same as m starts from 0 in the symmetric case and from 1 in the antisymmetric case.

Now let us consider the limiting behavior of our solutions in the special case when the eccentricity d goes to zero. In this case we must have

$$Q_{mn} = Q_{mn} \delta_{mn} \quad (16)$$

so that Eq. (10) does not vanish identically. Hence the condition for the solution of Eq. (1) simply becomes

$$J_m(kb) - F_m(ka) Y_m(kb) = 0, \quad (17)$$

which coincides with the equation obtained for solution in the concentric case.[6]

We wish to mention that we have firstly solved Eq. (1) with the origin at the center of the smaller circle and then transferred the solution to the origin at the center of the larger circle. If we could have proceeded the other way, $Y_m(x)$ would have been dropped from the very beginning as this function is not finite at $x = 0$. But the solution could not have been obtained in the closed analytical form that we have got. Moreover the limiting case ($d = 0$) of this solution would have corresponded to the system with only one circular boundary.

[1] S. Glasstone and M. C. Edlund, *The Elements of Nuclear Reactor Theory* (MacMillan, London, 1952), Chap. XI.
[2] See references in Ref. 3.
[3] K. Nagaya, J. Sound Vib. **50**, 545 (1977).
[4] The factor $\frac{1}{2}$ which appears in Eq. (8) had been missed in Eq. (10) of Ref. 3. However, this does not affect the discussions of that paper.
[5] G. N. Watson, *Treatise on the Theory of Bessel Functions* (Cambridge U. P., London, 1958), Chap. XI.
[6] See, for example, for the concentric control rod problem, Ref. 1, p. 318.

Convergence of a crossing-symmetric perturbation series for the four-point vertex

E. D. Davis[a] and W. D. Heiss
Physics Department and Nuclear Physics Research Unit, University of the Witwatersrand, Johannesburg, South Africa

(Received 30 June 1983; accepted for publication 2 December 1983)

A finite radius of convergence is established for a perturbation expansion of the crossing-symmetric vertex function in a discrete model.

PACS numbers: 02.30.Mv, 11.50.Jg

The convergence of perturbation expansions in field theory is not often ranked as a priority among physicists. In quantum electrodynamics the question appears nearly irrelevant, since the success of the low-order terms allows one to adhere to the hope that the expansion is at least asymptotically convergent.[1] In strongly-interacting systems the question is usually not raised[2,3] owing the relative success of partial summations such as the traditional random-phase approximation or its extensions.[4,5] However, when it is necessary to go beyond partial summations, the question whether the relevant physical quantities are regular at zero interaction strength is pertinent to virtually any approximation procedure.

A perturbation expression for a unitary crossing-symmetric vertex function Γ is given by

$$\Gamma = \Gamma^0 + \sum_{n=1}^{\infty} \sum_{m=1}^{N_n} \left(\left\{ \Gamma^0 \sum_{i=s,t,u}{}' (GG)_i \right\}_m^n \Gamma^0 \right), \qquad (1)$$

where Γ^0 is the irreducible part of Γ and G is the one-body Green's function. The significance of the remaining symbols is explained elsewhere.[6] The crossing symmetry of this series is not preserved by any partial summations.[7] It is our aim to find an approximate expression when the simplest possible forms for Γ^0 and G are adopted, i.e., Γ^0, henceforth to be denoted by λ, is both static and label independent, while G contains only one left-hand and one right-hand pole each with diagonal residues. Under these circumstances Eq. (1) is a formal power series in the variable λ. It is known[8] that the perturbation expansion of the exact and therefore crossing symmetric Γ for a soluble model[9] has a finite nonzero radius of convergence. In this paper we report a nonzero radius of convergence for the more general case outlined above.

Denoting by a_n the coefficient of λ^n, it is obviously sufficient to show $|a_n| \leqslant c^n$, where c is some constant. By Eq. (1), a_n is the sum of m_n Feynman diagrams, where, from Ref. 6,

$$m_n = \frac{3^{n-1}}{n-1} \sum_{k=1}^{n-1} \left(\frac{2}{3}\right)^{k-1} \binom{n-1}{k} \binom{n-1}{k-1}.$$

Since m_n is bounded by exponential growth, it has to be shown that any nth-order Feynman diagram in this perturbation series is bounded by exponential growth.

Expressed in Goldstone diagrams[10] an nth-order Feynman diagram is a sum with $n!$ terms. Thus one has to consider its Goldstone diagrams individually and show that, although their number is not bounded by exponential growth, the sum of their moduli is.

Consider the diagram in Fig. 1. When evaluated as a Goldstone diagram, the result for its modulus is

$$E = \left| \left\{ \prod_{k=1}^{r-1} \frac{1}{(\omega_2 - \epsilon_k)^{S_k}} \right\} \left(\frac{1}{s - r(\epsilon - \delta)} \right)^{S_r} \times \left\{ \prod_{k=1}^{r-1} \frac{1}{(\omega_3 - \tilde{\epsilon}_k)^{S_{2r-k}}} \right\} \right|,$$

where S_k is the number of propagations in the k th column of the diagram, $\epsilon_k = k(\epsilon - \delta) + \delta = \tilde{\epsilon}_k + \epsilon + \delta$, ϵ and δ being the pole positions of G, and ω_2, ω_3, and s are appropriate energy variables.[4] To establish the growth of E in terms of n, attention must be focused on ϵ_k, $\tilde{\epsilon}_k$, and the S_k's (which are subject to the constraint $\Sigma_{k=1}^{2r-1} S_k = n - 1$). For fixed values of the energy variables, we find the estimate

$$E \sim \frac{c_0^n}{r^{S_r}} \prod_{k=1}^{r-1} \frac{1}{(k)^{S_k + S_{2r-k}}} \equiv c_0^n \prod_{k=1}^{2r-1} \frac{1}{(P_k)^{S_k}}, \qquad (2)$$

where c_0 is dependent on the energy variables. The prominent feature of Eq. (2) is that, because $S_k \geqslant 1$ for any k, the denominator grows as $(r\,!)^2$ or faster; this growth is entirely due to the pole positions in the expression for the Goldstone diagram.

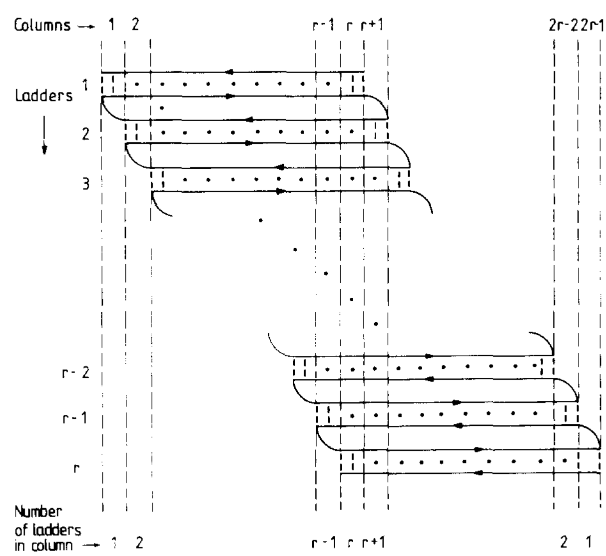


FIG. 1. A typical nth-order Feynman diagram in the perturbation series given in Eq. (1). The lines composed of long thin dashes demarcate the columns into which the diagram may be conveniently divided.

[a] Supported in part by a bursary from C.S.I.R., Pretoria, South Africa.

 0022-2488/84/040812-02$02.50

This observation prompts the conjecture that convergence can be established by exploiting the growth rate of the pole positions in the expressions for Goldstone diagrams. To demonstrate this we will consider the sum S over that class of Goldstone diagrams, corresponding to the Feynman diagram in Fig. 1, for which Eq. (2) is valid. It will be assumed that $r \sim n$ (so that the number of diagrams in S is not bounded by exponential growth) and, for convenience, that each ladder carries the same number of interactions denoted by $q+2$. [Note q is $O(1)$.]

This class of Goldstone diagrams is generated from the one drawn by changing the positions of interactions on any ladder relative to interactions on other ladders without changing either the relative positioning of interactions within a ladder or the number of ladders in any column. Such changes can be specified by stating: (a) the time ordering of the interactions at the ends of the ladders (the "end points" of ladders); (b) the number of interactions on each ladder within each column which intersects it; (c) the relative positioning of all interactions within a column.

Before an expression for S can be written down cognizance must be taken of the following result: each Goldstone diagram for which (a) and (b) above are identical has the same expression denoted by g. The number of such diagrams is $m = \prod_{k=1}^{2r-1} C(x_{i_1k}, x_{i_2k}, \ldots, x_{i_{o_k}k})$, where $C(x_1, \ldots, x_p)$ is the multinomial coefficient, x_{ik} is the number of interactions on the ith ladder within the kth column in each of these diagrams and $i_1, i_2, \ldots, i_{o_k}$ are the indices of those ladders which intersect the kth column. It follows that

$$S = \sum_{(a)} \sum_{(b)} m|g|,$$

where $\Sigma_{(a)}$ and $\Sigma_{(b)}$ are summations over the various possibilities under (a) and (b) above.

To estimate S we insert for $|g|$ Eq. (2). Now it can be shown that, for any permissible values of the x_{i_jk}'s,

$$\frac{C(x_{i_1k}, \ldots, x_{i_{o_k}k})}{(P_k)^{S_k}} \lesssim \frac{c_1^{S_k}}{(P_k)^{R_k}},$$

where c_1 is a constant, R_k is any number of $O(1)$ and the P_k's are defined in Eq. (2). If we combine this result with another, namely that if S_k is $O(1)$ then $C(x_{i_1k}, \ldots, x_{i_{o_k}k}) \lesssim (\tilde{C}_1)^{S_k}$, we realize that to estimate $C(x_{i_1k}, \ldots, x_{i_{o_k}k})/(P_k)^{S_k}$ it is sufficient to consider those distributions of the interactions in which each S_k is $O(1)$. To ensure deducing an upper bound, we consider those distributions of this kind for which Eq. (2) attains its maximum; in these distributions, which occur when the ladders are time ordered as in the diagram considered, $S_k = q+1$ if $k = 1, 2, \ldots, r/2$ or $k = 3r/2, 3r/2+1, \ldots, 2r-1$, otherwise $S_k = 1$. Thus

$$m|g| \leqslant \frac{r}{(r!)^2} \frac{c_2^n}{\{(r/2)!\}^{2q}}, \tag{3}$$

where $c_2 = c_0 \tilde{c}_1$.

Since Eq. (3) is independent of the summations in S, it only remains to find upper bounds to the numbers of terms in the summations, $n_{(a)}$ and $n_{(b)}$. For the sake of brevity, we merely quote the result that $n_{(b)} \leqslant (c_3 r)^{qr}$, c_3 being another constant. On the other hand, $n_{(a)}$ is simply the number of time orderings of the endpoints of the ladders compatible with the restriction that the number of ladders in any column is not changed. Clearly $n_{(a)} = (r!)^2$.

Combining these results we conclude that S is indeed bounded by exponential growth. In obtaining this result the essence of the proof of convergence has been presented. The Feynman diagram considered is typical of those generated in the perturbation series given in Eq. (1). Satisfactory upper bounds for arbitrary sums of Goldstone diagrams of any of these Feynman diagrams are set by adopting appropriate but straightforward modifications of the ideas employed above.

In conclusion we note that this proof of convergence remains valid when we relax the restrictions on Γ^0 and G in that G can have an arbitrary number of poles and Γ^0 can be energy dependent (as the exact expression would be). We emphasize, however, that the crucial prerequisite for our proof is the discreteness of the single-particle spectrum. While it is therefore relevant for models used in nuclear structure calculations, it does not shed light on situations where a continuum occurs.

[1]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Row, Peterson and Company, New York, 1961).

[2]M. B. Johnson and M. Baranger, Ann. Phys. (N.Y.) **62**, 172 (1971); P. F. Bortignon, R. A. Broglia, D. R. Bes, and R. Liotta, Phys. Rep. **30**, 305 (1977).

[3]T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood Academic, New York, 1981).

[4]C. A. Engelbrecht, F. J. W. Hahne, and W. D. Heiss, Ann. Phys. (N.Y.) **104**, 221 (1977).

[5]W. D. Heiss, C. A. Engelbrecht, and F. J. W. Hahne, Nucl. Phys. A **289**, 286 (1977); W. D. Heiss, J. Math. Phys. **21**, 848 (1980).

[6]W. D. Heiss, J. Math. Phys. **22**, 1182 (1981).

[7]R. J. Yaes, Phys. Rev. D **2**, 2457 (1970).

[8]M. Z. I. Gering and W. D. Heiss, submitted for publication to J. Math. Phys.

[9]H. J. Lipkin, N. Meshkov, and A. J. Glick, Nucl. Phys. **62**, 188 (1965).

[10]A. L. Fetter and J. D. Walecka,, *Quantum Theory of Many Particle Systems* (McGraw-Hill, New York, 1971).

An integral transformation in Lobachevsky space

Yi. A. Verdiyev
Institute of Physics of the Academy of Sciences of the AzSSR, Baku, USSR

(Received 24 February 1983; accepted for publication 22 April 1983)

The integral formulas of the one-to-one correspondence between the one-particle state $F_{s\lambda}(p)$ with nonzero mass $p^2 = E^2 - \bar{p}^2$, helicity $(-s \leqslant \lambda \leqslant s)$, and the state $F_\nu(k)$ $(\nu = \pm s, \pm (s-1),...,0$ or $\pm(\frac{1}{2})$ with mass equal to zero $(k^2 = \omega^2 - \bar{k}^2 = 0)$ is obtained.

PACS numbers: 02.30.Qy

It is the aim of this paper to obtain an integral transformation, which gives the relationship between one-particle states with nonzero mass and mass equal to zero. We shall base our calculations on the results of Refs. 1 and 2. In the latter, the realization of the representation $T_\chi(g)$, $\chi = (\sigma + 1 + \nu, \sigma + 1 - \nu)$, $g \in SO(3,1)$ in the space of the homogeneous function $f_\nu(k;\sigma)$ of degree σ on the light cone $[k,k] = \omega^2 - \bar{k}^2 = 0$, $\omega > 0$ is considered.

Expansion of the function $F_{s\lambda}(p)$ on the double sheeted hyperboloid $p^2 = [p,p] = E^2 - \bar{p}^2$, $E > 0$ in terms of irreducible unitary representations of the Lorentz group is given by the formula from Ref. 1.

$$F_{s\lambda}(p) = \frac{1}{8\pi^2} \int_{-\infty}^{\infty} \sum_{m/2=-s}^{+s} \sum_{JM} A_{JM}{}^{\chi} D_{s\lambda;JM}{}^{(\chi)}(h(p)) \times \tfrac{1}{4}(\rho^2 + m^2) d\rho, \tag{1}$$

where $D_{s\lambda;JM}{}^{(\chi)}(h(p))$ are the matrix elements of the principal series of the unitary representations $T_\chi(g)$, $\chi = ((i\rho + m)/2, (i\rho - m)/2)$, $g \in SO(3,1)$; $h(p)$ is the boost operator $-p = \mathring{p} h(p)$, $\mathring{p} = (1,0,0,0)$.

The inverse transformation has the form

$$A_{JM}{}^{\chi} = 8\pi \int \sum_{\lambda=-s}^{+s} F_{s\lambda}(p) D^{*}_{s\lambda;JM}{}^{(\chi)}(h(p)) dp, \tag{2}$$

where dp is the invariant measure over the double sheeted hyperboloid.

An expansion of the function $F_\nu(k)$ on the light cone $[k,k] = \omega^2 - \bar{k}^2 = 0$, $\omega > 0$ can be obtained from Eq. (1) by taking the limit $p^2 \to 0$. In this limit, the sum over the index $m/2$ in Eq. (1) is nonzero when $m/2 \equiv \nu = \lambda$.

One can obtain the same results by a method based on the idea that the space of the functions $F_\nu(k)$ has the subspaces of the homogeneous functions $f_\nu(k;\sigma)$ of degree σ: $f_\nu(tk;\sigma) = t^\sigma f_\nu(k;\sigma)$, $t > 0$. An expansion of the square-integrable function $F_\nu(k)$ in terms of the functions $f_\nu(k;\sigma)$ is given by the Mellin transformation[3]:

$$F_\nu(k) = \frac{1}{4\pi} \int_{-\infty}^{\infty} f_\nu(k;\rho) d\rho,$$

$$f_\nu(k;\rho) = \int_0^\infty F_\nu(tk) t^{-i\rho/2} dt, \tag{3}$$

where $f_\nu(k;\rho)$ is the homogeneous function of degree $\sigma = -1 + i\rho/2$. The homogeneous function $f_\nu(k;\sigma)$ is also determined by its values $f_\nu(\bar{n};\sigma)$ on the unit sphere $[\bar{n}\,\bar{n}] = n_1^2 + n_2^2 + n_3^2 = 1$. Let $D_{\nu M}{}^{(J)}(r(0,\bar{n}))$ be the orthonormal canonical basis in the space of the functions $f_\nu(\bar{n};\sigma)$:

$$f_\nu(\bar{n};\sigma) = \sum_{JM} A_{JM}{}^{\chi} D_{\nu M}{}^{(J)}(r(0,\bar{n})). \tag{4}$$

Action of the representation operator $T_\chi(g)$, $g \in SO(3,1)$ on this basis is given by

$$T_\chi(g) D_{\nu M}{}^{(J)}(r) = (\omega_g/\omega)^\sigma D_{\nu M}{}^{(J)}(r_g) = \sum_{J'M'} D_{\nu M'}{}^{(J')}(r) D_{J'M';JM}{}^{(\chi)}(g). \tag{5}$$

From Eq. (5) we have

$$D_{J'M';JM}{}^{\chi}(g) = \int \left(\frac{\omega_g}{\omega}\right)^\sigma D^{*}_{\nu M'}{}^{(J')}(r) D_{\nu M}{}^{J}(r_g) dr, \tag{6}$$

where dr is the invariant volume over the group rotation. Substituting (6) into (1), and summing over the indices J, M by taking into account Eq. (4), we get

$$F_{s\lambda}(p) = \frac{1}{4\pi^4} \int_{-\infty}^{+\infty} \sum_{\nu=-s}^{+s} \int f_\nu(\bar{n};\rho)(E - \bar{p}\bar{n})^{-1-i\rho/2} \times D^{*}_{\nu\lambda}{}^{(s)}(r_{h^{-1}(p)}) dr (\tfrac{1}{4}\rho^2 + \nu^2) d\rho. \tag{7}$$

Equation (7) may be rewritten in the form

$$F_{s\lambda}(p) = \frac{1}{8\pi^4} \int_{-\infty}^{+\infty} \sum_{\nu=-s}^{+s} f_\nu(k;\rho) \delta([pk] - 1) \times D^{*}_{\nu\lambda}{}^{(s)}(r_{h^{-1}(p)}) (\tfrac{1}{4}\rho^2 + \nu^2) d\rho\, dk,$$

where dk is the invariant measure on the upper cone $[k,k] = 0$, $\omega > 0$. Thus after the integration over variable ρ by Eq. (3), we derive the required transformation:

$$F_{s\lambda}(p) = \frac{2}{\pi^3} \int \sum_{\nu=-s}^{+s} F_\nu(k) [-\delta''([pk] - 1) + \nu^2 \delta([pk] - 1)] D^{*}_{\nu\lambda}{}^{(s)}(r_{h^{-1}(p)}) dk. \tag{8}$$

Analogously, we find that the inverse transformation is

$$F_\nu(k) = \int \sum_{\lambda=-s}^{+s} F_{s\lambda}(p) \delta([pk] - 1) D_{\nu\lambda}{}^{(s)}(r_{h^{-1}(p)}) dp. \tag{9}$$

The equation $[pk] = 1$ defines a two-dimensional surface–horosphere in the Lobachevsky space[3] $[p\,p] = p^2$. Therefore, Eq. (7) gives a one-to-one correspondence between the function $F_{s\lambda}(p)$ and its integral over the horosphere. Thus the problem of the integral geometry in Lobachevsky space for a nondegenerate representation of the Lorentz group is solved. In the maximally degenerate case, i.e., $s = 0$ from (8) and (9), one gets a Gel'fand–Grayev integral transformation.[3]

Generalization of Eqs. (8) and (9) for the SO$(n,1)$ group is straightforward.

[1]Yi. A. Verdiyev, Ann. Phys. (NY) **72**, 139 (1972).
[2]Yi. A. Verdiyev, Theor. Math. Phys. (USSR) **16**, 360 (1973).
[3]I. M. Gel'fand, M. I. Grayev, and N. Ya. Vilenkin, *Integral geometry and connection with its problems of the theory representations*, Generalized functions No. 5 (Moscow, 1962).

 0022-2488/84/040814-01$02.50

Planar coherent states

William Roger Fuller
Department of Mathematics, University of Portland, Portland, Oregon 97203-5798

(Received 1 February 1983; accepted for publication 4 November 1983)

Planar coherent states with domains in the complex plane are defined by means of ladder operators acting in a separable Hilbert space. Some properties of the states are derived and examples provided to indicate areas of possible applications.

PACS numbers: 02.30.Sa, 02.30.Dk, 02.50.Cw

1. INTRODUCTION

Field coherent states were first applied to the theory of the quantum radiation field by Glauber[1] in 1963. The concept of an over-complete function with a homogeneous manifold of the Heisenberg group for its domain was a fruitful one and during the next decade generalizations were made to the group SU(2) by Radcliffe,[2] and by Arrechi *et al.*,[3] and to arbitrary Lie groups by Perelomov.[4] During the last ten years most of the applications[5,6,7] of coherent states have been developed within the context of the theory of Lie groups.

On the other hand, coherent states may be formulated independently of the Lie theory. The purpose of the present discussion is to provide one such development. Planar coherent states with domains in the complex plane are defined by means of ladder operators acting in an arbitrary separable Hilbert space. Examples are provided to indicate areas of possible applications, and some properties of the states are derived. One advantage of this approach lies in its making problems in the theory of coherent states more accessible to the standard tools of analysis.

2. LADDER OPERATORS

Let L be a complex separable Hilbert space and fix an orthonormal basis $\Gamma = \{f_0, f_1, f_2, \ldots\}$ in L. For each sequence $c = \{c_{-1}, c_0, c_1, \ldots\}$ of complex numbers with $c_{-1} = 0$, we may define a raising operator $A_c^\dagger$ in L by setting $A_c^\dagger f_n = c_n f_{n+1}$, for $n = 0,1,2,\ldots$, and then extending $A_c^\dagger$ to the linear subspace L_Γ generated by Γ in the usual manner by linearity. Similarly, we may define a lowering operator A_c on L_Γ by the relations $A_c f_n = \bar{c}_{n-1} f_{n-1}$ for $n = 0,1,2,\ldots$. A_c annihilates f_0 and for this reason f_0 is called the vacuum associated with A_c and $A_c^\dagger$. Together A_c and $A_c^\dagger$ are referred to as ladder operators associated with the sequence c.

By construction $A_c^\dagger$ is the adjoint of A_c on L_Γ. The products $A_c A_c^\dagger$ and $A_c^\dagger A_c$ are, consequently, self-adjoint there and possess the elements of Γ as characteristic vectors: for $n = 0,1,2,\ldots$, $A_c A_c^\dagger f_n = |c_n|^2 f_n$ and $A_c^\dagger A_c f_n = |c_{n-1}| f_n$. Since the commutator of A_c and $A_c^\dagger$ is, in general, not a restriction of some multiple of the identity operator 1_L on L, the Lie algebra generated by the ladder operators associated with a given sequence c will not be isomorphic to that associated with the Heisenberg group.

Operators of the form $H = r + sA_c A_c^\dagger + tA_c^\dagger A_c$, with r, s, and t complex numbers, are of physical interest,[8] and will be considered later. Such operators have as eigenvalue equations

$$Hf_n = (r + s|c_n|^2 + t|c_{n-1}|^2)f_n \tag{1}$$

for $n = 0,1,2,\ldots$. Conversely, if H is a closed symmetric operator on L with point spectrum, then the set of orthonormal characteristic vectors belonging to H may be chosen to be the set Γ and the equations (1) solved (though not uniquely) for the parameters r, s, and t, and for the sequence c. H may then be written on L_Γ as a quadratic polynomial in the ladder operators associated with the sequence c.

3. PLANAR COHERENT STATES

Let μ be a positive σ-finite measure on the Borel sets of an open domain G in the complex plane, and let $F(z)$ be a continuous function from G into the unit sphere in L. We define the projection operator $\hat{F}(z)$ associated with $F(z)$ by $\hat{F}(z)f = (F(z),f)\,F(z)$, for all f in L. Since the inner product $(g,\hat{F}(z)f)$ in L is μ-measurable on G for all f and g in L, the Bochner integral $\int_G \hat{F}(z)\,d\mu(z)$ exists. If this integral is equal to the identity operator on L, then $F(z)$ is said to possess the over-completeness property and $F(z)$ is called a planar coherent state on G based on the measure μ.

One standard method of constructing planar coherent states is by exponentiation of ladder operators. For example, let G be the unit disc $D = \{z\in\mathbb{C}: |z| < 1\}$, and let c be a sequence with general term $c = \sqrt{(n+1)(n+2)}$. We consider

$$[\exp(zA_c^\dagger)]f_0 = \sum_{n=0}^{\infty}\left(\frac{z^n}{n!}\right)(A_c^\dagger)^n f_0 = \sum_{n=0}^{\infty} b_n z^n f_n. \tag{2}$$

The sequence b is given by

$$b_n = (n!)^{-1}\prod_{j=0}^{n-1} c_j \tag{3}$$

for $n = 0,1,2,\ldots$. The empty product is taken as unity. Here $b_n = (n+1)^{1/2}$ so that the series (2) has radius of convergence $R = 1$, and therefore, defines an analytic function on D. As a consequence, the function

$$F(z) = (1 - |z|^2)\sum_{n=0}^{\infty}(n+1)^{1/2} z^n f_n \tag{4}$$

is a planar coherent state with respect to the measure $d\mu(z) = 2r(1 - r^2)^{-2}\,dr\,d\varphi$. Here we take $z = r\exp(2\pi i\varphi)$, while both dr and $d\varphi$ are Lebesque measures on $[0,1)$. It is clear that planar coherent states of the form (2) need not exist for arbitrary sequences c, nor need there be a measure μ which can act as a base for $\exp(zA_c^\dagger)f_0$. The fundamental link between the sequence c, the planar coherent state $F(z)$, and the measure μ in this example is that the measure $2r\,dr$ is a probability measure on $[0,1)$ and that its moment sequence is intimately related to the sequence b in (3). This observation

 0022-2488/84/040815-05$02.50

serves as motivation for the following discussion. We restrict our considerations to two domains G: the unit disk D and then the entire plane.

Let ν be a probability measure on the Borel sets of the unit interval. Assume that ν possesses finite nonzero moments ν_n of all orders and that the cumulative distribution function associated with ν is strictly less than one on $[0,1)$. If we define a sequence c by

$$c_n = (n+1)(\nu_{2n}/\nu_{2n+2})^{1/2}, \tag{5}$$

then the series (2) does define an analytic function on D. Indeed, in this case the sequence b in (3) has general term $b_n = (\nu_{2n})^{-1/2}$ and consequently the radius of convergence R of (2) is given by

$$R = \lim_{n\to\infty} (\nu_{2n})^{1/2n} = \inf\left\{s\colon \int_0^s d\nu(r) = 1\right\} = 1.$$

For each z in D the norm of (2) depends only on $|z|^2$ and will be denoted by $N^{-2}(|z|^2)$.

It is now clear that the function

$$F(z) = N(|z|^2)\exp(zA_c^\dagger)f_0 \tag{6a}$$

is a planar coherent state on D based on the measure

$$d\mu(z) = N^{-2}(r^2)d\nu(r)d\varphi. \tag{6b}$$

Conversely, let a sequence c be given such that the sequence $\{\nu_0,\nu_2,\nu_4\dots\}$ with general term $\nu_{2n} = (n!)^2|c_0c_1\dots c_{n-1}|^{-2}$ has an extension to a completely monotone sequence $\{\nu_0,\nu_1,\nu_2,\dots\}$ with the property that $\lim_{n\to\infty}\inf(\nu_n)^{1/n} = 1$. The Hausdorff moment theorem now guarantees the existence of a unique measure ν on the unit interval such that the function (6a) is a planar coherent state on D based on the measure (6b).

When G is the entire plane, we may also consider states of the form (2). Let ν be a probability measure on the Borel sets of $[0,+\infty)$, then for the state (6a) to be over complete the sequence c must be given by (5). In this case, moreover, the radius of convergence of the series (2) is infinite because the norm of the function $h(r) = r$ in the Lebesgue space $L_{2n}(\nu)$ has for its limit, as n becomes infinite, the essential supremum of h on $[0,+\infty)$, and this is infinite. We thus have that (6a) is a planar coherent state based on the measure (6b). The converse proceeds as in the case of G being the unit disk, except that here we consider only sequences $\{\nu_0,\nu_2,\nu_4,\dots\}$ which have extensions to positive definite sequences $\{\nu_0,\nu_1,\nu_2,\dots\}$. The Stieltjes moment theorem may now be applied to yield the existence of a probability measure ν in $[0,+\infty)$ such that (6a) is a planar coherent state on the plane based on the measure (6b).

The first illustrations are planar coherent states on D. Let $d\nu(r)$ be the uniform measure on $[0,1)$. Its moments are $\nu_n = (n+1)^{-1}$ and the associated sequence c is specified by $c_n = 2^{1/2}(2n+1)^{-1/2}(n+1)^{3/2}$. The corresponding planar coherent state is

$$F(z) = (1-|z|^2)(1+|z|^2)^{-1/2}\sum_{n=0}^{\infty}(2n+1)^{1/2}z^nf_n \tag{7}$$

and is based on the measure $d\mu(z) = (1+r^2)(1-r^2)^{-2}dr\,d\varphi$. Similarly, use of the beta distributions on $[0,1)$ with parameters $\alpha>0$ and $\beta\geqslant 1$ yields the planar coherent state

$$F(z) = [{}_2F_1(1,\alpha+\beta;\beta;|z|^2)]^{-1/2} \times\sum_{n=0}^{\infty}\left[\frac{B(\beta,n+\beta+\alpha)}{B(\alpha+\beta,n+\beta)}\right]^{1/2}z^nf_n \tag{8}$$

based on the measure $d\mu(z) = 2B^{-1}(\alpha,\beta)\,{}_2F_1(1,\alpha+\beta;\beta;r^2)(1-r^2)^{\alpha-1}r^{2\beta-1}\,dr\,d\varphi$. Here $B(r,s)$ is the beta function depending on the parameters r and s, and ${}_2F_1(r,s;t;x)$ the Gauss hypergeometric function. The introductory example (4) is a special case of (8) with $\alpha=\beta=1$.

To obtain examples of states defined on the entire plane, we may utilize the gamma distributions on $[0,+\infty)$ with parameters $\alpha>0$ and $\beta\geqslant 1$. In this case the planar coherent state is

$$F(z) = N_*(|z|^2)\Gamma^{1/2}(\beta)\sum_{n=0}^{\infty}\alpha^{-n}\Gamma^{-1/2}(2n+\beta)z^nf_n \tag{9}$$

with a base measure $d\mu(z) = N_*^{-2}(r^2)\Gamma^{-1}(\beta)\alpha^\beta r^{\beta-1}e^{-\alpha r}\,dr\,d\varphi$. The normalization factor N_* may be expressed in terms of Kummer functions:

$$N_*^{-2}(r^2) = \tfrac{1}{2}[{}_1F_1(1,\beta,r/\alpha) + {}_1F_1(1,\beta,-r/\alpha)].$$

Related states are those given by

$$F(z) = {}_1F_1^{-1/2}\left(1,\beta;\frac{r^2}{\alpha}\right)\Gamma^{1/2}(\beta) \times\sum_{n=0}^{\infty}\alpha^{-n/2}\Gamma^{-1/2}(n+\beta)z^nf_n \tag{10}$$

and based on the measure $d\mu(z) = 2\,{}_1F_1(1,\beta;r^2/\alpha)\Gamma^{-1}(\beta)\alpha^\beta r^{2\beta-1}e^{-\alpha r^2}\,dr\,d\varphi$. Again we require $\alpha>0$ and $\beta\geqslant 1$. These states are historically important because that state with $\alpha=\beta=1$ was the state studied by Glauber in connection with his work on the quantum radiation field.

Next, let H be an operator on L. If H has a pure point spectrum with eigenvalues $p_n = p_0 + pn$ $(p\neq 0)$, then $H = p_0 + pA_c^\dagger A_c$, where the ladder operators are defined with respect to an orthonormal basis of eigenvalues of H and with respect to the sequence c whose nth term is $c_n = (n+1)^{1/2}$. The associated planar coherent state is the state (10) with $\alpha=\beta=1$. The typical physical application for this state is to systems of quantum oscillators. Here L is the space of complex square-integrable functions on the real line which vanish at infinity. H is the usual Schrödinger operator: $-(\hbar^2/2m)(\partial/\partial x)^2 + \frac{1}{2}m\omega^2x^2$. The energy spectrum is $E_n = (h\omega/2\pi)(n+\frac{1}{2})$, $n = 0,1,2,\dots$, so that the planar coherent state has explicit form:

$$F(z,x) = \left(\frac{2m\omega}{h}\right)^{1/4}\exp\left(-\tfrac{1}{2}|z|^2 - \frac{\pi m\omega}{h}x^2\right) \times\sum_{n=0}^{\infty}2^{-n/2}(n!)^{-1}z^nH_n\left[\left(\frac{2\pi m\omega}{h}\right)^{1/2}x\right],$$

where $H_n(u)$, $n = 0,1,2,\dots$, are the Hermite polynomials.

When H has eigenvalues $p_n = p_0 + pn + qn^2$, with both p and q positive, then a solution of Eq. (1) yields the sequence c with the general term $c_n = (n+1)^{1/2}(n+1+p/q)^{1/2}$; we consequently obtain the planar coherent state (6a) on D:

$$F(z) = (1 - |z|^2)^{1/2}\left(1 + \frac{p}{q}\right) \times \sum_{n=0}^{\infty} \left[\left(\frac{p}{q}\right) B\left(n + 1, \frac{p}{q}\right)\right]^{-1/2} z^n f_n \tag{11}$$

based on the measure $d\mu(z) = 2(p/q)r(1 - r^2)^{-2} dr\, d\varphi$. This is a special case of Eq. (8) for $\beta = 1$ and $\alpha = p/q$. One area of application of this state may be the case when L is the space of continuous real-valued functions on $[-1, +1]$ and when H is the Legendre operator $(x^2 - 1)(d^2/dx^2) + 2x(d/dx)$ with the Legendre polynomials as eigenfunctions. Another application may be to the quantum rigid rotator. Here L is the space of complex square integrable functions on $[0, +\infty)$ which vanish at infinity; H is the radial operator $(-\hbar^2/2m)(\partial^2/\partial r^2) + (h^2 n(n+1)/8\pi^2 m r^2)$; and the orthonormal system Γ is the collection of Bessel functions.

When the Hilbert space L is finite-dimensional, the expression (2) for $[\exp(zA_c^\dagger)] f_0$ is a polynomial in z; consequently, the discussion dealing with the state (6) needs only straightforward modifications. As an example, consider a quantum particle with integral or half-integral spin S and with spin operators S_1, S_2, and S_3. These operators form a Lie algebra with products $S_1S_2 - S_2S_1 = iS_3$ (cyclic). S has spectrum $n - S$, $n = 0,1,2,\ldots,2S$ and has for eigenfunctions the spherical harmonics $Y_S^{n-S}(\theta,\varphi)$, where θ and φ are the usual angle variables in spherical coordinates. The linear combination $S^\dagger = i2^{-1/2}(S_1 + iS_2)$ is a raising operator for S_3 with source sequence c given by $c_n = [S(S+1) - (n-S)(n-S+1)]^{1/2}$, for $n = 0,1,\ldots,2S$. The corresponding planar coherent state is defined on $\mathbb{C}$:

$$F(z;\theta,\varphi) = (1 + |z|^2)^{-S} \sum_{n=0}^{2S} \binom{2S}{n}^{1/2} z^n Y_S^{n-S}(\theta,\varphi) \tag{12}$$

and is based on the measure $d\mu(z) = 2(2S+1)r(1 + r^2)^{-2} dr\, d\varphi$. This is not, however, the only planar coherent state that may be associated with this quantum system. If we take the finite sequence c with terms $c = (n+1)^{1/2} n = 0,1,\ldots,2S-1$, then we obtain the state

$$F(z;\theta,\varphi) = \widehat{N}^{-1/2}(|z|^2)\exp(-\tfrac{1}{2}|z|^2) \times \sum_{n=0}^{2S} (n!)^{-1/2} z^n Y_S^{n-S}(\theta,\varphi) \tag{13}$$

based on the measure $d\mu(z) = 2r\widehat{N}(r^2) dr\, d\varphi$. The normalization factor is

$$\widehat{N}(r^2) = 1 - \frac{r^{2S+1}}{(2S+1)!}\, {}_1F_1(2S+1; 2S+2; r^2).$$

Finally we note that not all operators with point spectra possess a planar coherent state based on a measure $d\mu = N^{-2}(r^2) d\nu(r) d\varphi$, where ν is a probability measure concentrated on $[0,1)$. Indeed, the spectrum of the hydrogen atom is of the form $p_n = p_0(n+1)^{-2}$, $n = 0,1,2,\ldots$; and for all choices of r, s, and t in (1) we obtain that $\lim_{n\to\infty} (|c_n| n^{-1}) < 1$. If an appropriate measure ν were to exist in $[0,1)$, its moments would have to obey the condition (5) and this would imply that $\lim_{n\to\infty} [\nu_{2n+2}(\nu_{2n})^{-1}] > 1$, so that there could be no extension to a completely monotone sequence of moments.

4. PROPERTIES OF PLANAR COHERENT STATES

For each planar coherent state $F(z)$ defined on a domain G and based on the measure μ, an overlap function $K(w,z)$ may be defined on $G \times G$ by $K(w,z) = (F(w), F(z))$. When $F(z)$ is given by (6a), then the overlap function is $K(w,z) = N^{-2}(\bar{w}z)N(|w|^2)N(|z|^2)$. $K(w,z)$ possesses the property of reproducing under convolutions:

$$\int_G K(w,z)K(z,u)\,d\mu(z) = K(w,u).$$

We note also that $K(0,z) = N(|z|^2)$ is never zero for any value of z so that f_0 is not orthogonal to $F(z)$ for any z. A more general question is that of "exact" orthogonality for planar coherent states: Given a nonzero element w of G, find z in G such that $K(w,z) = 0$. It seems to be the case that on each domain G there are planar coherent states for which solutions may exist to this problem and also states for which there are no solutions. For example, no solutions exist on the unit disk for the state (7). On the other hand, if L is two-dimensional with basis $\{f_0, f_1\}$, then $F(z) = (1 + 2|z|^2)\{^1(f_0 + \sqrt{2}zf_1)$ is a planar coherent state on D based on the measure $d\mu(z) = (1 + 2r^2)^2\, dr\, d\varphi$ and for each $w = s\exp(2\pi i\varphi)$ such that $\frac{1}{2} < s < 1$, $z = -(2s)^{-1}\exp(2\pi i\varphi)$ makes $K(w,z) = 0$. Similarly on $\mathbb{C}$, the overlap $K(s\exp(2\pi i\varphi), z)$ for the state (9) with $\alpha = \beta = 1$ vanishes whenever $z = is^{-1}(k + \frac{1}{2})\exp(2\pi i\varphi)$, $k = 0, \pm 1, \pm 2, \ldots$. In contrast, the problem has no solutions when the planar coherent state is given by (10) with $\alpha = \beta = 1$. When the parameters take arbitrary values in the special cases (8), (9), and (10), solutions are not easy to obtain and the question in general is unresolved.

Instead of exact "orthogonality," approximate orthogonality is sometimes considered.[1] For each nonzero w in G a subset V_w of the boundary of G is sought such that as z in G approaches V_w, the overlap $K(w,z)$ tends to zero. The unit circle for states (4), (7), and (8) is V_w; it is the ideal point at infinity for states (9) and (10); and V_w is empty in the example (12).

Next consider the action of the lowering operator A_c on the planar coherent state (6);

$$A_cF(z) = N(|z|^2) \sum_{n=0}^{\infty} |c_n|^2 [(n+1)!]^{-1} z^{n+1} (A_c^\dagger)^n f_0. \tag{14}$$

It now follows that the Glauber state (10) with $\alpha = \beta = 1$, is an eigenfunction of the associated lowering operator with eigenvalue z. Similar results do not follow in general. Indeed, the action of the corresponding lowering operator on the state (11) is given by

$$A_cF(z) = (zA_c^\dagger + 1 + p/q)zF(z).$$

On the other hand, given a sequence c and an orthonormal basis Γ which determine a planar coherent state $F(z)$ by means of (6), an operator B_c exists such that $B_cF(z) = zF(z)$. In fact, each sequence $\{t_0, t_1, t_2, \ldots\}$ of real numbers gives rise to a shift operator defined by

$$B_c f_n = \begin{cases} [(\nu_{2n+2}/\nu_{2n})^{1/2} \exp(-2\pi i t_n)] f_{n-1}, & n = 1,2,3,\ldots, \\ 0, & n = 0. \end{cases} \tag{15}$$

Clearly B_c acts on $F(z)$ in the desired manner. The two operators A_c and B_c are equal if and only if the measure ν has even moments of the form $\nu_{2n} = n!$, and the Glauber state is the most important example whose base measure has this property.

The over-completeness property of the planar coherent states (6) may be used to obtain "kernel" functions associated with elements in and operators on L. If g is an element in L, the function $g(z) = \Sigma_{n=0}^{\infty} b_n (g, f_n) z^n$, where b_n is given in Eq. (3), may be utilized to write g as a Bochner integral:

$$G = \int_G g(\bar{z}) F(z) d\mu(z).$$

In this integral, and in the following discussion, the domain G is either the unit disk or the plane. To each continuous operator H on L, we may uniquely associate the function $T(H;z) = (F(z), HF(z))$; we may also consider this function for some expressions involving the ladder operators:

$$T(A_c^{\dagger};z) = \bar{z} N^2(|z|^2) \sum_{n=0}^{\infty} (n+1)\ (\nu_{2n+2})^{-1} |z|^{2n},$$

$$T(A_c A_c^{\dagger};z) = N^2(|z|^2) \sum_{n=0}^{\infty} (n+1)^2 (\nu_{2n+2})^{-1} |z|^{2n}, \tag{16}$$

$$T(A_c^{\dagger} A_c;z) = N^2(|z|^2) \sum_{n=0}^{\infty} (n+1)^2\ \nu_{2n} (\nu_{2n+2})^{-2} |z|^{2n+2}.$$

These expressions are particularly simple for the Glauber state: $T(A_c^{\dagger};z) = \bar{z}$; $T(A_c A_c^{\dagger};z) = 1 + |z|^2$; $T(A_c^{\dagger} A_c;z) = |z|^2$. For fixed z, $T(H;z)$ is linear in H; it is, moreover, a continuous linear functional on the Liouville space of L. Indeed, if H_1 and H_2 are any two Hilbert–Schmidt operators on L, $|T(H_1;z) - T(H_2;z)|$ is dominated by the Hilbert–Schmidt norm of $(H_1 - H_2)$. We note also that for each trace-class operator H on L,

$$\operatorname{Tr} H = \int_G T(H;z) d\mu(z).$$

On the other hand, let $S(H;z)$ be a continuous linear functional on the Liouville space of L such that for each trace-class operator H on L, $\operatorname{Tr} H = \int_G S(H;z) d\mu(z)$, where $d\mu(z) = (\Sigma_{n=0}^{\infty} \nu_{2n}^{-1} r_{2n}) d\nu(r) d\varphi$. Riesz's theorem and the linearity of the inner product on the Liouville space indicate that there is a family $\{S(z): z \in G\}$ of Hilbert–Schmidt operators on L such that $S(H;z) = \operatorname{Tr}[S^{\dagger}(z)H]$, for all z in G and all Hilbert–Schmidt operators H, and such that $\int_G S(z)\, d\mu(z) = 1_L$. When the planar coherent state $F(z)$ is given by formula (6), and when $S(H;z) = T(H;z)$, then $S(z) = F(z)$.

We may associate an uncertainty relation to each pair of ladder operators, A_c and $A_c^{\dagger}$, for which a corresponding state of the form (6a) exists. Namely, we set $P = A_c^{\dagger} + A_c$, $Q = \iota(A_c^{\dagger} - A_c)$, and $Z = [A_c, A_c^{\dagger}]$. P, Q, and Z are Hermitian operators and $[P,Q] = 2\iota Z$. If we define $\langle P^2 \rangle = T(P^2;z) - T^2(P;z)$ and similarly for Q and Z then we obtain

$$\langle P^2 \rangle \langle Q^2 \rangle \geqslant \langle Z^2 \rangle. \tag{17}$$

Equality in (17) holds if and only if $T(PQ + QP;z) = 2T(P;z)T(Q;z)$, which in turn holds if and only if $\operatorname{Im} T(A_c^{\dagger 2};z) = \operatorname{Im} T^2(A_c^{\dagger};z)$. This last relation is equivalent to a denumerable family of restrictions on the even moments of the radial portion of the measure (6b): for $n = 0, 1, 2, \ldots,$

$$\sum_{k=0}^{n} \left[\frac{(k+1)(n-k+1)}{(\nu_{2k+2})(\nu_{2n-2k+2})} \right] = \sum_{k=0}^{n} \left[\frac{(k+1)(k+2)}{(\nu_{2k+4})(\nu_{2n-2k})} \right]. \tag{18}$$

Solution of (18) yields that $\nu_{2n} = (n!)(\nu_2)^n$. The states (10) with $\beta = 1$ are, therefore, examples of minimum uncertainty states in the sense of (17). On the other hand, every planar coherent state (6) yields equality in the relation

$$\langle \hat{P}^2 \rangle \langle \hat{Q}^2 \rangle \geqslant \langle \hat{Z}^2 \rangle, \tag{19}$$

where $\hat{P} = B_c^{\dagger} + B_c$; $\hat{Q} = \iota(B_c^{\dagger} - B_c)$, and $\hat{Z} = [B_c, B_c^{\dagger}]$. Here B_c is a lowering operator of the form (15) and $B_c^{\dagger}$ is the corresponding raising operator. Indeed, equality holds in (19) if and only if $\operatorname{Im} T(B_c^2;z) + \operatorname{Im} T^2(B_c;z)$ and this follows immediately from the definition of B_c.

Consider next the space of continuous complex-valued functions on D along with the inner product $\langle f,g \rangle = \int_D \overline{f(z)}\, g(z)\ d\mu(z)$, where $d\mu(z) = (\Sigma_{n=0}^{\infty} \nu_{2n}^{-1} r^{2n})\, d\nu(r) d\varphi$. With respect to this inner product, the set of functions

$\Psi = \{\psi_n(z) = \nu_{2n}^{-1/2} (\Sigma_{n=0}^{\infty} \nu_{2n}^{-1} r^{2n})^{-1/2} z^n : n = 0,1,2,\ldots\}$ forms an orthonormal set. Since for each z in D, $\Sigma_{n=0}^{\infty} |\psi_n(z)|^2 = 1$, each function of the form $h(z) = \Sigma_{n=0}^{\infty} s_n\ \psi_n(z)$ with $s = (s_0, s_1, s_2, \ldots)$ in $l_2(\mathbb{C})$, converges uniformly and absolutely on each compact subset of D. The set M of all such functions forms a Hilbert space with the inner product $\langle f,g \rangle$ and possesses ψ as an orthonormal basis; moreover, L and M are isometrically isomorphic under the correspondence

$$f = \sum_{n=0}^{\infty} s_n f_n \rightarrow \sum_{n=0}^{\infty} s_n\ \psi_n(z). \tag{20}$$

Here $\Gamma = \{f_0 f_1 f_2, \ldots\}$ is the orthonormal basis of L fixed earlier. We shall denote the image of g under (20) by $f(z)$.

The correspondence (20) permits another characterization of a planar coherent state on the unit disk. For each fixed z in D, the assignment $f \rightarrow f(z)$ defines a continuous linear functional on L. By Riesz's theorem, there is a vector $F^*(z)$ in L such that $f(z) = (F^*(z), f)$. For the given basis Γ, we have that $f_n(z) = \psi_n(z)$, for all z in D, and therefore, $F^*(z)$ has the form given in (6) for all points in the disk.

Let μ be an arbitrary positive, σ-finite measure on D and let $F(z)$ be a continuous function on D into L. Choose an orthonormal basis $\Gamma^* = \{f_0^* f_1^* f_2^*, \ldots\}$ in L. With respect to Γ^*, $F(z)$ is a planar coherent state based on the measure μ if and only if there exists an orthonormal basis $\Lambda = \{\lambda_0(z), \lambda_1(z), \lambda_2(z), \ldots\}$ for M such that for all z in D

$$1)\ \sum_{n=0}^{\infty} |\lambda_n(z)|^2 = 1; \tag{21}$$

and

$$2)\ F(z) = \sum_{n=0}^{\infty} \lambda_n(z) f_n^*. \tag{22}$$

Necessity follows by direct verification. On the other hand, if $F(z)$ is a planar coherent state, the set $\{\lambda_n(z) = (F(z), f_n^*) : n = 0,1,2,\ldots\}$ forms an orthonormal basis for M and the two properties, (21) and (22), follow immediately.

If ν is an arbitrary probability measure on $[0,1)$, if the sequence c is given by (5), and if an orthonormal basis $\Gamma = \{f_0 f_1 f_2, \ldots\}$ for L is given, the function (6a) is a planar coherent state based on the measure (6b). On the other hand, if $F(z)$ is an arbitrary planar coherent state based on a measure $d\mu(z) = (\Sigma_{n=0}^{\infty} \nu_{2n}^{-1} r^{2n}) d\nu(r) d\varphi$, with ν a probability measure, then there exists an orthonormal basis Γ in L and a sequence c such that $F(z) = (\exp zA_c^\dagger) f_0$. Indeed, let $\Gamma^* = \{f_0^*, f_1^*, f_2^*, \ldots\}$ be an arbitrary orthonormal basis for L. There is a corresponding orthonormal basis $\Lambda = \{\lambda_0(z), \lambda_1(z), \ldots\}$ for M such that for all z in D Eq. (22) holds. Let Φ be the isometric isomorphism from L onto M defined by its values on this basis: $\Phi(f_n^*) = \lambda_n(z)$, $n = 0,1,2,\ldots$, and let U be the unitary transformation that effects a change of basis from Λ to Ψ. The set $\{f_n = \Phi^{-1}(U(\Phi(f_n^*))) : n = 0,1,2,\ldots\}$ is an orthonormal basis for L and relative to this basis $F(z) = \Sigma_{n=0}^{\infty} \psi_n(z) f_n$. With the sequence c specified by (5), $F(z)$ has the desired form. Thus the planar coherent states which are based on measures of the form (6a) are precisely those which may be put into the form (6a).

Finally, we observe that the continuity of both $f(z)$ and $F(z)$ implies that $\overline{f(z)}F(z)$ is Bochner integrable. The overcompleteness property of $F(z)$ indicates that

$$f = \int_D \overline{f(z)} F(z) \, d\mu(z). \tag{23}$$

This relation is just the inverse of $f(z) = (F(z), f)$, which was obtained earlier. An immediate consequence of (23) is that the overlap function $K(z,w)$ is the kernel of the identity operator on M. For each w in D,

$$f(w) = \int_D f(z) K(z,w) \, d\mu(z).$$

5. AN APPLICATION TO SPIN SYSTEMS

In this section we shall use planar coherent states to obtain a lower bound for the partition function of a system of generalized quantum "spins" of dimension $S+1$. Since the generalization is based on probability measures, we shall call these spins probability spins, or p-spins. The state for a single p-spin is $\mathbb{C}^{S+1}$, and we take $\{K_n : n = 0,1,2,\ldots,S\}$ to be the usual Kronecker basis for $\mathbb{C}^{S+1}$. As in the previous sections of the paper, let $d\nu(r)d\varphi$ be a measure either on the unit disk D or on the entire plane $\mathbb{C}$ such that ν is a probability measure on $[0,1)$ or $[0,\infty)$, respectively, and let the sequence $C = \{c_n\}$ be defined by (5) for $n = 0,1,2,\ldots,S-1$. We set both C_{-1} and C_S to zero, and we also take A_c and $A_c^\dagger$ to be the general spin operators. The planar coherent state (6) is now formed based on the measure $d\mu(z) = N_s^{-2}(r^2) d\nu(r) d\varphi$, where

$$N_S^{-2}(|z|^2) = \|\exp(zA_c^\dagger) K_0\|^2.$$

In view of the planar coherent states, (12) and (13), it is clear that this p-spin model is natural to consider since it is just an explicit development of what was implicit in the discussion of those states.

The system we shall consider is an anisotropic chain of p-spins, each of dimension $S^j + 1$, and each associated with the measure $d\mu^j(z) = d\nu^j(r) d\varphi$. Here and throughout this discussion, the superscript j labels quantities belonging to the jth p-spin. There are M p-spins in the chain and the Hamiltonian is taken to be

$$H = -\sum_{j=1}^{M-1} [A_c^j + (A_c^j)^\dagger] \, [A_c^{j+1} + (A_c^{j+1})^\dagger].$$

The Hamiltonian acts on the Hilbert space for the chain

$$L_M = \bigotimes_{j=1}^{M} \mathbb{C}^{S^j+1},$$

so that the partition function is

$$Z = \mathrm{Tr} \exp(-\beta H), \tag{24}$$

where the trace is evaluated with respect to bases for L_M.

The trace in (24) may be replaced by an integration over the measure $d\mu_* = d\mu^1(z^1) \ldots d\mu^M(z^M)$ by the use of the following identity. Let A be any operator on the finite-dimensional Hilbert space L, and let $F(z)$ be a planar coherent state in L based on some measure $d\mu(z)$. Then for any orthonormal basis $\{f_n\}$ in L:

$$\begin{aligned}
\mathrm{Tr} A &= \sum_n (f_n, Af_n) \\
&= \sum_n (A^\dagger f_n, \hat{F}(z) f_n) d\mu(z) \\
&= \int \sum_n (F(z), f_n)(f_n, AF(z)) d\mu(z) \\
&= \int T(A,z) d\mu(z).
\end{aligned}$$

In terms of the products $F_M = \bigotimes_{j=1}^{M} F^j(z^j)$, the partition function may be written

$$Z \int (F_M e^{-\rho H} F_M) d\mu_*.$$

The lower bound for z is now obtained by an application of the Peierls–Bogoliubov inequality:

$$Z \geqslant \int e^{-\rho(F_M, HF_M)} d\mu_*.$$

The inner product in the argument of the exponential may be recast by use of the first formula in (16):

$$\begin{aligned}
(F_M, HF_M) &= \sum_{j=1}^{M} T(A_c^j + (A_c^j)^\dagger; z^j) T(A_c^{j+1} + (A_c^{j+1})^\dagger; z^{j+1}) \\
&= 2 \sum_{j=1}^{m} [g^j(r^j) x^j] \, [g^{j+1}(r^{j+1}) x^{j+1}],
\end{aligned}$$

where

$$g^j(t) = \left(\sum_{k=0}^{S^j-1} (k+1)(\nu_{2k+2}^j)^{-1} t^{2k} \right) \left(\sum_{k=0}^{S^j} (\nu_{2k}^j)^{-1} t^{2k} \right)^{-1}.$$

Each $g^j(r)$ is dominated by $2^{-1/2} S^j (S^j + 1)$ so that

$$Z \geqslant \int \exp -\beta \sum_{j=1}^{M} [S^j(S^j+1) x^j] \, [S^{j+1}(S^{j+1}+1) x^{j+1}] \, d\mu_*. \tag{25}$$

When all the measures $d\mu_M$ are identical, the integral on the right-hand side of (25) is just the classical partition function for a particular system of M identical planar rotors with interactions indicated in the exponent of the integrand.

[1] R. Glauber, Phys. Rev. **131**, 2766 (1963).
[2] J. M. Radcliffe, J. Phys. A **4**, 313 (1971).
[3] F. T. Arrechi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2211 (1972).
[4] A. M. Perelomov, Commun. Math. Phys. **26**, 222 (1972).
[5] E. H. Lieb, Commun. Math. Phys. **31**, 327 (1973).
[6] W. Fuller and A. Lenard, Commun. Math. Phys. **67**, 69 (1979).
[7] P. Gulshani and A. Volkov, J. Phys. A **13**, 3195 (1980).
[8] L. de la Pena and R. Montemayor, Am. J. Phys. **48**, 855 (1980).

General transformation matrix for Dirac spinors and the calculation of spinorial amplitudes

Kiwon Nam and Michael J. Moravcsik
Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon 97403

(Received 8 September 1982; accepted for publication 3 June 1983)

A general transformation matrix $T(\tilde{p},\tilde{s};p,s)$ is constructed which transforms a Dirac spinor $\psi(p,s)$ into another Dirac spinor $\psi(\tilde{p},\tilde{s})$ with arbitrarily given momenta and polarization states by exploiting the so-called Stech operator as one of generators for those transformations. This transformation matrix is then used in a calculation to yield the spinorial matrix element $M = \bar{\psi}(\tilde{p},\tilde{s})\Gamma\psi(p,s)$ for any spin polarization state. The final expressions of these matrix elements show the explicit structure of spin dependence for the process described by these spinorial amplitudes. The kinematical limiting cases such as very low energy or high energy of the various matrix elements can also be easily displayed. Our method is superior to the existing one in the following points. Since we have a well-defined transformation operator between two Dirac spinor states, we can evaluate the necessary phase factor of the matrix elements in an unambiguous way without introducing the coordinate system. This enables us to write down the Feynman amplitudes of complicated processes in any spin basis very easily in terms of previously calculated matrix elements of $\bar{\psi}\Gamma\psi$ which are building blocks of those Feynman amplitudes. In contrast, in the existing method, one has to figure out the suitable multiplication factor for each individual case. Furthermore, in the previous method one needs a coordinate system to evaluate phase factors, thus making the calculation more cumbersome. Also, the expression of the matrix element in the previous method is in the fractional form whose denominator may have a singularity structure (and not always an isolated one at that), while the spinorial amplitudes in our method take a much simpler form, free of coordinate systems in any spin basis. The usefulness of the results is illustrated on Compton scattering and on the elastic scattering of two identical massive leptons where the phase factor is important. It is also shown that the Stech operator as a polarization operator is simply related to the operator $K = \beta(\mathbf{\Sigma}\cdot\mathbf{L} + 1)/2$, which is often used in bound state problems.

PACS numbers: 02.30.Tb, 12.20.Ds, 11.10.Qr

I. INTRODUCTION

The explicit expressions of the spinorial matrix elements $\bar{\psi}(\tilde{p},\tilde{s})\Gamma\psi(p,s)$ in terms of the momenta and the physical, three-dimensional spin vectors[1] of the Dirac particles play an important role in particle reactions especially now that a multitude of polarization experiments are feasible and therefore the experimental determination of reaction amplitude could be achieved. Even though observables can be calculated by the traditional techniques using spin projection operators, the evaluation of the matrix elements do not lend themselves to this method of calculation. Also, it is known that measurements of all spin projections of particles in the scattering process along only one direction do not allow a complete determination of the amplitudes up to certain overall phase factors.[2] The complete determination of the scattering amplitude requires the measurements of spin projection along at least two nonparallel axes. It is therefore convenient to have the expression of matrix elements with respect to most general arbitrary polarization states. The aim of this paper is, therefore, to develop a simple and unambiguous calculational scheme for providing such matrix elements which form parts of a Feynman amplitude, and thereby of the S-matrix elements. In general, the unimodular phase factor of the Dirac spinor depends on the way one defines the spinor state. Therefore, the unimodular phase factor of a given Dirac spinor state is to a certain degree arbitrary so that we need a convention in order to fix the phase. The most commonly used one is Jacob–Wick helicity formalism,[3] and all other spin bases may be obtained from this by suitable transformations.

But if we can relate an arbitrary given Dirac spinor $\psi(p,s)$ to another given Dirac spinor $\psi(\tilde{p},\tilde{s})$ in a consistent and well-defined way, we can calculate the spinorial amplitude unambiguously for any spin basis using the spin projection operator. The general transformation operator $T(\tilde{p},\tilde{s};p,s)$ which transforms a given Dirac spinor $\psi(\tilde{p},\tilde{s})$ into any other spinor $\psi(\tilde{p},\tilde{s})$ is obtained by combining several simpler operators. This transformation operator is then used to calculate the matrix elements in a straightforward manner. The construction of the operator is presented in Sec. II, while the matrix elements are given in Sec. III, where the important phase factor is also briefly discussed. The more detailed discussion about the phase factor is in Appendix C. Since we can figure out the relevant phase factor, it is a matter of a simple algebra to write down any S-matrix element using tabulated spinorial amplitudes $M = \bar{\psi}(\tilde{p},\tilde{s})\Gamma\psi(p,s)$ for any spin states. There is another scheme[4] to calculate spinorial matrix elements M. Since $M = \bar{u}(\tilde{p},\tilde{s})\Gamma u(p,s)$ does not change if we multiply it by a somewhat simpler matrix element such as $S = \bar{u}(\tilde{p},\tilde{s})u(p,s)$ and divide it by the same factor to get $M = \bar{u}(\tilde{p},\tilde{s})\Gamma u(p,s)\bar{u}(p,s)u(\tilde{p},\tilde{s})/\bar{u}(\tilde{p},\tilde{s})u(p,s)$, we can use the spin projection operator to calculate the numerator. But one has to specify the coordinate system to obtain the denominator.[4]

Our method and final expressions are superior to those in Ref. 3 in the following ways.

(i) In principle, we can calculate $M = \bar{u}\Gamma u$, for any Γ from $|M^2|$ except for the unimodular phase factors. Since these phase factors are not determined this way, one has to use an explicit representation of Dirac spinors and Γ matrices in a specific coordinate system to obtain the phase factors.[4] But these phase factors may depend on the representation of Dirac spinors and Γ matrices and also on the coordinate system employed. What Fearing and Silbar actually did was to obtain the phase of any $M = \bar{u}\Gamma u$ relative to the simpler one such as $\Gamma = 1$ and to find $M = \bar{u}u$ in the helicity basis by employing an explicit coordinate system. In our method all the final results are representation-independent and free of a coordinate system.

(ii) For any spin basis other than the helicity basis there does not seem to exist in the Fearing–Sibar method any unambiguous way to evaluate the phase factor even after introducing a coordinate system. One may have to express the spinorial amplitudes in a general spin basis with respect to the standard such as the Jacob–Wick helicity basis to have the correct phase factors. This could be a complicated procedure. In contrast our results can be applicable to any spin basis with the knowledge of the necessary phase.

(iii) Their final results are in a fractional form which could be rather inconvenient in algebraic manipulations. Furthermore, the denominator (which is the multiplication factor) may have singularities which are not necessarily isolated zeros[5] so that one has to be careful in choosing a suitable multiplication factor.

(iv) They have to figure out the suitable multiplication factor for every individual situation and also its relevant phase factor. But in our algorithm we can use the calculated results of $M = \bar{u}\Gamma u$ to write down any S-matrix element straightforwardly in any spin basis, and it is easy to evaluate the necessary phase factor in a well-defined way. As an illustration of the usefulness of our results, the scattering amplitudes for Compton scattering and for Möller scattering are evaluated to the lowest order in QED in Sec. IV. As a by-product of the discussion, it is shown, in Appendix B, that the Stech operator[6] is directly related to the so-called K operator[7] commonly used in bound state problems.

II. TRANSFORMATION OPERATORS

In this section we provide a general transformation operator $T(\tilde{p},\tilde{s};p,s)$ that transforms a Dirac spinor $\psi(p,s)$ into another Dirac spinor $\psi(\tilde{p},\tilde{s})$ with arbitrarily given momenta and polarization states, i.e., $\psi(\tilde{p},\tilde{s}) = T(\tilde{p},\tilde{s};p,s)\psi(p,s)$. The result that we have obtained is, in the form of 2×2 block matrix,

$$T(\tilde{p},\tilde{s};p,s) = \eta\left[\frac{(\tilde{E}+m)(E+m)}{4m^2}\right]^{1/2} \times \frac{1}{2\sqrt{2(1+\hat{s}\cdot\hat{\tilde{s}})}}\begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22}\end{pmatrix},$$

where

$$\begin{aligned} T_{11} = -T_{22} = {} & (1+\hat{s}\cdot\hat{\tilde{s}})\left[1+\frac{\mathbf{p}\cdot\tilde{\mathbf{p}}}{(\tilde{E}+m)(E+m)}\right] \\ & + \frac{\hat{s}\times\hat{\tilde{s}}\cdot\mathbf{p}\times\tilde{\mathbf{p}}}{(\tilde{E}+m)(E+m)} \\ & - i(1+\hat{s}\cdot\hat{\tilde{s}})\frac{\boldsymbol{\sigma}\cdot\mathbf{p}\times\tilde{\mathbf{p}}}{(\tilde{E}+m)(E+m)} \\ & - i\left[1-\frac{\mathbf{p}\cdot\tilde{\mathbf{p}}}{(\tilde{E}+m)(E+m)}\right]\boldsymbol{\sigma}\cdot\hat{s}\times\hat{\tilde{s}} \\ & - i\frac{(\tilde{\mathbf{p}}\cdot\hat{s}\times\hat{\tilde{s}}\boldsymbol{\sigma}\cdot\mathbf{p} + \mathbf{p}\cdot\hat{s}\times\hat{\tilde{s}}\boldsymbol{\sigma}\cdot\tilde{\mathbf{p}})}{(\tilde{E}+m)(E+m)} \end{aligned}$$

and

$$\begin{aligned} T_{12} = -T_{21} = {} & -(1+\hat{s}\cdot\hat{\tilde{s}})\left(\frac{\boldsymbol{\sigma}\cdot\tilde{\mathbf{p}}}{\tilde{E}+m}+\frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m}\right) \\ & + i\left(\frac{\mathbf{p}\cdot\hat{s}\times\hat{\tilde{s}}}{E+m}+\frac{\tilde{\mathbf{p}}\cdot\hat{s}\times\hat{\tilde{s}}}{\tilde{E}+m}\right) \\ & + \left[\frac{\boldsymbol{\sigma}\cdot\mathbf{p}\times(\hat{s}\times\hat{\tilde{s}})}{E+m}-\frac{\boldsymbol{\sigma}\cdot\tilde{\mathbf{p}}\times(\hat{s}\times\hat{\tilde{s}})}{\tilde{E}+m}\right]. \end{aligned} \tag{2.1}$$

Here η is the phase factor of modulus 1 which will be discussed in Appendix C. $\boldsymbol{\sigma}$'s are the 2×2 Pauli matrices in the vector form. We denote $\psi(p,s)$ as a simultaneous eigenspinor of two commuting operators[8] $\not{p} = p_\mu\gamma^\mu = E\gamma_0 - \mathbf{p}\cdot\boldsymbol{\gamma}$ and $\boldsymbol{\pi}\cdot\hat{s}$, where $\hat{s}$ is a three-dimensional unit vector describing the spin state. $\boldsymbol{\pi}$ is a vector operator and is given by

$$\begin{aligned} \boldsymbol{\pi} &= \boldsymbol{\pi}^{(+)}\left(\frac{\not{p}+m}{2m}\right) + [-\boldsymbol{\pi}^{(-)}]\left(\frac{-\not{p}+m}{2m}\right) \\ &= \begin{pmatrix} \boldsymbol{\sigma}+\dfrac{\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}}{m(E+m)} & -\dfrac{\mathbf{p}}{m} \\ \dfrac{\mathbf{p}}{m} & -\boldsymbol{\sigma}-\dfrac{\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}}{m(E+m)} \end{pmatrix}, \end{aligned} \tag{2.2}$$

with

$$\boldsymbol{\pi}^{(+)} = \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & -\boldsymbol{\sigma}+2\boldsymbol{\sigma}\cdot\hat{p}\hat{p} \end{pmatrix}$$

and

$$\boldsymbol{\pi}^{(-)} = \begin{pmatrix} -\boldsymbol{\sigma}+2\boldsymbol{\sigma}\cdot\hat{p}\hat{p} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix}. \tag{2.3}$$

The 4×4 matrix $\boldsymbol{\pi}^{(+)}$ is known as the Stech operator, which is used as a polarization operator. Its heuristic derivation is given in Appendix A and some of its properties are discussed in Appendix B. The unit vector $\hat{s}$ is the same as the one appearing in the covariant 4-vector $S^\mu = ((\mathbf{p}\cdot\mathbf{s})/m, \hat{s} + (\mathbf{p}\mathbf{p}\cdot\hat{s})/m(E+m))$. Specifically, we can see that the projection operator for a positive energy state is

$$\begin{aligned} u(p,s)\bar{u}(p,s) &= \left(\frac{\not{p}+m}{2m}\right)\left(\frac{1+\gamma_5\not{S}}{2}\right) \\ &= \left(\frac{\not{p}+m}{2m}\right)\left(\frac{1+\boldsymbol{\pi}^{(+)}\cdot\hat{s}}{2}\right). \end{aligned} \tag{2.4}$$

We will use $\psi(|\mathbf{p}|\hat{p},\hat{s})$ instead of $\psi(p,s)$ whenever it is convenient to do so in the following discussion. The construction of $T(\tilde{p},\tilde{s};p,s)$ can be done as follows. First of all, we need a certain transformation operator $R(\hat{\tilde{s}};\hat{s})$ which changes only the spin state of a Dirac spinor, that is,

$$\psi(|\mathbf{p}|\hat{p},\hat{\tilde{s}}) = R(\hat{\tilde{s}};\hat{s})\psi(|\mathbf{p}|\hat{p},\hat{s}).$$

This can be obtained by treating the operator $\boldsymbol{\pi}$ as a generator of $R\,(\underset{\sim}{\hat{s}};\hat{s})$;

$$R\,(\underset{\sim}{\hat{s}};\hat{s}) = e^{i(1/2)\boldsymbol{\pi}\cdot\boldsymbol{\theta}} = \cos(\theta/2) + i\boldsymbol{\pi}\cdot\hat{\theta}\sin(\theta/2),$$

where

$$\cos\theta = \hat{s}\cdot\underset{\sim}{\hat{s}}, \quad \hat{\theta} = (\hat{s}\times\underset{\sim}{\hat{s}})/|\hat{s}\times\underset{\sim}{\hat{s}}|. \tag{2.5}$$

After some algebra, we obtain

$$R\,(\hat{s};\underset{\sim}{\hat{s}}) \equiv R\,(\hat{p},\hat{s};\,\hat{p},\hat{s}) = \eta\,\frac{1}{\sqrt{2m^2(1+\hat{s}\cdot\underset{\sim}{\hat{s}})}} \times\begin{pmatrix} E\,(1+\hat{s}\cdot\underset{\sim}{\hat{s}}) - i(E-m)\hat{p}\cdot\hat{s}\times\underset{\sim}{\hat{s}}\boldsymbol{\sigma}\cdot\hat{p} - im\boldsymbol{\sigma}\cdot\hat{s}\times\underset{\sim}{\hat{s}} & i\mathbf{p}\cdot\hat{s}\times\underset{\sim}{\hat{s}} - (1+\hat{s}\cdot\underset{\sim}{\hat{s}})\boldsymbol{\sigma}\cdot\mathbf{p} \\ -i\mathbf{p}\cdot\hat{s}\times\underset{\sim}{\hat{s}} + (1+\hat{s}\cdot\underset{\sim}{\hat{s}})\boldsymbol{\sigma}\cdot\mathbf{p} & -E\,(1+\hat{s}\cdot\underset{\sim}{\hat{s}}) + i(E-m)\hat{p}\cdot\hat{s}\times\underset{\sim}{\hat{s}}\boldsymbol{\sigma}\cdot\hat{\mathbf{p}} + im\boldsymbol{\sigma}\cdot\hat{s}\times\underset{\sim}{\hat{s}} \end{pmatrix}, \tag{2.6}$$

where η is an appropriate phase factor. Note that for a particle with $|\mathbf{p}| = 0$, i.e., in the particle's rest frame, we have

$$R\,(\underset{\sim}{\hat{s}};\hat{s})_0 = \eta\,\frac{1}{\sqrt{2(1+\hat{s}\cdot\underset{\sim}{\hat{s}})}} \times\begin{pmatrix} 1+\hat{s}\cdot\underset{\sim}{\hat{s}} - i\boldsymbol{\sigma}\cdot\hat{s}\times\underset{\sim}{\hat{s}} & 0 \\ 0 & -(1+\hat{s}\cdot\underset{\sim}{\hat{s}}) + i\boldsymbol{\sigma}\cdot\hat{s}\times\underset{\sim}{\hat{s}} \end{pmatrix}. \tag{2.7}$$

Then we have for the most general transformation operator

$$T\,(\tilde{p},\tilde{s};\,p,s) = R\,(\tilde{p},\tilde{s};\,\tilde{p},s)L\,(\tilde{p})L^{-1}(p) \tag{2.8}$$

or

$$T\,(\tilde{p},\tilde{s};\,p,s) = L\,(\tilde{p})R\,(\tilde{s};s)_0L^{-1}(p), \tag{2.8}$$

where $L\,(p)$ is a Lorentz transformation operator for a Dirac spinor,

$$\psi(p,s) = L\,(p)\psi(s),$$

$$L\,(p) = \left(\frac{E+m}{2m}\right)^{1/2}\begin{pmatrix} 1 & (\boldsymbol{\sigma}\cdot\mathbf{p})/(E+m) \\ (\boldsymbol{\sigma}\cdot\mathbf{p})/(E+m) & 1 \end{pmatrix} \tag{2.9}$$

and $L^{-1}(p)$ is the inverse matrix of $L\,(p)$. The relation (2.9) will produce the result (2.1). This construction of the general transformation operator may seem trivial but it is not because of the following observation: One might have been able to use

$$\tilde{R}\,(\underset{\sim}{\hat{s}};\hat{s})_0 = \eta\,\frac{1}{\sqrt{2(1+\hat{s}\cdot\underset{\sim}{\hat{s}})}} \times\begin{pmatrix} 1+\hat{s}\cdot\underset{\sim}{\hat{s}} - i\boldsymbol{\sigma}\cdot\hat{s}\times\underset{\sim}{\hat{s}} & 0 \\ 0 & 1+\hat{s}\cdot\underset{\sim}{\hat{s}} - i\boldsymbol{\sigma}\cdot\hat{s}\times\underset{\sim}{\hat{s}} \end{pmatrix} \tag{2.10}$$

instead of (2.7) because both R_0 and $\tilde{R}_0$ give the same result for the positive energy Dirac spinor in its rest frame. But we have found that $T\,(|\mathbf{p}|\underset{\sim}{\hat{p}},\underset{\sim}{\hat{s}};|\mathbf{p}|\hat{p},\underset{\sim}{\hat{s}})\tilde{R}\,(\underset{\sim}{\hat{s}};\hat{s})$ is not equal to $\tilde{R}\,(\underset{\sim}{\hat{s}};\hat{s})T\,(|\mathbf{p}|\underset{\sim}{\hat{p}},\hat{s};|\mathbf{p}|\hat{p},\hat{s})$ while $T\,(|\mathbf{p}|\underset{\sim}{\hat{p}},\underset{\sim}{\hat{s}};|\mathbf{p}|\hat{p},\underset{\sim}{\hat{s}})R\,(\underset{\sim}{\hat{s}};\hat{s})$ is equal to $R\,(\underset{\sim}{\hat{s}};\hat{s})T\,(|\mathbf{p}|\underset{\sim}{\hat{p}},\underset{\sim}{\hat{s}};|\mathbf{p}|\hat{p},\underset{\sim}{\hat{s}})$. That is, we can define $T\,(|\mathbf{p}|\underset{\sim}{\hat{p}},\underset{\sim}{\hat{s}};|\mathbf{p}|\hat{p},\hat{s})$ uniquely only if we use $R\,(\underset{\sim}{\hat{s}};\hat{s})$. This is the main reason why we prefer $R\,(\underset{\sim}{\hat{s}};\hat{s})$ to $\tilde{R}\,(\underset{\sim}{\hat{s}};\hat{s})$.

III. CALCULATION OF SPINORIAL AMPLITUDES

Consider the matrix element $M = \bar{u}(\tilde{p},\tilde{s})\Gamma u(p,s)$. Here Γ is any 4×4 matrix that can be represented as a linear combination of 16 basis matrices $\{1,\gamma_5,\gamma_\mu,\gamma_5\gamma_\mu,\sigma_{\mu\nu}\}$. Equipped with the transformation operator in the last section, we can rewrite M as follows: Since

$$\bar{u}(\tilde{p},\tilde{s}) = \bar{u}(p,s)[\gamma_0T^\dagger(\tilde{p},\tilde{s};\,p,s)\gamma_0],$$

we have

$$\begin{aligned} M &= (\bar{u}(\tilde{p},\tilde{s}))_\alpha\Gamma_{\alpha\beta}(u(p,s))_\beta \\ &= (\bar{u}(p,s))_\xi\,[\gamma_0T^\dagger(\tilde{p},\tilde{s};\,p,s)\gamma_0]_{\xi\alpha}\Gamma_{\alpha\beta}(u(p,s))_\beta \\ &= \mathrm{Tr}\{\gamma_0T^\dagger(\tilde{p},\tilde{s};\,p,s)\gamma_0\Gamma\,[u(p,s)\bar{u}(p,s)]\}. \end{aligned} \tag{3.1}$$

Alternatively, if we use $u(p,s) = T\,(p,s\,\tilde{p},\tilde{s})u(\tilde{p},\tilde{s})$, we have

$$M = \mathrm{Tr}\{[u(\tilde{p},\tilde{s})\bar{u}(\tilde{p},\tilde{s})]\Gamma T\,(p,s;\,\tilde{p},\tilde{s})\}. \tag{3.2}$$

Though it is algebraically tedious, we can show that both expressions (3.1) and (3.2) produce the same results.

Since any Γ matrix can be expressed as a linear combination of $\{1,\gamma_5,\gamma_\mu,\gamma_5\gamma_\mu,\sigma_{\mu\nu}\}$, or equivalently $\{1,\gamma_5,\gamma_0,\boldsymbol{\gamma},\gamma_5\gamma_0,\gamma_5\boldsymbol{\gamma},\gamma_0\boldsymbol{\gamma},\gamma_5\gamma_0\boldsymbol{\gamma}\}$, we only need to calculate matrix elements with respect to this basis. We have obtained the following results:

$$c\bar{u}(\tilde{p},\tilde{s})u(p,s) = \left(\frac{1+\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}\left[1 - \frac{\mathbf{p}\cdot\underset{\sim}{\mathbf{p}}}{(\underset{\sim}{E}+m)(E+m)}\right] + i\,\frac{\mathbf{p}\times\underset{\sim}{\mathbf{p}}}{(\underset{\sim}{E}+m)(E+m)}\cdot(\hat{a}+i\mathbf{b}),$$

$$c\bar{u}(\tilde{p},\tilde{s})\gamma_5u(p,s) = \left(\frac{\mathbf{p}}{E+m} - \frac{\underset{\sim}{\mathbf{p}}}{\underset{\sim}{E}+m}\right)\cdot(\hat{a}+i\mathbf{b}),$$

$$
\begin{aligned}
c\bar{u}(\tilde{p},\tilde{s})\gamma_0 u(p,s) &= \left(\frac{1+\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}\left[1+\frac{\mathbf{p}\cdot\underset{\sim}{\mathbf{p}}}{(E+m)(\underset{\sim}{E}+m)}\right] - i\,\frac{\mathbf{p}\times\underset{\sim}{\mathbf{p}}}{(E+m)(\underset{\sim}{E}+m)}\cdot(\hat{a}+i\mathbf{b}),\\
c\bar{u}(\tilde{p},\tilde{s})\boldsymbol{\gamma} u(p,s) &= \left(\frac{1+\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}\left(\frac{\mathbf{p}}{E+m}+\frac{\underset{\sim}{\mathbf{p}}}{\underset{\sim}{E}+m}\right) + i\left(\frac{\mathbf{p}}{E+m}-\frac{\underset{\sim}{\mathbf{p}}}{\underset{\sim}{E}+m}\right)\times(\hat{a}+i\mathbf{b}),\\
c\bar{u}(\tilde{p},\tilde{s})\gamma_0\gamma_5 u(p,s) &= \left(\frac{\mathbf{p}}{E+m}+\frac{\underset{\sim}{\mathbf{p}}}{\underset{\sim}{E}+m}\right)\cdot(\hat{a}+i\mathbf{b}),\\
c\bar{u}(\tilde{p},\tilde{s})\boldsymbol{\gamma}\gamma_5 u(p,s) &= \left[1-\frac{\mathbf{p}\cdot\underset{\sim}{\mathbf{p}}}{(E+m)(\underset{\sim}{E}+m)}\right](\hat{a}+i\mathbf{b})\\
&\quad + i\left(\frac{1+\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}\frac{\mathbf{p}\times\underset{\sim}{\mathbf{p}}}{(E+m)(\underset{\sim}{E}+m)} + \frac{\mathbf{p}\underset{\sim}{\mathbf{p}}+\underset{\sim}{\mathbf{p}}\mathbf{p}}{(\underset{\sim}{E}+m)(E+m)}\cdot(\hat{a}+i\mathbf{b}),\\
c\bar{u}(\tilde{p},\tilde{s})\gamma_0\boldsymbol{\gamma} u(p,s) &= \left(\frac{1+\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}\left(\frac{\mathbf{p}}{E+m}-\frac{\underset{\sim}{\mathbf{p}}}{\underset{\sim}{E}+m}\right) + i\left(\frac{\mathbf{p}}{E+m}+\frac{\underset{\sim}{\mathbf{p}}}{\underset{\sim}{E}+m}\right)\times(\hat{a}+i\mathbf{b}),\\
c\bar{u}(\tilde{p},\tilde{s})\gamma_5\gamma_0\boldsymbol{\gamma} u(p,s) &= \left[1+\frac{\mathbf{p}\cdot\underset{\sim}{\mathbf{p}}}{(E+m)(\underset{\sim}{E}+m)}\right](\hat{a}+i\mathbf{b})\\
&\quad - i\left(\frac{1+\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}\frac{\mathbf{p}\times\underset{\sim}{\hat{\mathbf{p}}}}{(\underset{\sim}{E}+m)(E+m)} - \frac{\mathbf{p}\underset{\sim}{\mathbf{p}}+\underset{\sim}{\mathbf{p}}\mathbf{p}}{(\underset{\sim}{E}+m)(E+m)}\cdot(\hat{a}+i\mathbf{b}).
\end{aligned}
\tag{3.3}
$$

Here

$$
\begin{aligned}
c^{-1} &= \eta\left[\frac{(\underset{\sim}{E}+m)(E+m)}{4m^2}\right]^{1/2},\\
\hat{a} &\equiv \frac{\hat{s}+\underset{\sim}{\hat{s}}}{|\hat{s}+\underset{\sim}{\hat{s}}|},\quad \mathbf{b}\equiv\frac{\hat{s}\times\underset{\sim}{\hat{s}}}{|\hat{s}\times\underset{\sim}{\hat{s}}|}\left(\frac{1-\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}.
\end{aligned}
\tag{3.4}
$$

The above expressions show the explicit, characteristic spin structure for each spinorial amplitude which we can interpret as a vertex amplitude. Note that two vectors $\hat{a}$ and $\mathbf{b}$ are well defined except the case in the fixed quantization scheme. For the fixed quantization scheme we can obtain the suitable $\hat{a}$ and $\mathbf{b}$ using a limiting procedure: Let $\hat{s}=\alpha\hat{e}$ and $\underset{\sim}{\hat{s}}=\tilde{\alpha}\hat{e}$. Then we choose

$$
\hat{a}=\frac{\hat{s}+\underset{\sim}{\hat{s}}}{|\hat{s}+\underset{\sim}{\hat{s}}|}=\frac{1}{2}(\alpha+\tilde{\alpha})\hat{e}+\frac{1}{2}(\alpha\tilde{\alpha}-1)\hat{g}
$$

and

$$
\mathbf{b}=\frac{\hat{s}\times\underset{\sim}{\hat{s}}}{|\hat{s}\times\underset{\sim}{\hat{s}}|}\left(\frac{1-\hat{s}\cdot\underset{\sim}{\hat{s}}}{2}\right)^{1/2}=\frac{1}{2}(\alpha-\tilde{\alpha})\hat{f}. \tag{3.5}
$$

We can see that the set of three vectors $\{\hat{e},\hat{f},\hat{g}\}$ form basis vectors of a right-handed Cartesian coordinate system, i.e., $\hat{e}\times\hat{f}=\hat{g}$. All unit vectors $\hat{f}$ and $\hat{g}$ satisfying this property are acceptable, but a judicious choice of $\hat{f}$ and $\hat{g}$ will provide a simpler expression of matrix elements. The results (3.4) may be used in the very well defined manner for certain kinematical limiting cases, such as at the very low energy or in the extremely high energy limit, since all the amplitudes are expressed as functions of $|\mathbf{p}|/(E+m)$ and $|\underset{\sim}{\mathbf{p}}|/(\underset{\sim}{E}+m)$. The above matrix elements are a little complicated due to the fact that they contain all spin information, but they take on a simple form if we specify the spin basis explicitly. For example, for the case $|\underset{\sim}{\mathbf{p}}|=|\mathbf{p}|$ in the helicity basis where $\hat{S}=\lambda\hat{p}$ and $\underset{\sim}{\hat{S}}=\tilde{\lambda}\underset{\sim}{\hat{p}}$, we obtain the following expressions: $c^{-1}=\eta_{\lambda\tilde{\lambda}}(E+m)/2m$, where the phase factor $\eta_{\lambda\tilde{\lambda}}$ in this case is given by

$$
\begin{aligned}
\eta_{\lambda\tilde{\lambda}} &= \tfrac{1}{2}(\lambda-\tilde{\lambda})+\tfrac{1}{2}(1+\lambda\tilde{\lambda}),\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})u(\hat{p},\lambda) &= \left(\frac{1+\lambda\tilde{\lambda}\hat{p}\cdot\underset{\sim}{\hat{p}}}{2}\right)^{1/2}\left[1-\lambda\tilde{\lambda}\left(\frac{E-m}{E+m}\right)\right],
\end{aligned}
$$

$$
\begin{aligned}
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\gamma_5 u(\hat{p},\lambda) &= (\lambda-\tilde{\lambda})\frac{|\mathbf{p}|}{E+m}\left(\frac{1-\hat{\mathbf{p}}\cdot\underset{\sim}{\hat{\mathbf{p}}}}{2}\right)^{1/2},\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\gamma_0 u(\hat{p},\lambda) &= \left(\frac{1+\lambda\tilde{\lambda}\hat{p}\cdot\underset{\sim}{\hat{p}}}{2}\right)^{1/2}\left[1+\lambda\tilde{\lambda}\left(\frac{E-m}{E+m}\right)\right],\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\boldsymbol{\gamma} u(\hat{p},\lambda) &= \frac{|\mathbf{p}|}{E+m}\left(\frac{2}{1+\hat{p}\cdot\underset{\sim}{\hat{p}}}\right)^{1/2}\\
&\quad\times\left[\left(\frac{1+\lambda\tilde{\lambda}}{2}\right)(\hat{p}+\underset{\sim}{\hat{p}})+i\left(\frac{\lambda+\tilde{\lambda}}{2}\right)\hat{p}\times\underset{\sim}{\hat{p}}\right],\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\gamma_0\gamma_5 u(\hat{p},\lambda) &= (\lambda+\tilde{\lambda})\frac{|\mathbf{p}|}{E+m}(1+\hat{p}\cdot\underset{\sim}{\hat{p}}^2)^{1/2},\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\boldsymbol{\gamma}\gamma_5 u(\hat{p},\lambda) &= \left[1+\lambda\tilde{\lambda}\left(\frac{E-m}{E+m}\right)\right]\\
&\quad\times\frac{\lambda\hat{p}+\tilde{\lambda}\underset{\sim}{\hat{p}}+i\lambda\tilde{\lambda}\hat{p}\times\underset{\sim}{\hat{p}}}{\sqrt{2(1+\lambda\tilde{\lambda}\hat{p}\cdot\underset{\sim}{\hat{p}})}},\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\gamma_0\boldsymbol{\gamma} u(\hat{p},\lambda) &= \frac{|\mathbf{p}|}{E+m}\left(\frac{2}{1-\hat{p}\cdot\underset{\sim}{\hat{p}}}\right)^{1/2}\\
&\quad\times\left[\left(\frac{1-\lambda\tilde{\lambda}}{2}\right)(\hat{p}-\underset{\sim}{\hat{p}})\right.\\
&\quad\left.-i\left(\frac{\lambda-\tilde{\lambda}}{2}\right)\hat{p}\times\underset{\sim}{\hat{p}}\right],\\
c\bar{u}(\underset{\sim}{\hat{p}},\tilde{\lambda})\gamma_5\gamma_0\boldsymbol{\gamma} u(\hat{p},\lambda) &= \left[1-\lambda\tilde{\lambda}\left(\frac{E-m}{E+m}\right)\right]\\
&\quad\times\frac{\lambda\hat{p}+\tilde{\lambda}\underset{\sim}{\hat{p}}+i\lambda\tilde{\lambda}\hat{p}\times\underset{\sim}{\hat{p}}}{\sqrt{2(1+\lambda\tilde{\lambda}\hat{p}\cdot\underset{\sim}{\hat{p}})}}.
\end{aligned}
\tag{3.6}
$$

These expressions are exact results. Note that in this particular example with $|\underset{\sim}{\mathbf{p}}|=|\mathbf{p}|$ some matrix elements vanish identically when either $\tilde{\lambda}=\lambda$ or $\tilde{\lambda}=-\lambda$. Specifically, there cannot be any helicity flip part in $\bar{u}\boldsymbol{\gamma}u$ with these kinematics which occurs such in the elastic scattering. We can see explicitly that helicities are conserved in the high energy limit for the vector and the axial vector vertex cases while the helicity must change at the scalar, pseudoscalar, and tensor vertices of Dirac particles in the high energy asymptotic limit. One interesting observation could be made here. We expect that in the very high energy limit two helicity states become decoupled from each other as in the case of massless particles. Hence it is suggestive to know that only the vector and the axial vector vertex are compatible with this expectation. A scalar, or a pseudoscalar particle exchange between the Dirac particles in the high energy is unacceptable. We can easily find that for some cases such as $\Gamma = \gamma_5, \gamma_0\boldsymbol{\gamma}, \boldsymbol{\gamma}$, and $\gamma_0\gamma_5$ the matrix elements identically vanish for the helicity flip or the helicity conserving cases for any scattering angle. The same thing occurs in other quantization schemes, viz., $\bar{u}(\tilde{p},\tilde{s})u(p,s)$ vanishes identically in the transversity flip case. Hence the points where certain matrix elements such as $\bar{u}\gamma_5 u$ in the helicity scheme or $\bar{u}u$ in the transversity scheme vanish are not isolated zeros as is claimed in Ref. 3.

IV. EXAMPLES: COMPTON SCATTERING AND $e^-e^- \to e^-e^-$

A. Compton scattering

For the illustration of the application of the matrix elements obtained in the previous section, we consider Compton scattering of a lepton with momentum p_μ by a photon of momentum k_μ and a polarization ϵ_μ. We will specify the final state of the lepton by $\tilde{P}_\mu$ and a photon by $\tilde{\kappa}_\mu$, $\tilde{\epsilon}_\mu$. By adding the contribution of the two lowest-order Feynman diagrams, we obtain the following Feynman amplitude:

$$i\,\frac{2p\cdot\kappa p\cdot\tilde{\kappa}}{e^2}(F_{\rm I} + F_{\rm II}) = \bar{u}(\tilde{p},\tilde{s})[p\cdot\tilde{\kappa}(2p\cdot\epsilon\tilde{\not{\epsilon}}^* + \tilde{\not{\epsilon}}^*\not{\kappa}\not{\epsilon}) - p\cdot\kappa(2p\cdot\tilde{\epsilon}^*\not{\epsilon} - \not{\epsilon}\tilde{\not{\kappa}}\tilde{\not{\epsilon}}^*)]u(p,s). \tag{4.1}$$

Using the identity $\gamma_\mu\gamma_\nu\gamma_\lambda = g_{\mu\nu}\gamma_\lambda + g_{\nu\lambda}\gamma_\nu + i\epsilon_{\mu\nu\lambda\rho}\gamma_5\gamma^\rho$, we have

$$\begin{aligned} i\,\frac{2p\cdot\kappa p\cdot\tilde{\kappa}}{e^2}F &= \bar{u}(\tilde{p},\tilde{s})[(2p\cdot\epsilon p\cdot\tilde{\kappa} + p\cdot\kappa\tilde{\kappa}\cdot\epsilon)\tilde{\not{\epsilon}}^* + (p\cdot\tilde{\kappa}\kappa\cdot\tilde{\epsilon}^* - 2p\cdot\tilde{\epsilon}^*p\cdot\kappa)\not{\epsilon} \\ &\quad - (p\cdot\tilde{\kappa} + p\cdot\kappa)\epsilon\cdot\tilde{\epsilon}^*\not{\kappa} + i\epsilon_{\mu\nu\lambda\rho}\gamma_5\gamma^\rho(p\cdot\tilde{\kappa}\tilde{\epsilon}^{*\mu}\kappa^\nu\epsilon^\lambda \\ &\quad + p\cdot\kappa\epsilon^\mu\tilde{\kappa}^\nu\tilde{\epsilon}^{*\lambda})]u(p,s). \end{aligned} \tag{4.2}$$

In the laboratory frame where the lepton is at rest, $p = (m,\mathbf{0})$, $\tilde{\kappa} = (\underset{\sim}{\omega},\underset{\sim}{\mathbf{k}})$, $\epsilon_\mu = (0,\boldsymbol{\epsilon})$, and $\tilde{\epsilon} = (0,\underset{\sim}{\boldsymbol{\epsilon}})$. Also we have $\kappa\cdot\epsilon = 0$, $\tilde{\kappa}\cdot\tilde{\epsilon} = 0$ as well as $p\cdot\epsilon = 0$, $p\cdot\tilde{\epsilon} = 0$. Then we get

$$\begin{aligned} i\,\frac{2m^2\omega\underset{\sim}{\omega}}{e^2}F_L &= \bar{u}(\tilde{p},\tilde{s})[m\omega\boldsymbol{\epsilon}\cdot\underset{\sim}{\boldsymbol{\kappa}}\boldsymbol{\epsilon}^*\cdot\boldsymbol{\gamma} + m\underset{\sim}{\omega}\underset{\sim}{\boldsymbol{\epsilon}}^*\cdot\boldsymbol{\kappa}\boldsymbol{\epsilon}\cdot\boldsymbol{\gamma} \\ &\quad + m(\omega + \tilde{\omega})\boldsymbol{\epsilon}\cdot\underset{\sim}{\boldsymbol{\epsilon}}^*(\omega\gamma_0 - \mathbf{k}\cdot\boldsymbol{\gamma})]u(p,s). \end{aligned} \tag{4.3}$$

Note that terms proportional to $\epsilon_{\mu\nu\lambda\rho}$ cancel each other in the laboratory frame. In the center of mass frame this cancellation does not occur and we have a term proportional to $\boldsymbol{\epsilon}\times\tilde{\boldsymbol{\epsilon}}^*\cdot\gamma_5\boldsymbol{\gamma}$. Using the result in the previous section, we obtain

$$\begin{aligned} i\,\frac{2m^2\omega\underset{\sim}{\omega}}{e^2}\left(\frac{2m}{\underset{\sim}{E}+m}\right)^{1/2}F_L &= \Bigg[\omega(\omega + \underset{\sim}{\omega})\boldsymbol{\epsilon}\cdot\underset{\sim}{\boldsymbol{\epsilon}}^* + \left(\frac{\omega - \underset{\sim}{\omega}}{\underset{\sim}{E}+m}\right)\boldsymbol{\epsilon}\cdot\underset{\sim}{\boldsymbol{\kappa}}\underset{\sim}{\boldsymbol{\epsilon}}^*\cdot\boldsymbol{\kappa} \\ &\quad - \left(\frac{\omega + \underset{\sim}{\omega}}{\underset{\sim}{E}+m}\right)\boldsymbol{\epsilon}\cdot\boldsymbol{\epsilon}^*\boldsymbol{\kappa}\cdot\underset{\sim}{\mathbf{p}}\Bigg]\left(\frac{1 + \hat{s}\cdot\hat{\underset{\sim}{s}}}{2}\right)^{1/2} \\ &\quad - i\left[\frac{\hat{s} + \hat{\underset{\sim}{s}}}{|\hat{s} + \hat{\underset{\sim}{s}}|} + i\,\frac{\hat{s}\times\hat{\underset{\sim}{s}}}{|\hat{s}\times\hat{\underset{\sim}{s}}|}\left(\frac{1 - \hat{s}\cdot\hat{\underset{\sim}{s}}}{2}\right)^{1/2}\right] \\ &\quad \cdot\left[\frac{\omega\underset{\sim}{\boldsymbol{\epsilon}}^*\times\underset{\sim}{\mathbf{p}}\boldsymbol{\epsilon}\cdot\boldsymbol{\kappa} + \underset{\sim}{\omega}\boldsymbol{\epsilon}\times\underset{\sim}{\mathbf{p}}\underset{\sim}{\boldsymbol{\epsilon}}^*\cdot\boldsymbol{\kappa}}{\underset{\sim}{E}+m} + (\boldsymbol{\kappa}\times\mathbf{p})\,\frac{(\omega + \underset{\sim}{\omega})\boldsymbol{\epsilon}\cdot\underset{\sim}{\boldsymbol{\epsilon}}^*}{\underset{\sim}{E}+m}\right]. \end{aligned} \tag{4.4}$$

We can see that the structure of the Feynman amplitude becomes simpler (i.e., it has fewer terms) for photon states of linear polarization along the direction orthogonal to the scattering plane. We can neglect the overall phase factor in the above expressions since the observables in this case are independent of this phase.

B. $e^-e^- \to e^-e^-$

In this example, we have to evaluate the relevant phase factor due to the existence of the exchange diagram. We only consider the lowest-order Feynman diagrams in QED. Then the Feynman amplitude can be immediately written down in the c.m. frame using the results in Sec. III:

$$\begin{aligned} \tilde{F}_{\rm c.m.} &\equiv \frac{2i}{e^2}\left(\frac{2m|\mathbf{p}|}{E+m}\right)^2[1 - (\hat{p}\cdot\hat{\underset{\sim}{p}})^2]F_{\rm c.m.} \\ &= (1 + \hat{p}\cdot\hat{\underset{\sim}{p}})\bar{u}(\tilde{p},\tilde{\alpha})\gamma_\mu u(p,\alpha)\bar{u}(\tilde{q},\tilde{\beta})\gamma^\mu u(q,\beta) - (1 - \hat{p}\cdot\hat{\underset{\sim}{p}})\bar{u}(\tilde{q},\tilde{\beta})\gamma_\mu u(p,\alpha)\bar{u}(\tilde{p},\tilde{\alpha})\gamma^\mu u(q,\beta), \end{aligned} \tag{4.5}$$

where $p = (E,|\mathbf{p}|\hat{p})$, $q = (E, -|\mathbf{p}|\hat{p})$, $\tilde{p} = (E,|\mathbf{p}|\hat{\underset{\sim}{p}})$, and $\tilde{q} = (E, -|\mathbf{p}|\hat{\underset{\sim}{p}})$. For the most general polarization states, we obtain

$$
\begin{aligned}
\tilde{F}_{\text{c.m.}} = {} & (1+\hat{p}\cdot\underset{\sim}{\hat{p}})\Big\{\Big(\frac{1+\hat{\alpha}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}\Big(1+\frac{E-m}{E+m}\hat{p}\cdot\underset{\sim}{\hat{p}}\Big)-i\Big(\frac{E-m}{E+m}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\cdot\Big[\frac{\hat{\alpha}+\underset{\sim}{\hat{\alpha}}}{|\hat{\alpha}+\underset{\sim}{\hat{\alpha}}|}+i\frac{\hat{\alpha}\times\underset{\sim}{\hat{\alpha}}}{|\hat{\alpha}\times\underset{\sim}{\hat{\alpha}}|}\Big(\frac{1-\hat{\alpha}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}\Big]\Big\} \\
& \cdot\Big\{\Big(\frac{1+\hat{\beta}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}\Big(1+\frac{E-m}{E+m}\hat{p}\cdot\underset{\sim}{\hat{p}}\Big)-i\Big(\frac{E-m}{E+m}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\cdot\Big[\frac{\hat{\beta}+\underset{\sim}{\hat{\beta}}}{|\hat{\beta}+\underset{\sim}{\hat{\beta}}|}+i\frac{\hat{\beta}\times\underset{\sim}{\hat{\beta}}}{|\hat{\beta}\times\underset{\sim}{\hat{\beta}}|}\Big(\frac{1-\hat{\beta}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}\Big]\Big\} \\
& -(1+\hat{p}\cdot\underset{\sim}{\hat{p}})\Big(\frac{E-m}{E+m}\Big)\Big\{\Big(\frac{1+\hat{\alpha}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}(\hat{p}+\underset{\sim}{\hat{p}})+i(\hat{p}-\underset{\sim}{\hat{p}})\times\Big[\frac{\hat{\alpha}+\underset{\sim}{\hat{\alpha}}}{|\hat{\alpha}+\underset{\sim}{\hat{\alpha}}|}+i\frac{\hat{\alpha}\times\underset{\sim}{\hat{\alpha}}}{|\hat{\alpha}\times\underset{\sim}{\hat{\alpha}}|}\Big(\frac{1-\hat{\alpha}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}\Big]\Big\} \\
& \cdot\Big\{\Big(\frac{1+\hat{\beta}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}(-\hat{p}-\underset{\sim}{\hat{p}})+i(-\hat{p}-\underset{\sim}{\hat{p}})\times\Big[\frac{\hat{\beta}+\underset{\sim}{\hat{\beta}}}{|\hat{\beta}+\underset{\sim}{\hat{\beta}}|}+i\frac{\hat{\beta}\times\underset{\sim}{\hat{\beta}}}{|\hat{\beta}\times\underset{\sim}{\hat{\beta}}|}\Big(\frac{1-\hat{\beta}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}\Big]\Big\} \\
& -\eta_{\text{ex}}(1-\hat{p}\cdot\underset{\sim}{\hat{p}})\Big(\Big\{\Big(\frac{1+\hat{\beta}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}\Big(1-\frac{E-m}{E+m}\hat{p}\cdot\underset{\sim}{\hat{p}}\Big)+i\Big(\frac{E-m}{E+m}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\cdot\Big[\frac{\hat{\beta}+\underset{\sim}{\hat{\alpha}}}{|\hat{\beta}+\underset{\sim}{\hat{\alpha}}|}+i\frac{\hat{\beta}\times\underset{\sim}{\hat{\alpha}}}{|\hat{\beta}+\underset{\sim}{\hat{\alpha}}|}\Big(\frac{1-\hat{\beta}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}\Big]\Big\} \\
& \cdot\Big\{\Big(\frac{1+\hat{\alpha}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}\Big(1-\frac{E-m}{E+m}\hat{p}\cdot\underset{\sim}{\hat{p}}\Big)+i\Big(\frac{E-m}{E+m}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\cdot\Big[\frac{\hat{\alpha}+\underset{\sim}{\hat{\beta}}}{|\hat{\alpha}+\underset{\sim}{\hat{\beta}}|}+i\frac{\hat{\alpha}\times\underset{\sim}{\hat{\beta}}}{|\hat{\alpha}\times\underset{\sim}{\hat{\beta}}|}\Big(\frac{1-\hat{\alpha}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}\Big]\Big\} \\
& -\Big(\frac{E-m}{E+m}\Big)\Big\{\Big(\frac{1+\hat{\beta}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}(-\hat{p}+\underset{\sim}{\hat{p}})-i(\hat{p}+\underset{\sim}{\hat{p}})\times\Big[\frac{\hat{\beta}+\underset{\sim}{\hat{\alpha}}}{|\hat{\beta}+\underset{\sim}{\hat{\alpha}}|}+i\frac{\hat{\beta}\times\underset{\sim}{\hat{\alpha}}}{|\hat{\beta}\times\underset{\sim}{\hat{\alpha}}|}\Big(\frac{1-\hat{\beta}\cdot\underset{\sim}{\hat{\alpha}}}{2}\Big)^{1/2}\Big]\Big\} \\
& \cdot\Big\{\Big(\frac{1+\hat{\alpha}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}(\hat{p}-\underset{\sim}{\hat{p}})+i(\hat{p}+\underset{\sim}{\hat{p}})\times\Big[\frac{\hat{\alpha}+\underset{\sim}{\hat{\beta}}}{|\hat{\alpha}+\underset{\sim}{\hat{\beta}}|}+i\frac{\hat{\alpha}\times\underset{\sim}{\hat{\beta}}}{|\hat{\alpha}\times\underset{\sim}{\hat{\beta}}|}\Big(\frac{1-\hat{\alpha}\cdot\underset{\sim}{\hat{\beta}}}{2}\Big)^{1/2}\Big]\Big\}\Big).
\end{aligned}
\tag{4.6}
$$

The above expression is rather complicated due to the fact it contains all information on the spin structure. One can specify the spin basis to obtain a simpler expression.

In the helicity basis where $\hat{\alpha}=\alpha\hat{p}$, $\underset{\sim}{\hat{\alpha}}=\tilde{\alpha}\underset{\sim}{\hat{p}}$, $\hat{\beta}=\beta(-\hat{p})$, and $\underset{\sim}{\hat{\beta}}=\tilde{\beta}(-\underset{\sim}{\hat{p}})$, $\tilde{F}^{(H)}_{\text{c.m.}}$ is given in the following form:

$$
\begin{aligned}
\tilde{F}^{(H)}_{\text{c.m.}} \equiv {} & \frac{2i}{e^2}\Big(\frac{2m|\mathbf{p}|}{E+m}\Big)^2[1-(\hat{p}\cdot\underset{\sim}{\hat{p}})^2]F^{(H)}_{\text{c.m.}} = (1+\hat{p}\cdot\underset{\sim}{\hat{p}})\Big\{\Big(\frac{1+\alpha\tilde{\alpha}\hat{p}\cdot\underset{\sim}{\hat{p}}}{2}\Big)^{1/2}\Big[1+\alpha\tilde{\alpha}\Big(\frac{E-m}{E+m}\Big)\Big] \\
& \cdot\Big(\frac{1+\beta\tilde{\beta}\hat{p}\cdot\underset{\sim}{\hat{p}}}{2}\Big)^{1/2}\Big[1+\beta\tilde{\beta}\Big(\frac{E-m}{E+m}\Big)\Big]+\Big(\frac{E-m}{E+m}\Big)\Big(\frac{2}{1+\hat{p}\cdot\underset{\sim}{\hat{p}}}\Big)\Big[\Big(\frac{1+\alpha\tilde{\alpha}}{2}\Big)(\hat{p}+\underset{\sim}{\hat{p}})+i\Big(\frac{\alpha+\tilde{\alpha}}{2}\Big)(\hat{p}\times\underset{\sim}{\hat{p}})\Big] \\
& \cdot\Big[\Big(\frac{1+\beta\tilde{\beta}}{2}\Big)(\hat{p}+\underset{\sim}{\hat{p}})-i\Big(\frac{\beta+\tilde{\beta}}{2}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\Big]\Big\}-\eta_{\text{ex}}(1-\hat{p}\cdot\underset{\sim}{\hat{p}})\Big\{\Big(\frac{1-\alpha\tilde{\beta}\hat{p}\cdot\underset{\sim}{\hat{p}}}{2}\Big)^{1/2}\Big[1+\alpha\tilde{\beta}\Big(\frac{E-m}{E+m}\Big)\Big] \\
& \cdot\Big(\frac{1-\beta\tilde{\alpha}\hat{p}\cdot\underset{\sim}{\hat{p}}}{2}\Big)^{1/2}\Big[1+\beta\tilde{\alpha}\Big(\frac{E-m}{E+m}\Big)\Big]+\Big(\frac{E-m}{E+m}\Big)\Big(\frac{2}{1-\hat{p}\cdot\underset{\sim}{\hat{p}}}\Big)\Big[\Big(\frac{1+\alpha\tilde{\beta}}{2}\Big)(\hat{p}-\underset{\sim}{\hat{p}})-i\Big(\frac{\alpha+\tilde{\beta}}{2}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\Big] \\
& \cdot\Big[\Big(\frac{1+\beta\tilde{\alpha}}{2}\Big)(\hat{p}-\underset{\sim}{\hat{p}})+i\Big(\frac{\beta+\tilde{\alpha}}{2}\Big)\hat{p}\times\underset{\sim}{\hat{p}}\Big]\Big\}.
\end{aligned}
\tag{4.7}
$$

Here the relative phase factor η_{ex} is due to the fact that $u(\tilde{p},\tilde{\alpha})$ and $u(\tilde{q},\tilde{\beta})$ are obtained from $u(p,\alpha)$ and $u(q,\beta)$ along different paths in the exchange diagram than in the direct diagram. We obtain $\eta_{\text{ex}}=-1$ for the cases $(\alpha,\beta;\tilde{\alpha},\tilde{\beta})=(+,+;+,+)$, $(++;--)$, $(--;++)$, and $(-,-;-,-)$ and $\eta_{\text{ex}}=+1$ for all the other cases. For the specific helicity configuration we have obtained the following results:

$$
\begin{aligned}
& \frac{2i}{e}(2m|\mathbf{p}|)^2[1-(\hat{p}\cdot\underset{\sim}{\hat{p}})^2]F^{(H)}_{\text{c.m.}}(\alpha,\beta;\tilde{\alpha},\tilde{\beta})\equiv\tilde{F}^{(H)}_{\text{c.m.}}(\alpha,\beta;\tilde{\alpha},\tilde{\beta}), \\
& \tilde{F}^{(H)}_{\text{c.m.}}(+,+;+,+)=16|\mathbf{p}|^2+4m^2(1+\hat{p}\cdot\underset{\sim}{\hat{p}})^2, \\
& \tilde{F}^{(H)}_{\text{c.m.}}(+,+;-,-)=4m^2[1-(\hat{p}\cdot\underset{\sim}{\hat{p}})^2], \\
& \tilde{F}^{(H)}_{\text{c.m.}}(+,-;+,-)=4(1+\hat{p}\cdot\underset{\sim}{\hat{p}})(|\mathbf{p}|^2+E^2\hat{p}\cdot\underset{\sim}{\hat{p}}), \\
& \tilde{F}^{(H)}_{\text{c.m.}}(+,-;-,+)=-4(1-\hat{p}\cdot\underset{\sim}{\hat{p}})(|\mathbf{p}|^2-E^2\hat{p}\cdot\underset{\sim}{\hat{p}}), \\
& \tilde{F}^{(H)}_{\text{c.m.}}(+,-;+,+)=-4Em\hat{p}\cdot\underset{\sim}{\hat{p}}|\hat{p}\times\underset{\sim}{\hat{p}}|.
\end{aligned}
\tag{4.8}
$$

Note that $\tilde{F}^{(H)}_{\text{c.m.}}(++;--)$ is isotropic and $\tilde{F}^{(H)}_{\text{c.m.}}(+-;++)$ vanishes at 90° c.m. angle scattering as it should be in this identical particle scattering.

ACKNOWLEDGMENT

This work was supported in part by the U.S. Department of Energy under Contract DE-AT06-76ER-70004.

APPENDIX A: THE STECH OPERATOR, $\pi^{(+)}$

Consider the Foldy–Wouthuysen transformation of a free particle for the following operators: $\Sigma_A = \Sigma = \begin{pmatrix}\sigma & 0\\ 0 & \sigma\end{pmatrix}$, $\Sigma_B = \beta\Sigma = \begin{pmatrix}\sigma & 0\\ 0 & -\sigma\end{pmatrix}$. In the standard representation of Dirac γ matrices, we have the following form of the Foldy–Wouthuysen transformation:

$$U_{\rm FW} = \left(\frac{E+m}{2m}\right)^{1/2}\begin{pmatrix} 1 & \dfrac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m} \\ -\dfrac{\boldsymbol{\sigma}\cdot\mathbf{p}}{E+m} & 1\end{pmatrix}.$$

With this we have

$$\Sigma_A = U_{\rm FW}\Sigma_A U_{\rm FW}^{-1} = \begin{pmatrix} \dfrac{m}{E}\boldsymbol{\sigma} + \dfrac{\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}}{E(E+m)} & -i\dfrac{\boldsymbol{\sigma}\times\mathbf{p}}{m} \\ i\dfrac{\boldsymbol{\sigma}\times\mathbf{p}}{m} & \dfrac{m}{E}\boldsymbol{\sigma} + \dfrac{\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}}{E(E+m)}\end{pmatrix}$$

and

$$\Sigma_B = U_{\rm FW}\Sigma_B U_{\rm FW}^{-1} = \begin{pmatrix} \boldsymbol{\sigma} - \dfrac{\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}}{E(E+m)} & \dfrac{\mathbf{p}}{E} \\ \dfrac{\mathbf{p}}{E} & -\boldsymbol{\sigma} + \dfrac{\mathbf{p}\boldsymbol{\sigma}\cdot\mathbf{p}}{E(E+m)}\end{pmatrix}.$$

Now we obtain $\boldsymbol{\pi}^{(+)}$ as follows:

$$\boldsymbol{\pi}^{(+)} = (\Sigma_A\cdot\mathbf{p})\hat{p} + \hat{p}\times(\Sigma_B\times\hat{p}) = \begin{pmatrix}\boldsymbol{\sigma} & 0\\ 0 & -\boldsymbol{\sigma} + 2\boldsymbol{\sigma}\cdot\hat{p}\hat{p}\end{pmatrix}.$$

We construct $\boldsymbol{\pi}^{(+)}$ to have a form for the polarization operator along the momentum direction which is different from that for the polarization operator along the direction orthogonal to the momentum vector. We want to associate $\boldsymbol{\pi}^{(+)}$ with Σ_A along the direction of the momentum and with Σ_B along the direction orthogonal to the momentum.

APPENDIX B: SOME PROPERTIES OF $\pi^{(+)}$

The Stech operator $\boldsymbol{\pi}^{(+)}$ commutes with both $\not{P} = E\gamma_0 - \mathbf{p}\cdot\boldsymbol{\gamma}$ and the free Dirac particle Hamiltonian $H_0 = \boldsymbol{\alpha}\cdot p + \beta m$ while neither Σ nor $\beta\Sigma$ commutes with both of them. Since $\boldsymbol{\pi}^{(+)}$ commutes with $\not{P}$, we can construct the Dirac spinors which are simultaneous eigenstates of $\boldsymbol{\pi}^{(+)}\cdot\hat{s}$ and $\gamma^\mu P_\mu$. That is, we have a positive energy Dirac spinor $u(p,s)$ that satisfies both equations $Pu(p,s) = mu(p,s)$ and $(\boldsymbol{\pi}^{(+)}\cdot\hat{s})u(p,s) = u(p,s)$. The $u(p,s)$ is the one with momentum $P^\mu = (E,\mathbf{p})$ and covariant spin vector

$$S^\mu = \left(\frac{\mathbf{p}\cdot\hat{s}}{m}, \hat{s} + \frac{\mathbf{p}\hat{s}\cdot\mathbf{p}}{m(E+m)}\right).$$

Because $\boldsymbol{\pi}^{(+)}$ also commutes with the Lorentz transformation matrix (2.10), the unit vector $\hat{s}$ is Lorentz invariant. Thus one may say that the polarization is the same in any Lorentz frame if we define the polarization vector by the operator $\boldsymbol{\pi}^{(+)}\cdot\hat{s}$. There are two situations that show $\boldsymbol{\pi}^{(+)}$ being a useful polarization operator. The first one is the observation of the following relation: $(\boldsymbol{\pi}^{(+)}\cdot\hat{x})[U(|\mathbf{p}|\hat{z},\hat{z}) \pm U(|\mathbf{p}|\hat{z},-\hat{z})] = \pm[U(|\mathbf{p}|\hat{z},\hat{z}) \pm U(|\mathbf{p}|\hat{z},-\hat{z})]$. That is, we can interpret the state $[U(|\mathbf{p}|\hat{z},\hat{z}) \pm U(|\mathbf{p}|\hat{z},-\hat{z})]$ to be proportional to $[U(|\mathbf{p}|\hat{z},\pm\hat{x})]$.

This is similar to the nonrelativistic spin description where we can obtain the spin state polarized along the $\hat{x}$ direction by superposing two oppositely polarized spin states along the $\hat{z}$ direction.

The second situation is somewhat nontrivial because it comes from a system bound by a central potential such as the hydrogen atom. In the case of a Coulomb potential we have the Hamiltonian $H_C = \boldsymbol{\alpha}\cdot\mathbf{p} + \beta m + \beta V(r)$ which commutes with the total angular momentum operator $\mathbf{J} = \mathbf{L} + \mathbf{S} = \mathbf{r}\times\mathbf{p} + 1/2\Sigma$. We understand that for free Dirac particles the linear momentum $\mathbf{p}$ is a constant of the motion and its direction vector $\mathbf{p}$ can serve as a good reference direction for the polarization vector. In the bound state we have $\mathbf{J}$ as a conserved vector while $\mathbf{p}$ is not. Then we could use the direction of $\mathbf{J}$ as a reference direction for the description of the spin state of the bound Dirac particle. Let us consider an operator

$$\boldsymbol{\pi}^{(+)}\cdot\mathbf{J} = \begin{pmatrix}\frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{L} + \frac{3}{4} & 0\\ 0 & -\frac{1}{2}\boldsymbol{\sigma}\cdot\mathbf{L} - \frac{1}{4}\end{pmatrix} = \tfrac{1}{2}\beta(\Sigma\cdot\mathbf{L} + 1) + \tfrac{1}{4}.$$

We can recognize the first term in the above expression as the operator $K = \beta(\Sigma\cdot\mathbf{L} + 1)$, which is used to solve the problem of the hydrogen atom. In the usual discussion of the hydrogen atom, the introduction of the K operator is rather mathematical without having any clear physical interpretation. But we can see that K is related to the polarization operator for the bound system if we understand $\boldsymbol{\pi}^{(+)}$ as a polarization operator.

APPENDIX C: THE PHASE FACTOR η

The phase factor η which appears in the expression of $T(\tilde{p},\tilde{s};p,s)$ and in the spinorial amplitude is needed because the phase of the spinor $\psi(\tilde{p},\tilde{s})$ depends upon the path along which we construct $T(\tilde{p},\tilde{s};p,s)$. In most cases this phase factor can be neglected since we need only one kind of path to reach $\psi(\tilde{p},\tilde{s})$ from $\psi(p,s)$ and thus the observables are independent of η. But when we deal with the scattering of identical particles such as $e^-e^-\rightarrow e^-e^-$, we have to compare two different paths due to the presence of the exchange diagram. Let $M = \bar{\psi}(\tilde{p},\tilde{\alpha})\Gamma_1\psi(p,\alpha)\bar{\psi}(\tilde{q},\tilde{\beta})\Gamma_2\psi(q,\beta)$ and $\widetilde{M} = \bar{\psi}(\tilde{q},\tilde{\beta})\Gamma_1\psi(p,\alpha)\bar{\psi}(\tilde{p},\tilde{\alpha})\Gamma_2\psi(q,\beta)$. We can write down matrix elements M and $\widetilde{M}$ very easily using already evaluated basic spinorial amplitudes except the relevant phase factors. Since only the relative phase between M and $\widetilde{M}$ appears in interference phenomena, we just need to calculate this relative phase factor. This relative phase arises because we compare $\psi(\tilde{p},\tilde{\alpha})$ states transformed from $\psi(p,\alpha)$ and $\psi(\tilde{p},\tilde{\alpha})$ states coming from $\psi(q,\beta)$. The situation is the same between $\psi(\tilde{q},\tilde{\beta})$ and $\psi(p,\alpha)$ and $\psi(q,\beta)$. Note that we do not need to compare the initially given states $\psi(p,\alpha)$ and $\psi(q,\beta)$. Since $T(\tilde{p},\tilde{s};p,s)$ transforms a Dirac spinor $\psi(p,s)$ into the state in its rest frame as an intermediate step, we can evaluate the relevant phase simply in the following way. Draw a sphere of unit radius and specify all four spin states $\hat{\alpha}$, $\hat{\beta}$, $\underset{\sim}{\hat{\alpha}}$, and $\underset{\sim}{\hat{\beta}}$ as points on its surface. Note that our transformation operator of spin states can be represented as an arc of the great circle of this sphere connecting two spin states. After specifying all the paths representing the transformation, we find the path difference between M and $\widetilde{M}$ which form a closed circuit when neglecting the path direction. Using the fact that each spin state is an eigenstate of the polarization operator $\boldsymbol{\sigma}\cdot\hat{s}$, we can evaluate the relative phase factor.

The same idea can be applied to obtain relative phases among different spin configurations for only M. In the helicity case, the closed path just becomes the whole great circle and the relative phase factor is -1. Another rather nontrivial example of the phase factor is $\eta = \frac{1}{2}(1 + \hat{p}\cdot\underset{\sim}{\hat{p}})^{1/2} + \frac{1}{2}i(1 - \hat{p}\cdot\underset{\sim}{\hat{p}})^{1/2}$ for the case when $\hat{\alpha} = \hat{p}$, $\underset{\sim}{\hat{\alpha}} = \hat{p}$, $\hat{\beta} = (-\hat{p}\times\underset{\sim}{\hat{p}})/|\hat{p}\times\underset{\sim}{\hat{p}}|$, and $\underset{\sim}{\beta} = (\hat{p}\times\underset{\sim}{\hat{p}})/|\hat{p}\times\underset{\sim}{\hat{p}}|$.

Note that we do not need to introduce the coordinate system for the evaluation of the phase factors.

[1]We are using the three-dimensional polarization vector $\hat{s}$, which is shown to be equivalent to the more conventional covariant vector S_μ. For the helicity description we have $\hat{S} = \lambda\hat{p}$, where $\lambda = \pm 1$. Note that the values of λ are twice as big as the ordinary value of the helicity of a spin-$\frac{1}{2}$ particle.

[2]G. R. Goldstein and M. J. Moravcsik, Ann. Phys. (N.Y.) **126**, 176 (1980); M. Borghini *et al.*, Phys. Rev. D **17**, 24 (1978).

[3]M. Jacob and G. C. Wick, Ann. Phys. (N.Y.) **7**, 404 (1959).

[4]H. W. Fearing and R. R. Silbar, Phys. Rev. D **6**, 471 (1972). The results of spinorial amplitudes in this paper are given in the form of a fraction whose denominator has a complicated structure with possible zeros, and there are also phase ambiguities.

[5]$\bar{u}(|\mathbf{p}|\underset{\sim}{\hat{p}},\lambda)\gamma_5 u(|\mathbf{p}|\hat{p},\lambda)$ in the helicity scheme and $\bar{u}(|\mathbf{p}|\underset{\sim}{\hat{p}}, -\hat{s})u(|\mathbf{p}|\hat{p},\hat{s})$ in the transversity scheme vanish identically for all angles.

[6]B. Stech, Z. Phys. **144**, 214 (1956); D. M. Fradkin and R. H. Good, Jr., Rev. Mod. Phys. **33**, 343 (1961).

[7]P. A. M. Dirac, *The Principle of Quantum Mechanics* (Oxford U.P., Oxford, 1958), 4th ed.; J. J. Sakurai, *Advanced Quantum Mechanics* (Addison–Wesley, Reading, MA, 1967).

[8]We are using the notation in J. D. Bjorken and S. D. Drell's *Relativistic Quantum Mechanics* (McGraw–Hill, New York, 1964). Specifically our metric is $(1, -1, -1, -1)$ and our γ matrices are in the standard representation. But the final results of spinorial amplitudes are, of course, representation-independent, as they should be.

Orthogonal decompositions of certain finite-dimensional vector spaces

Werner Hässelbarth
Institut für Quantenchemie, Freie Universität Berlin, Holbeinstrasse 48, 1000 Berlin 45, West Germany

(Received 8 August 1983; accepted for publication 6 October 1983)

The vector space of real functions, defined on the set of all mappings of a finite set P into another finite set L, splits into a sum of orthogonal subspaces, one for each subset of P. The orthogonal projections onto these subspaces merely involve averaging operations. Certain linear functional identities are equivalents of k-representability, i.e., of location in the span of those subspaces that belong to subsets of cardinality k. Potential applications refer to complex systems where these results could be used to analyze empirically how their properties depend on properties of their components as well as on the interactions between them. Roughly this amounts to estimating the internal structure of a "black box" from measured properties.

PACS numbers: 02.40. + m, 02.30.Bi

INTRODUCTION

This work will be concerned with the (vector space) structure of the set $\mathbb{R}^{\Omega}$ of real functions, defined on the set $\Omega = L^P$ of all mappings $P \rightarrow L$ between two finite sets of cardinality p and l, respectively. We are going to interpret $\Omega = L^P$ as the *state space* of a composite system and $\mathbb{R}^{\Omega}$ as the space of its *properties* as follows: consider a composite system, built up from p subsystems labeled $1,2,...,p$ each with a finite state space[1] $L_1, L_2, ..., L_p$. Let this system be such that each of its states is completely characterized by specifying the states of all the subsystems. Then its state space is the Cartesian product $L_1 \times L_2 \times \cdots \times L_p$ of the state spaces of its components. Let us further assume that for any state of a subsystem there is an analogous state of each other subsystem. Thus all the state spaces L_i can be mutually identified, and the state space of the composite system takes the particular form of a Cartesian product $L \times L \times \cdots \times L$ of p copies of the same state space L, common to all the subsystems $1,2,...,p$. This object, however, is readily identified with the set L^P of mappings from the set of subsystems $P = \{1,2,...,p\}$ into the set L.

Remark: It is only for the sake of simplicity that we restrict ourselves to Cartesian powers of the same set L. The subsequent results are immediately generalized to $L_1 \times L_2 \times \cdots \times L_p$ instead of $L \times L \times \cdots \times L \cong L^P$.

With L^P taking the part of the state space of a composite system, real functions $f: L^P \rightarrow \mathbb{R}$ are readily interpreted as (real number valued) properties of the system in the sense that for $\varphi \in L^P$ the number $f(\varphi)$ is the value of the property f for the system in its state φ. We may, e.g., consider f to represent a measuring apparatus and $f(\varphi)$ to be the result of the corresponding measurement performed on the system in its state φ.

Our main result will be an orthogonal decomposition,

$$f = \sum_{Q \subseteq P} \tilde{f}_Q ,$$

of any property f into a sum of components $\tilde{f}_Q$, one for each subset $\varnothing \subseteq Q \subseteq P$. With regard to the interpretation sketched above we will refer to the subsets of P as *clusters*, i.e., aggregates of subsystems. Thus by virtue of the foregoing decomposition, a property of the compound system is split up into a sum of contributions due to its clusters of subsystems. Rewritten in the form

$$f = \tilde{f}_{\varnothing} + \sum_{i}^{1,p} \tilde{f}_i + \sum_{i<j}^{1,p} \tilde{f}_{ij} + \sum_{i<j<k} \tilde{f}_{ijk} + \cdots ,$$

the cluster decomposition may be considered as an expansion into a sum of individual contributions $\tilde{f}_i$ of the subsystems, followed by corrections $\tilde{f}_{ij}, \tilde{f}_{ijk}, ...$ due to interactions of increasing complexity: interactions between pairs of subsystems, triples, etc.

For a preliminary impression, let us consider the case of $p = 3$, i.e., a real valued function f of three variables that run through the same finite set L. Its cluster expansion reads as follows:

$$\begin{aligned} f(x, y, z) = \tilde{f}_{\varnothing} &+ \tilde{f}_1(x) + \tilde{f}_2(y) + \tilde{f}_3(z) \\ &+ \tilde{f}_{12}(x, y) + \tilde{f}_{13}(x, z) + \tilde{f}_{23}(y, z) \\ &+ \tilde{f}_{123}(x, y, z) , \end{aligned}$$

where the cluster functions are given by

$$\begin{aligned} \tilde{f}_{\varnothing} &= \langle f(x, y, z) \rangle_{x, y, z} , \\ \tilde{f}_1(x) &= \langle f(x, y, z) \rangle_{y, z} - \tilde{f}_{\varnothing} , \\ \tilde{f}_{12}(x, y) &= \langle f(x, y, z) \rangle_{z} - \tilde{f}_1(x) - \tilde{f}_2(y) - \tilde{f}_{\varnothing} , \end{aligned}$$

with analogous expressions for $\tilde{f}_2, \tilde{f}_3$, and for $\tilde{f}_{13}, \tilde{f}_{23}$, respectively. Here the brackets denote mean values, and the subscripts indicate the variables over which to average.

Finally, to provide an idea of what linear identities and k-representability are about, let us call the previous function of three variables 1-representable if there are real valued functions a,b,c of one variable such that

$$f(x, y, z) = a(x) + b(y) + c(z) .$$

Then f is 1-representable if and only if

$$\begin{aligned} f(x, y, z) = &-2 f(v,v,v) + f(x,v,v) \\ &+ f(v, y, v) + f(v,v,z) \end{aligned}$$

holds as an identity in $x, y, z \in L$, while $v \in L$ can be fixed arbitrarily.

The motivation of studying these objects originates in the theory of chirality functions by E. Ruch and A. Schönhofer,[2] which is concerned with the pseudoscalar properties of chemical compounds that derive from an achiral parent compound by substitution. Here sums of contributions from minimal chiral molecular fragments have been proposed as an approximation, and an associated linear identity was used as a preliminary test of such an approximation for the optical rotatory power of certain methane derivatives.[3]

CLUSTER DECOMPOSITION

Let L and P denote two finite sets of cardinality l and p, respectively. Furthermore, let $\Omega = L^P$ and $X = \mathbb{R}^\Omega$ denote the sets of mappings from P into L and of real functions on Ω. Elements of Ω will be denoted be small Greek letters like $\mu, \rho, \ldots$, those of X by small Latin letters $f, g, \ldots$. With addition of functions and their multiplication by real numbers defined pointwise in the usual fashion, X becomes a vector space over $\mathbb{R}$ of dimension $|\Omega| = l^p$. Moreover we may endow X with the customary scalar product,

$$(f,g) := \sum_{\mu \in \Omega} f(\mu) g(\mu) ,$$

turning it into a Euclidean space. We begin by associating with each subset $Q \subseteq P$ a subset $X_Q \subseteq X$ defined as follows:

$$X_Q := \{ f \in X \mid \mu|_Q = \rho|_Q \Rightarrow f(\mu) = f(\rho) \quad \text{for all} \ \mu, \rho \in \Omega \} ,$$

where the subscript $\cdot|_Q$ denotes the restriction of a mapping from its domain P to $Q \subseteq P$. Evidently, all the X_Q are linear subspaces of X. In particular, $X_P = X$, and $X_\varnothing$ is the subspace of constant functions. The following properties are equally obvious:

$$R \subseteq Q \Rightarrow X_R \subseteq X_Q ,$$

and, as a stronger version,

$$X_Q \cap X_R = X_{Q \cap R} .$$

In the sense of the preceding interpretation X_Q is the subspace of all those properties of the compound system which only depend on the state of the cluster Q, i.e., which are independent of the state of the complementary cluster $P \setminus Q$. How can dependency on the state of a given cluster be removed? By averaging over its states! Accordingly, as candidates for the orthogonal projectors onto the subspaces X_Q, we define mappings $\mathscr{P}_Q$ from X into X_Q by

$$\mathscr{P}_Q : f \mapsto f_Q ,$$

where

$$f_Q(\mu) := \frac{1}{l^{p-q}} \sum_{\substack{\rho \in \Omega \\ \rho|_Q = \mu|_Q}} f(\rho) .$$

Here l, p, and q denote the cardinalities of L, P, and Q. Evidently any such $\mathscr{P}_Q$ is a linear map onto X_Q. Moreover easy calculations show that $\mathscr{P}_Q$ is symmetric, and that the following relation holds:

$$\mathscr{P}_Q \mathscr{P}_R = \mathscr{P}_{Q \cap R} \quad \text{for all} \ Q, R \subseteq P .$$

In particular, this implies that these maps are idempotent. Collecting the results we have proved that $\mathscr{P}_Q$ is the orthogonal projector onto X_Q for all $Q \subseteq P$.

In view of our intention—namely to attribute to any cluster a specific contribution to a property—the orthogonal projections f_Q are not yet what we are looking for. A subspace $X_Q \subseteq X$ contains all the subspaces X_R, $R \subset Q$ in turn. Thus let Σ_Q be their sum (span)

$$\Sigma_Q := \sum_{R \subset Q} X_R ,$$

and let $\tilde{X}_Q$ denote the orthogonal complement of Σ_Q in X_Q, i.e., $\tilde{X}_Q$ is defined by

$$X_Q = \tilde{X}_Q \oplus \Sigma_Q$$

being a direct sum of orthogonal subspaces. Accordingly, any projection $f_Q = \mathscr{P}_Q f$ splits into a sum of two components $\tilde{f}_Q \in \tilde{X}_Q$, $f_\Sigma \in \Sigma_Q$, where f_Σ is that part of f_Q that can be attributed to smaller clusters $R \subset Q$, while $\tilde{f}_Q$ may be considered as a candidate for the specific contribution of Q. The subsequent results confirm that this is the right guess. Explicitly $\tilde{X}_Q$ is given by

$$\tilde{X}_Q = \{ f \in X_Q \mid \mathscr{P}_R f = 0 \quad \text{for all} \ R \subset Q \} .$$

From this characterization it is an easy conclusion that all these spaces are mutually orthogonal

$$\tilde{X}_Q \perp \tilde{X}_S \quad \text{for} \ Q \neq S .$$

With arbitrary elements $\tilde{q} \in \tilde{X}_Q$, $\tilde{s} \in \tilde{X}_S$ it follows from

$$(\tilde{q}, \tilde{s}) = (\mathscr{P}_Q \tilde{q}, \mathscr{P}_S \tilde{s}) = \begin{cases} (\tilde{q}, \mathscr{P}_Q \mathscr{P}_S \tilde{s}) = (\tilde{q}, \mathscr{P}_{Q \cap S} \tilde{s}) \\ (\mathscr{P}_S \mathscr{P}_Q \tilde{q}, \tilde{s}) = (\mathscr{P}_{Q \cap S} \tilde{q}, \tilde{s}) \end{cases} ,$$

that this scalar product vanishes for $Q \neq S$, since $\mathscr{P}_{Q \cap S} \tilde{s}$ vanishes unless $Q \supseteq S$, and $\mathscr{P}_{Q \cap S} \tilde{q}$ vanishes unless $Q \subseteq S$. The same type of argument also shows that

$$\tilde{X}_Q \perp X_S \quad \text{for} \ Q \not\subseteq S .$$

This information is enough to prove our next assertion: For any $Q \subseteq P$ the space X_Q is the direct sum of its subspaces $\tilde{X}_R$, $R \subseteq Q$.

$$X_Q = \sum_{R \subseteq Q}^{\oplus} \tilde{X}_R .$$

The proof starts from the orthogonal decomposition

$$X_Q = \tilde{X}_Q \oplus \Sigma_Q ,$$

where Σ_Q is the span of all the subspaces X_R, $R \subset Q$. First we notice that Σ_Q is already spanned by the subspaces X_R for subsets $R \subset Q$ of cardinality $|Q| - 1$, because any subspace that belongs to a smaller subset of Q is contained in at least one of them. Thus we have

$$X_Q = \tilde{X}_Q \oplus \sum_{\substack{R \subset Q \\ |Q| - |R| = 1}} X_R .$$

The same argument applies to the subspaces X_R

$$X_R = \tilde{X}_R \oplus \sum_{\substack{S \subset R \\ |Q| - |S| = 2}} X_S .$$

Inserting these decompositions into the previous one we arrive at

$$X_Q = \sum_{\substack{R \subseteq Q \\ |Q| - |R| < 1}}^{\oplus} \tilde{X}_R \oplus \sum_{\substack{S \subset Q \\ |Q| - |S| = 2}} X_S .$$

The direct sums are due to the facts that all the subspaces $\tilde{X}_R$ are mutually orthogonal and that they are orthogonal to any X_S, S of smaller cardinality as well. This line of reasoning may be repeated until finally the remainder reduces to $X_\varnothing$, thus finishing the proof.

Next we should like to determine the orthogonal projectors $\tilde{\mathscr{P}}_Q$ onto the subspaces $\tilde{X}_Q$. An immediate consequence of the previous result is

$$\mathscr{P}_Q = \sum_{R\subseteq Q} \tilde{\mathscr{P}}_R .$$

These equations can be inverted, using the Möbius function[4] of the lattice of subsets of P

$$\tilde{\mathscr{P}}_Q = \sum_{R\subseteq Q} (-1)^{|Q|-|R|} \mathscr{P}_R .$$

Direct calculation confirms that these maps are indeed the orthogonal projectors onto the subspaces in question. For example they satisfy

$$\tilde{\mathscr{P}}_Q \tilde{\mathscr{P}}_R = \begin{cases} \tilde{\mathscr{P}}_Q & \text{for } Q=R \\ 0 & \text{for } Q\neq R \end{cases}.$$

Our main interest concerns the particular case of $Q=P$, where $X_P = X$, and $\mathscr{P}_P = 1$ the identity operator of X. Summing up we have proved the following:

Theorem: The Euclidean space $X=\mathbb{R}^{\Omega}$ of real functions, defined on the set $\Omega = L^P$ of all mappings between two finite sets P and L, splits into a sum of orthogonal subspaces $\tilde{X}_Q$, one for each subset $Q\subseteq P$. The family of projection maps $\tilde{\mathscr{P}}_Q$ constitutes the corresponding resolution of the identity map on X

$$X = \sum_{Q\subseteq P}^{\oplus} \tilde{X}_Q , \quad 1 = \sum_{Q\subseteq P} \tilde{\mathscr{P}}_Q .$$

Thus the cluster decomposition of a function $f\in X$ is given by

$$f = \sum_{Q\subseteq P} \tilde{f}_Q ,$$

with

$$\tilde{f}_Q := \sum_{R\subseteq Q} (-1)^{q-r} f_R ,$$

and

$$f_R(\mu) = \frac{1}{l^{p-r}} \sum_{\substack{\rho\in\Omega \\ \rho|_R = \mu|_R}} f(\rho)$$

(l, p, q, r denote the cardinalities of L, P, Q, R.) For computational purposes, i.e., when the function f is given by its graph such as, e.g., a table of measured values is, another expression for its cluster components will be preferred:

$$\tilde{f}_Q(\mu) = \frac{(-1)^q}{l^p} \sum_{r=0}^{q} (1-l)^r \sum_{\substack{\rho\in\Omega \\ |Q\cap D(\mu,\rho)| = r}} f(\rho) .$$

with

$$D(\mu,\rho) := \{ i\in P \mid \mu(i) = \rho(i) \}$$

denoting the subset of P where the maps μ and ρ coincide. Finally we give another characterization of the projection maps $\mathscr{P}_Q$ and $\tilde{\mathscr{P}}_Q$ from which it is obvious how to generalize this theory for $L_1\times L_2\times\cdots\times L_p$ taking the place of L^P. For this purpose we identify a map μ from $P=\{1,2,\ldots,p\}$ into L with the row vector $(\mu_1,\mu_2,\ldots,\mu_p)$ of its images $\mu_i := \mu(i)$. Then

$$[\mathscr{A}_i f](\mu_1,\mu_2,\ldots,\mu_p) := \frac{1}{|L|} \sum_{z\in L} f(\mu_1,\ldots,\mu_{i-1},z,\mu_{i+1},\ldots,\mu_p)$$

defines an averaging operator $\mathscr{A}_i$ with respect to $i\in P$. With these objects, the projectors $\mathscr{P}_Q$ and $\tilde{\mathscr{P}}_Q$ are given by

$$\mathscr{P}_Q = \prod_{j\in P\setminus Q} \mathscr{A}_j ,$$

$$\tilde{\mathscr{P}}_Q = \prod_{i\in Q} (1-\mathscr{A}_i) \prod_{j\in P\setminus Q} \mathscr{A}_j .$$

k-REPRESENTABILITY AND LINEAR IDENTITIES

Let us call a property $f\in X$ to be k-representable if it is a sum of contributions from clusters of size k. Formally, for $0\leqslant k\leqslant p$, $p=|P|$

$$f \text{ is } k\text{-representable} :\Leftrightarrow f\in X^{(k)} ,$$

where

$$X^{(k)} := \sum_{\substack{Q\subseteq P \\ |Q| = k}} X_Q .$$

Equivalently, $f\in X$ is k-representable if it is a sum of contributions from clusters of maximum size k, since $X^{(k)}$ is also given by

$$X^{(k)} = \sum_{\substack{Q\subseteq P \\ |Q| \leqslant k}} X_Q .$$

Replacing the X_Q by $\tilde{X}_Q$, the sum is turned into a direct one

$$X^{(k)} = \sum_{\substack{Q\subseteq P \\ |Q| \leqslant k}}^{\oplus} \tilde{X}_Q .$$

This follows by stepwise reduction to a direct sum according to

$$X^{(k)} = \sum_{\substack{Q\subseteq P \\ |Q| = k}}^{\oplus} \tilde{X}_Q \oplus X^{(k-1)} .$$

This decomposition of $X^{(k)}$ implies that

$$\mathscr{P}^{(k)} := \sum_{\substack{Q\subseteq P \\ |Q| \leqslant k}} \tilde{\mathscr{P}}_Q$$

is the orthogonal projector onto $X^{(k)}$, thus providing us with a test of k-representability

$$f \text{ is } k\text{-representable} \Leftrightarrow f = \mathscr{P}^{(k)} f .$$

These are linear identities of the form

$$\sum_{\sigma\in\Omega} c(\mu,\sigma) f(\sigma) = 0 \quad \text{for all } \mu\in\Omega ,$$

where the coefficients are nonzero for almost all the pairs $\mu,\sigma\in\Omega$ as a rule. There are, however, much "shorter" identities such as the one given in the Introduction. A general method for constructing identities like that as equivalents of k-representability starts from the trivial observation that a vector is located in a subspace if and only if it coincides with its projection onto that subspace, where the projection map is arbitrary. Thus we look for other projectors onto the subspaces $X^{(k)}$ to replace the orthogonal ones.

Fixing an arbitrary map $\rho \in \Omega$, we define a family of maps $\mathscr{R}_Q: X \to X_Q$ by

$$[\mathscr{R}_Q f](\mu) := f(\mu|_Q + \rho|_{P \setminus Q}),$$

where $+$ is abused to denote the composition of maps with disjoint domain, i.e.,

$$[\mu|_Q + \rho|_{P\setminus Q}](i) := \begin{cases} \mu(i) & \text{for } i \in Q \\ \rho(i) & \text{for } i \in P \setminus Q \end{cases}.$$

Evidently, $\mathscr{R}_Q$ is a linear idempotent map onto X_Q, i.e., a projector onto X_Q. Moreover these maps multiply as follows:

$$\mathscr{R}_Q \mathscr{R}_S = \mathscr{R}_{Q \cap S} \quad \text{for all } Q, S \subseteq P,$$

thus exhibiting close analogy with the orthogonal projectors $\mathscr{P}_Q$ (except for the loss of orthogonality). Therefore we proceed along the same lines by defining a family of subspaces $\overline{X}_Q$ of X

$$\overline{X}_Q := \{ f \in X_Q | \mathscr{R}_S f = 0 \quad \text{for all } S \subset Q \}.$$

Consider a set of vectors $\bar{q} \in \overline{X}_Q$, one for each subset $Q \subseteq P$, together with a linear relation with coefficients $c_Q \in \mathbb{R}$

$$\sum_{Q \subseteq P} c_Q \bar{q} = 0.$$

For a subset $S \subseteq P$, application of its projection $\mathscr{R}_S$ leads to

$$0 = \sum_{Q \subseteq P} c_Q \mathscr{R}_S \bar{q} = \sum_{Q \subseteq P} c_Q \mathscr{R}_S \mathscr{R}_Q \bar{q}$$

$$= \sum_{Q \subseteq P} c_Q \mathscr{R}_{Q \cap S} \bar{q} = \sum_{Q \subseteq S} c_Q \bar{q}.$$

Within the collection of subsets $Q \subseteq P$ with $c_Q \neq 0$, let S be a minimal element. Then the preceding line gives $c_S = 0$, thus proving by contradiction that there are no linear dependencies among vectors of different subspaces $\overline{X}_Q$. Thus any sum of subspaces $\overline{X}_Q$ is a direct one. Moreover any sum of a subspace X_S and of subspaces $\overline{X}_Q$ such that $Q \not\subseteq S$ is a direct one. This again suffices to prove that X_Q is the direct sum of its subspaces $\overline{X}_S$, $S \subseteq Q$.

$$X_Q = \sum_{S \subseteq Q}^{\oplus} \overline{X}_S.$$

In contrast with the former situation where we could make use of a well-known fact about orthogonal projections (namely that the projection onto a sum of orthogonal subspaces is the sum of their projections) and afterwards apply Möbius inversion, this time one has to show that the operators

$$\overline{\mathscr{R}}_Q := \sum_{S \subseteq Q} (-1)^{|Q| - |S|} \mathscr{R}_S$$

are indeed projection maps onto the subspaces $\overline{X}_Q$, and that they add up according to

$$\mathscr{R}_Q = \sum_{S \subseteq Q} \overline{\mathscr{R}}_S.$$

This is readily verified by direct calculation. The analogy of results also extends to

$$X^{(k)} = \sum_{\substack{Q \subseteq P \\ |Q| < k}}^{\oplus} \overline{X}_Q,$$

and to

$$\mathscr{R}^{(k)} := \sum_{\substack{Q \subseteq P \\ |Q| < k}} \overline{\mathscr{R}}_Q$$

being a projection map onto $X^{(k)}$. In consequence

$$f \text{ is } k\text{-representable} \Leftrightarrow f = \mathscr{R}^{(k)} f.$$

After some arithmetic one obtains the following expression[5] for $\mathscr{R}^{(k)}$, $0 \leqslant k \leqslant p-1$:

$$\mathscr{R}^{(k)} = \sum_{q=0}^{k} (-1)^{k-q} \binom{p-q-1}{k-q} \sum_{\substack{Q \subseteq P \\ |Q| = q}} \mathscr{R}_Q,$$

which provides a family of linear identities, equivalent to k-representability.

Theorem:

f is k-representable $(0 \leqslant k \leqslant p-1)$

$$\Updownarrow$$

$$f(\mu) = \sum_{q=0}^{k} (-1)^{k-q} \binom{p-q-1}{k-q} \sum_{\substack{Q \subseteq P \\ |Q| = q}} f(\mu|_Q + \rho|_{P \setminus Q})$$

for all $\mu \in \Omega$.

The map $\rho \in \Omega$ is arbitrary. The introductory example refers to a constant map $\rho = (v,v,v)$. With the same ρ, a function of three variables is 2-representable, i.e., there exist functions a', b', c' of two variables such that

$$f(x,y,z) = a'(x,y) + b'(x,z) + c'(y,z)$$

if and only if

$$f(x,y,z) = f(v,v,v) - f(x,v,v) - f(v,y,v) - f(v,v,z)$$
$$+ f(x,y,v) + f(x,v,z) + f(v,y,z).$$

These identities can be used for a preliminary test of how accurate an approximation to a function f its k-representable part $f^{(k)} := \mathscr{P}^{(k)} f$ is, namely by sample checking of how badly the associated identity is broken.

SUMMARY

The Euclidean space $X = \mathbb{R}^{\Omega}$ of real functions, defined on the set $\Omega = L^P$ of all mappings from a finite set P into another finite set L, splits into a direct sum of subspaces $\widetilde{X}_Q$, one for each subset $Q \subseteq P$. The associated orthogonal projections are given in closed form, involving merely averaging operations. Certain linear identities are equivalents of k-representability, i.e., of location in the span of those subspaces that belong to subsets $Q \subseteq P$ of cardinality k.

A typical application that we have in mind would be an analysis of large sets of experimental data of a molecular property, yielding empirical rules on how its numerical values depend on molecular structure. The cluster decomposition then provides an empirical definition of the contributions due to various molecular fragments, and linear identities may be used for estimating the degree of complexity of those fragments that have to be accounted for when using a truncated cluster expansion as an approximation.

[1]Finite space and finite set are used synonymously.
[2]E. Ruch and A. Schönhofer, Theor. Chim. Acta **19**, 225 (1970).
[3]W. J. Richter, B. Richter, and E. Ruch, Angew, Chem. Int. Ed. Engl. **12**, 30 (1973).
[4]G.-C. Rota, Z. Wahrscheinlichkeitstheorie **2**, 340 (1964) as cited in C. Berge, *Principles of Combinatorics* (Academic, New York 1971).
[5]Where $\binom{0}{0} := 1$.

Generalized supermanifolds. I. Superspaces and linear operators

J. Hoyos
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Matemáticas, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

M. Quirós
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain

J. Ramírez Mittelbrunn and F. J. de Urríes
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Fiśica, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

(Received 26 November 1982; accepted for publication 30 September 1983)

Generalized superspaces are defined taking as a starting point the concept of Grassmann–Banach algebra. The category of B-spaces with L_B-maps as morphisms is introduced. This allows one to discuss linear and tensor algebra on generalized superspaces.

PACS numbers: 02.40. + m, 11.30.Pb

1. INTRODUCTION

In the last years there has been a lot of interest in symmetry transformations mixing commuting and anticommuting objects. The two main examples of this kind of transformation are supersymmetries and Becchi–Rouet–Stora (BRS) transformations. The later are relevant in quantized nonabelian gauge theories, while supersymmetric transformations provide the possibility of enclosing particles of different spin in the same multiplet. The common feature from the mathematical point of view is that both of them are generated by a graded Lie algebra spanned by commuting and anticommuting generators. Accordingly, finite transformations should have parameters belonging to a Grassmann algebra.

The role of Grassmann algebras was strengthened by the superspace approach to supersymmetries in terms of superfields, introduced by Salam and Strathdee.[1] The concept of superspace, heuristically introduced in Ref. 1 as a collection of eight coordinates belonging to a Grassmann algebra, has been formalized and generalized by several authors.

There are essentially two approaches to a formalized concept of superspaces and supermanifolds. The first is due to Berezin and Leites[2] and Kostant.[3] The method followed by them consists in replacing the algebra $C^\infty(X)$ of a real C^∞-manifold X by a graded algebra $A(X) \simeq C^\infty(X) \otimes \Lambda^q$, where Λ^q is the exterior algebra expanded by q generators.[4] All the elements needed to do differential geometry as vector fields, differential operators, etc., are extended from $C^\infty(X)$ to $A(X)$ following an algebraic path. Therefore, supermanifolds are just ordinary real manifolds with a ring of functions enriched with anticommuting elements. So, such things as "points" of superspace with commuting and anticommuting coordinates do not exist in that approach.

The second approach to superspaces and supermanifolds has been introduced by Rogers.[5] Rogers's superspaces are products of even and odd parts of Grassmann algebras. Analysis on them is a generalization of real analysis, for which linear approximation of functions (differential) is done with respect to the Grassmann product. Accordingly G^∞-supermanifolds are topological space locally G-diffeomorphic to superspaces. This point of view provides a very natural way for generalizing the results of ordinary differential geometry to differential "supergeometry." It has also the advantage that supermanifolds admit nontrivial topologies in the anticommuting sector.[6] On the other hand, there is room for infinite-dimensional Grassmann algebra that avoids undesirable restrictions (vanishing for a sufficiently high number of fermionic fields) for Green's functions of fields valued in those algebras, as stressed in Ref. 5. Finally Berezin–Leites–Kostant manifolds can be realized as G^∞-supermanifolds.

We shall take Rogers' approach as a starting point for our generalized supermanifolds. However, we shall improve some unsatisfactory points of that work. In particular, the extension theorem is valid only for a particular class of open sets, as was pointed out by Jadczyk and Pilch.[7] On the other hand, the existence of a body manifold is not guaranteed for every G^∞-supermanifold.

In this series of papers we shall generalize the concepts of superspace and supermanifold as follows: Instead of considering a superspace as a direct product of m even and n odd copies of a given Grassmann algebra, we shall allow different coordinates to belong to different Grassmann algebras. In this way we arrive at superspaces and supermanifolds with inhomogeneous coordinates which include the usual ones as particular cases. Also the class of B-spaces, wider than that of superspaces, is introduced to deal with tensor products of superspaces. Although this generalization could be well justified from the mathematical point of view we can further argue that it offers a much broader field of application in physics. In fact, the generalized superspaces have proved useful in the study of the geometrical structure of quantized gauge (BRS invariant) theories.[6] We hope that the generalized supermanifolds will be useful in supersymmetric theories.

Because of the length of this work, we shall divide it into three papers. We shall refer to them from now on as I, II, and III, respectively. In Paper I we study generalized superspaces, B-spaces, and their linear algebra using as a basic building block a Grassmann–Banach algebra. Paper II is devoted to the analysis on generalized superspaces with emphasis on the concept of G-differentiability and in the expansion of G^∞-functions in terms of C^∞-functions (superfield expansion). In Paper III, the class of ρ-supermanifolds (manifolds with a well-behaved body manifold) is defined, and the geometrical objects associated with them as vector superfields, etc., are studied. Also the elements of the theory of Lie supergroups and principal superfiber bundles is developed within the context of ρ-supermanifolds.

In Sec. 2 of Paper I, we introduce the concept of Grassmann–Banach algebra as a Grassmann algebra endowed with a Banach algebra structure in such a way that the Grassmann product is continuous with respect to the Banach norm. The natural concept of Grassmann–Banach algebra, here introduced, is in contrast with that of Banach–Grassmann algebra recently introduced by Jadczyk and Pilch,[7] for which Grassmann structure (existence of anticommuting generators) is not explicitly required. In fact, finite-dimensional Grassmann algebras do not fulfill the conditions of Jadczyk and Pilch, but they are particular cases of our definition of Grassmann–Banach algebras. However, infinite-dimensional Grassmann algebras given in Ref. 5 do satisfy both definitions. In Sec. 3 we introduce the concept of graded B-module and define generalized superspaces and B-spaces modeled on B-modules. In Sec. 4, B-linear operators between B-spaces and their matrix representations are studied with special attention for the case of superspaces. Finally, in Sec. 5 we deal with supertensor algebra and B-multilinear operators.

2. THE GRASSMANN–BANACH ALGEBRA

As is well known a finite-dimensional Grassmann algebra is an algebra generated by n anticommuting elements. So long as we are involved in the study of functions defined on Grassmann algebras, we need some topology on them. So we also will consider them as Banach algebras[8] (Banach spaces with associative continuous product). Both structures will be married by the concept of Grassmann–Banach algebras which will be developed in this section.

Definition 2.1: A Z_2-graded Banach algebra is an associative Banach algebra B with unity 1, such that B can be decomposed as the direct sum $B = B^0 \oplus B^1$ of two closed Banach subspaces and $1 \in B^0$, $B^r B^s \subset B^{r+s}$ [r,s and sums of them are always considered mod 2]. We shall call $a \in B$ homogeneous if $a \in B^r$. We define the degree of an homogeneous element $a \in B$ as $|a| = r$ if $a \in B^r$. B is commutative if $ab = (-1)^{|a|\,|b|} ba$, for $a,b \in B$ homogeneous.

We only will be concerned with Banach spaces with countable (possibly finite) bases. Nevertheless, we are aware that the case of noncountable basis, and hence nonseparable Banach spaces, might also be considered. So, from here on, J will be a fixed (finite or countably infinite) set of indices, $P(J)$ the set of parts of J, and $F = F(J)$ the set of finite parts of J. Since it is countable, it can be given by $J = \{1,2,\dots,n\}$ or $J = \{1,2,\dots\} = \mathbb{N}$. This order on J induces an order on F via the injection $j: F \to \mathbb{N} \cup \{0\}$ given by $j(\varnothing) = 0$, $j(\{r_1,\dots,r_k\}) = \frac{1}{2}\sum_{i=1}^{k} 2^{r_i}$; j is onto if $J = \mathbb{N}$. A property of this order is: if $M \in F(J_m)$, $(J_m) = 1,\dots,m$ and $N \notin F(J_m)$, then $M < N$.

Definition 2.2: A Grassmann–Banach algebra is a Z_2-graded commutative Banach algebra, with the following properties:

(i) There is a linearly independent subset $\{\beta_i\}_{i \in J} \subset B^1$ such that $\{\beta_M\}_{M \in F}$ is a Banach space basis of B, and for each $M = \{i_1,\dots,i_r\} \in F$, $i_1 < \cdots < i_r$, $\beta_M = \beta_{i_1} \beta_{i_2} \cdots \beta_{i_r} \neq 0$, and $\beta_\varnothing = 1$.

(ii) For each $H \subset F$ and $a = \sum_{M \in F} a_M \beta_M \in B$ we define $p_H(a) = \sum_{M \in H} a_M \beta_M$ (i.e., we take $a_M = 0$ if $M \notin H$). Then $p_H: B \to B$ is a linear function from B to B.

It would seem natural to impose $\|\beta_M\| = 1$ and p_H continuous with $\|p_H\| = 1$. However, we will not need these conditions for our purpose and, we will see anyway that p_H always is a continuous projection. We shall refer in general to p_H projections as β-projections. All Grassmann algebras considered by Rogers[4] endowed with the structure of Banach algebra given there are examples of Grassmann–Banach algebras with the above properties.

Proposition 2.3: For each $H \subset F$, $p_H: B \to B$ is a continuous linear projection. Furthermore, $B = p_H(B) \oplus p_{H'}(B)$, $H' = F - H$, and $p_{H'} = I - p_H$, I is the identity on B.

Proof: Because the coordinate maps $(a \to a_M)$ are continuous linear,[8] we have that $p_H(B)$ and $p_{H'}(B)$ are closed subspaces of B. Also $p_H(B) \cap p_{H'}(B) = \{0\}$ and p_H is a linear projection. Hence, we only need to prove that $p_H(B) + p_{H'}(B) = B$. Let $a = \sum_{M \in F} a_M \beta_M \in B$; by property (ii) of Definition 2.2, $a_1 = \sum_{M \in H} a_M \beta_M \in p_H(B) \subset B$, $a_2 = \sum_{M \in H'} a_M \beta_M \in p_{H'}(B) \subset B$ and $a = a_1 + a_2$ taking limits on finite approximations. Q. E. D.

Let us point out some consequences of Definition 2.2.

(1) All β_M are homogeneous and $\beta_M \in B^r$ if $r = \operatorname{card}(M)$.

(2) $\beta_M \beta_N = \begin{cases} 0 & \text{if } M \cap N \neq \varnothing, \\ \pm \beta_{M \cup N} & \text{if } M \cap N = \varnothing, \end{cases}$

the sign being determined by the permutation which restores natural order of the β_i's.

(3) If $H = \{M \in F \mid \operatorname{card}(M) = 0\}$, then $B^0 = p_H(B)$, $B^1 = p_{H'}(B)$.

(4) If $H = \{\varnothing\} \subset F$ then $p_H(B) = \mathbb{R}$. Let $B' = p_{H'}(B)$; then $B = \mathbb{R} \oplus B'$ and $B' = B^{0'} \oplus B^1$, where $B^{0'} = B' \cap B^0$. We will call $r \equiv p_H$ ($s \equiv p_{H'}$) the body (soul) map and for each element $a \in B$, $r(a)$ ($s(a)$) will be called the body (soul) of a. We also have that r is an algebra homomorphism.

Let B'' be the set of nilpotent elements of B; since r is an algebra homomorphism, $B'' \subset B'$.

Proposition 2.4: B'' is a dense subset of B'. Furthermore, $B^1 \subset B''$ and for $a \in B^1$, $a^2 = 0$.

Proof: The last statement is a direct consequence of the commutativity property of Definition 2.1.

To prove the first part, we observe that if $H \subset F(J_m)$, $m \in N$, then $p_H(B) \cap B' \subset B''$, since for $a \in p_H(B) \cap B'$ $a^{m+1} = 0$. Now, let $a \in B'$ and $\epsilon > 0$; then there exists $M_\epsilon \in F$ such that $\|a - \sum_{M < M_\epsilon} a_M \beta_M\| \leqslant \epsilon$ and if $m_\epsilon = \max M_\epsilon$ then $\{M \in F / M \leqslant M_\epsilon\} \subset F(J_{m_\epsilon})$. Hence $\sum_{M < M_\epsilon} a_M \beta_M \in B''$. Q.E.D.

Proposition 2.5: (i) For each $a \in B'$ and $0 < \theta < 1$, $\exists\, \lambda \geqslant 0$ such that $\forall\, n \geqslant 0$ $\|a^n\| \leqslant \lambda \theta^n$.

(ii) $a \in B$ is invertible iff $r(a) \neq 0$. Furthermore,

$$a^{-1} = r(a)^{-1} \sum_{n=0}^{\infty} (-1)^n \left(\frac{s(a)}{r(a)}\right)^n. \tag{1}$$

Proof: In Lemma 2.7b of Ref. 4 it has been proved that for each $a \in B'$ $\exists\ 0<\theta<1, \lambda \geqslant 0$ such that $\|a^n\| \leqslant \lambda\theta^n\ \forall\ n \geqslant 0$. We now see that this implies (i). Let $a \in B'$, $0<\theta<1$. Then $b = ta \in B'$ so there exists $0<\theta'<1$, $\lambda \geqslant 0$ such that $\|b^n\| \leqslant \lambda\theta'^n$. Hence $\|a^n\| \leqslant \lambda(\theta'/t)^n \leqslant \lambda\theta^n$ for $t \geqslant \theta^{-1}$. For (ii) see Proposition 3.1 of Ref. 5. Q.E.D.

In this work we shall consider a special class of subalgebras associated with each subset $K \subset J$, defined as follows: $B_K = p_{F(K)}(B)$. In addition we have that $p_{F(K)}: B \rightarrow B_K$ is an algebra homomorphism and B_K is a Grassmann–Banach algebra.

We conclude this section with a lemma that will be useful for the expansion of superfunctions (see II).

Let K_i be subsets of J for $i = 1,...,s$ and $A = B^1_{K_1} B^1_{K_2} \cdots B^1_{K_s} = \{a_1,...,a_s \in B \mid a_i \in B^1_{K_i}\}$. Let E be the annihilator of A, i.e., $E = \{b \in B \mid ab = 0\ \forall\ a \in A\}$.

Lemma 2.6: E is a closed ideal of B and there exists a continuous projection p from B onto E. In fact, p is a β-projection and B and $E = p_H(B)$ for a subset $H \subset F(J)$.

Proof: From the definition of E, it follows immediately that E is a closed ideal of B. To prove the existence of p we will show that a $H \subset F(J)$ exists such that $E = p_H(B)$. Then $p \equiv p_H$.

In particular when a subset K_i is infinite $E\{0\}$ and $p = 0$, and when $K_i = \varnothing$, $A = \{0\}$, $E = B$, and $p = I$.

Let $H = \{K \in F(J) \mid \beta_K \in E\}$. Then $p_H(B) \subset E$ follows since E is a closed subspace of B. To prove that $E \subset p_H(B)$, we only have to show that for each $a = \Sigma_{M \in F(J)}\, a_M\, \beta_M \in E$, $a_K = 0$ if $K \notin H$. In fact, if $K \notin H$, there exists $b \in A$ such that $\beta_K b \neq 0$. Then, as each element $b \in A$ is of the form $b = \Sigma_{j=1}^{s} \Sigma_{i_j \in K_j} b_{i_1,...,i_s}\, \beta_{i_1},...,\beta_{i_s}$ with $b_{i_1,...,i_s} \in B^0$, if $K \notin H$ there exists $i_j \in K_j$ for $j = 1,...,s$ such that $\beta_k\, \beta_{i_1},...,\beta_{i_s} \neq 0$, and since $\beta_{i_1},...,\beta_{i_s} \in A$, $0 = a\beta_{i_1},...,\beta_{i_s} = \Sigma_{M \in F(J)}\, a_M(\beta_M\, \beta_{i_1},...,\beta_{i_s})$ follows. Now $\beta_M\, \beta_{i_1},...,\beta_{i_s} = 0$ if $M \cap \{i_1,...,i_s\} \neq \varnothing$ and $\beta_M\, \beta_{i_1},...,\beta_{i_s} = \pm\beta_{M \cup \{i_1,...,i_s\}}$ if $M \cap \{i_1,...,i_s\} = \varnothing$. Then, from the uniqueness of the expression in the basis of the Grassmann–Banach algebra $a_M = 0$ if $\beta_M\, \beta_{i_1},...,\beta_{i_s} \neq 0$. Hence, $a_K = 0$. Q. E. D.

3. SUPERSPACES AND *B*-SPACES

Let B be a Grassmann–Banach algebra.

Definition 3.1: A Banach Z_2-graded left B-module is a Banach space V that is also a left B-module and can be decomposed as a direct sum of Banach subspaces $V = V^0 \oplus V^1$, with $B^r V^s \subset V^{r+s}$. In addition $\|av\| \leqslant \|a\|\ \|v\|$, $a \in B$, $v \in V$.

Let us remark that V^r are not B-submodules but only B^0-submodules. In a similar way we could define a Banach Z_2-graded right B-module. Moreover, since B is graded commutative, each right B-module can be considered as a left B-module, and conversely, in the form

$$a_r v_s = (-1)^{rs} v_s a_r, \quad a_r \in B^r,\ v_s \in V^s.$$

Definition 3.2: A free Z_2-graded B-module is a B-module V having a basis whose elements are homogeneous.

A finite-dimensional free Z_2-graded B-module is said to be (m,n)-generated if it has a basis with m even and n odd elements. From now on B-modules will always be understood as Banach Z_2-graded free B-modules, if the other is not specified.

Let V be a left B-module with the basis $\{e_1,...,e_m,e_{m+1},...,e_{m+n}\}$, where e_i is even for $1 \leqslant i \leqslant m$ and odd for $m+1 \leqslant i \leqslant m+n$. Then, the even and odd parts of V can be expressed as

$$V^r = \left\{\sum_{i=1}^{m+n} a^i e_i \,\middle|\, a^i \in B \text{ and } |a^i| + |e_i| = r\right\}. \tag{2}$$

The simplest example of a B-module is B^{m+n} with the canonical basis $e_i = (\delta_i^j)_{j=1}^{m+n}$, considering that e_i is even for $1 \leqslant i \leqslant m$ and odd for $m+1 \leqslant i \leqslant m+n$ and the even and odd parts of B^{m+n} are defined by (2). The Banach structure of B^{m+n} is the usual product of Banach spaces. Moreover, from the property $\|av\| \leqslant \|a\|\ \|v\|$ of Definition 3.1 and Banach's open mapping theorem (8), V is naturally isomorphic to B^{m+n}.

Definition 3.3: The body map associated with a fixed basis $\{e_i\}_{i=1}^{m+n}$ of V is the map

$$R\left(\sum_{i=1}^{m+n} a^i e_i\right) = (r(a^1),...,r(a^{m+n})). \tag{3}$$

R is linear, continuous, and $R(av) = r(a)R(v)$ for $a \in B$, $v \in V$. We observed that if the basis $\{e_i\}_{i=1}^{m+n}$ is changed, the map R, associated with $\{e_i\}_{i=1}^{m+n}$, is not given by (3) in the new basis.

Now we define the class of B-spaces, containing the superspaces as a subclass. The need for B-spaces will become clear when dealing with linear operators between superspaces and tensor algebras. In fact, we are going to define a category whose objects are B-spaces and whose morphisms are L_B-operators.

Definition 3.4: Given a $m+n$ sequence $\pi = (p_1, p_2,...,p_{m+n})$ of β-projections of B (see 2.2). Let $V(\pi)$ be the Banach subspace of V,

$V(\pi) = \left\{\sum_{i=1}^{m+n} a^i e_i \in V \,\middle|\, a_i \in \operatorname{Im} p_i\right\}$. A B-space dimension is a couple $(\pi,\sigma) = ((p_i),(q_i))$ of $m+n$ sequences of β-projections such that $q_i \cdot p_i = q_i$ (i.e., $\operatorname{Im} q_i \subset \operatorname{Im} p_i$). A B-space of dimension (π,σ) is the quotient Banach space $V(\pi,\sigma) = V(\pi)/V(\sigma)$.

Remark: $V(\pi,\sigma)$ inherits a structure of Z_2-graded vector space from V. In fact $V(\pi)^0 = V(\pi) \cap V^0$ and $V(\pi)^1 = V(\pi) \cap V^1$. Also $V(\pi,\sigma)^r = V(\pi)^r / V(\sigma)$, so $V(\pi,\sigma) = V(\pi,\sigma)^0 \oplus V(\pi,\sigma)^1$. Clearly $V(\pi,\sigma)^0$ and $V(\pi,\sigma)^1$ are B-spaces themselves. Let us observe that when some coordinates of $V(\pi,\sigma)$ vanish, it is possible to consider it modeled in a B-module of lower dimension.

Definition 3.5: Let $V(\pi,\sigma)$ be a B-space, F_c a real vector space, and t a linear map from R^{m+n} onto F_c; then the linear map $t \cdot R|_{V(\pi)}$ from $V(\pi)$ into E factorizes to a linear map c from $V(\pi,\sigma)$ into F_c. We say that c is a k-body map of $V(\pi,\sigma)$ and that F_c is a body of $V(\pi,\sigma)$ if c is onto and $k = \dim F_c$.

$$\begin{tikzcd} V(\pi) \arrow[r, "t \cdot R|_{V(\pi)}"] \arrow[d, "p"'] & F_c \\ V(\pi,\sigma) \arrow[ur] & \end{tikzcd} \tag{4}$$

p is the quotient projection.

We call B-space with body to a couple $(V(\pi,\sigma),c)$. In what follows only bodied B-spaces will be considered. We observe that (V,R) is a bodied B-space with dimension $p_i = I$, $q_i = 0$, $1 \leqslant i \leqslant m+n$. Next we define the superspaces that will be the basic objects of our theory, as bodied B-spaces.

Definition 3.6: A superspace is a bodied B-space (S,r), where $S = V(\pi,\sigma)^0$, $p_i = P_{F(K_i)}$, $\varnothing \neq K_i \subset J$ for $m+1 \leqslant i \leqslant m+n$, and r is defined by $t\colon \mathrm{R}^{m+n} \to R^m$, $t(x_1,\ldots,x_{m+n}) = (x_1,\ldots,x_m)$.

The superspace S is just the Banach subspace $\{\Sigma_{i=1}^{m+n} a^i e_i \in V \mid a^i \in B_{K_i}^{r_i}\}$ of V. So S is isomorphic to $B_{K_1}^0 \times \cdots \times B_{K_m}^0 \times B_{K_{m+1}}^1 \times \cdots \times B_{K_{m+n}}^1$. Then a superspace S has m even and n odd coordinates, and its dimension can be characterized by the sequence $\mathscr{K} = (K_1,\ldots,K_{m+n})$. Consequently we shall often use for a superspace of dimension $\mathscr{K}$ the notation $S^{m,n}(\mathscr{K})$.

4. L_B-OPERATORS

In this section we shall introduce the morphisms of the category of B-spaces, and study their main properties. We begin with B-linear operators between B-modules.

A. B-Linear operators between B-modules

Let V and V' be B-modules with bases $\{e_i\}_{i=1}^{m+n}$ and $\{e'_j\}_{j=1}^{m'+n'}$, respectively.

Definition 4.1. A B-linear operator from V into V' is a continuous linear operator $T \in L(V,V')$ such that $(av)T = a(vT)$, $a \in B$, $v \in V$. The set of B-linear operators from V into V' will be denoted by $L_B(V,V')$.

Let us remark that a linear operator from V to V' with B-linearity conditions, is automatically continuous.

We give to $L_B(V,V')$ a B-module structure. The scalar multiplication is given by $u(Ta) = (uT)a$. The degree is determined by $|T| = r$ iff $V^s T \subset V'^{r+s}$, $r,s = 0,1$. A basis is given by T_j^i, $e_k T_j^i = \delta_k^i e'j$. $L_B(V,V')$ is a closed subspace of $L(V,V')$ and so it is a Banach space. The dimension of $L_B(V,V')$ as free B-module is $(mm' + nn', mn' + m'n)$. We observe that the composition of operators is written in the following way: $u(S \cdot T) = (uS)T$ (Russian notation).

As a B-module $L_B(V,V')$ has the body map R defined on it, given $T \in L_B(V,V')$ we can consider $R(T)$ as a linear map from R^{m+n} into $R^{m'+n'}$ that we shall call the body operator. More precisely:

Lemma 4.3: Let $T \in L_B(V,V')$ and $u,v \in V$ such that $R(u) = R(v)$; then $R(uT) = R(vT)$.

Proof: It is enough to prove that if $R(u) = 0$ then $R(uT) = 0$. Let $u = \Sigma_{i=1}^{m+n} a^i e_i$; then $R(u) = 0$ is equivalent to $r(a^i) = 0$ $1 \leqslant i \leqslant m+n$. But $uT = \Sigma_{i=1}^{m+n} a^i(e_i T)$ and $R(uT) = \Sigma_{i=1}^{m+n} r(a^i) R(e_i T) = 0$. Q. E. D.

Definition 4.4: Let $T \in L_B(V,V')$. We define the body operator of T, $T_R \in L(R^{m+n}, R^{m'+n'})$, by $T_R(x) = R(uT)$ if $x = R(u) \in R^{m+n}$, $u \in V$.

This definition is consistent with Lemma 4.3 and the idea behind both is expressed by the following commutative diagram:

$$
\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "R"'] & V' \arrow[d, "R"] \\
R^{m+n} \arrow[r] & R^{m'+n'}
\end{tikzcd}
\tag{5}
$$

Next we shall give matrix expressions for B-linear operators. As usual $e_i T = \Sigma_{j=1}^{m'+n'} a_i^j e'_j$ so that the matrix $A = (a_i^j)$ represents the operator T. Here j is the column index and i the row index, so vectors in V (and V') are expressed as a row. Matrices with entries in a superalgebra are usually referred to in the literature as supermatrices. The rules for calculus with supermatrices can be found in the paper by Leĭtes.[2]

We also have that the body of a B-linear operator T, represented by the matrix $A = (a^j{}_i)$, is represented by the matrix $R(A) = (r(a^j{}_i))$. R is an algebra homomorphism, i.e., $R(A+B) = R(A) + R(B)$ and $R(AB) = R(A)R(B)$.

A B-isomorphism from V onto V' is a B-linear operator. $T \in L_B(V,V')$ which is one-to-one and onto. This is just a linear isomorphism that is in $L_B(V,V')$. Also, T is a B-isomorphism iff there exists $S \in L_B(V,V')$ such that $T \cdot S = I$, $S \cdot T = I$. We can characterize B-isomorphisms by means of their body. The following theorem generalizes Proposition 2.5.

Theorem 4.5: Let $T \in L_B(V)$ and A be the matrix expression of T. Then

(i) If $R(A) = 0$, $\forall\, 0 < \theta < 1$ $\exists\, \alpha \geqslant 0$ such that $\|A^n\| \leqslant \alpha\theta^n$ $\forall\, n \geqslant 0$.

(ii) T is a B-isomorphism iff T_R is a linear isomorphism. In addition, if $S(A) = A - R(A)$,

$$
\begin{aligned}
A^{-1} &= (R(A))^{-1} \sum_{n=0}^{\infty} (-1)^n (S(A)\, R(A)^{-1})^n \\
&= \left\{ \sum_{n=0}^{\infty} (-1)^n (R(A)^{-1} S(A))^n \right\} R(A)^{-1}.
\end{aligned}
\tag{6}
$$

Proof: (i) Given $\delta > 0$ we can find $p \in \mathrm{N}$ such that A can be decomposed as $A = A_1 + A_2$, where entries of matrix A_1 belong to the finite-dimensional subalgebra B_{J_p} and $\|A_2\| < \delta$. Now

$$
\begin{aligned}
\|(A_1 + A_2)^{p+h}\| &\leqslant \Sigma \|A_1^{i_1} A_2^{i'_1} \cdots A_1^{i_r} A_2^{i'_r}\| \\
&\leqslant \Sigma' \|A_1\|^{\Sigma_{j=1}^r i_j} \|A_2\|^{\Sigma_{j=1}^r i'_j} \\
&< \sum_{k=0}^{p} \binom{p+h}{k} \|A_1\|^k \delta^{p+h-k},
\end{aligned}
\tag{7}
$$

where Σ is extended over the sequences of nonnegative numbers $(i_1, i'_1, \ldots i_r, i'_r)$ such that $\Sigma_{j=1}^r (i_j + i'_j) = p + h$ and Σ' over the sequences in Σ such that $\Sigma_{j=1}^r i_j \leqslant p$. The restriction from Σ to Σ' is due to the fact that $A_1^{i_1} A_2^{i'_1} \ldots A_1^{i_r} A_2^{i'_r} = 0$ if $\Sigma i_j \geqslant p + 1$, since the elements of A_1 only contain p generators. From here on the proof follows as in Lemma 2.7 of Ref. 5 and in Proposition 2.5.

(ii) If T is a B-isomorphism, there exists a B such that $A \cdot B = I$ and $B \cdot A = I$. So, since R is an algebra homomorphism $R(A)R(B) = I$ and $R(B)R(A) = I$ and, hence, T_R is a linear isomorphism.

Conversely, if T_R is a linear isomorphism, the expression for the inverse matrix A^{-1} (Ref. 6), has a sense and the series converges by (i).

Proposition 4.6 (dimension theorem): Given a B-isomorphism $T \in L_B(V,V')^0$ $[T \in L_B(V,V')^1]$ we have $m' = m$

and $n' = n$ [$m' = n$ and $n' = m$].

Proof: If T is a B-isomorphism by (ii) of Theorem 4.5 $m' + n' = m + n$. Now if T is even the matrix A can be expressed in boxes as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{8}$$

where A_{11} and A_{22} have even entries and dimensions $m \times m'$ and $n \times n'$, respectively, while A_{12} and A_{21} have odd entries. Therefore,

$$R(A) = \begin{pmatrix} R(A_{11}) & 0 \\ 0 & R(A_{22}) \end{pmatrix} \tag{9}$$

and, since $R(A)$ is a linear isomorphism, it follows that $m' = m$ and $n' = n$. For odd T, $R(A_{11}) = R(A_{22}) = 0$. Hence $m' = n$ and $n' = m$. Q.E.D.

B. L_B-operators between B-spaces

We now proceed to define B-linear operators between B-spaces.

Definition 4.7: Given $F = V(\pi,\sigma)$ and $F' = V'(\pi',\sigma')$ B-spaces and $T \in L(F,F')$ we say that T is L_B iff there exists $T' \in L(V(\pi), V(\pi))$ and $\overline{T} \in L_B(V,V')$ such that $pT = T'p'$ and $i\overline{T} = T'i'$, where i,i' and p,p' are, respectively, the inclusions and projections of the diagram below.

$$\begin{tikzcd} V \arrow[r, "\overline{T}"] & V' \\ V(\pi) \arrow[u, "i"] \arrow[r, "T'"] \arrow[d, "p"'] & V(\pi') \arrow[u, "i'"'] \arrow[d, "p'"] \\ F \arrow[r, "T"] & F' \end{tikzcd} \tag{10}$$

As a direct consequence of this definition T' is L_B when we consider $V(\pi)$ and $V' = (\pi')$ as B-spaces.

Now we identify canonically $L_B(F,F')$ with a B-space modeled in $L_B(V,V')$. To do this, we introduce first $L_B(V, V', V(\pi), V'(\pi')) = \{T \in L_B(V,V') \mid V(\pi)T \subset V'(\pi')\}$, which is a B-space modeled on $L_B(V,V')$ whose dimension is given by $H_i * H_j = \cap_{M \in H_i} H'_j(M)$, with $H'_j(M) = \{L \in F(J) \mid L \cap M \neq \varnothing \text{ or } L \cup M \subset H'_j\}$, and where H_i, H'_j define the dimensions of $V(\pi)$ and $V'(\pi')$, respectively. Then we consider $L_B(V,V';F,F') = L_B(V,V';V(\pi),V'(\pi')) \cap L_B(V,V';V(\sigma),V'(\sigma'))$ and the linear continuous map $q: L_B(V,V';F,F') \to L(F,F')$, defined by $[v](q(\overline{T})) = [v\overline{T}] \in F'$ for $[v] \in F$. Then, it is easy to see that q is well defined on $L_B(V,V';F,F')$ and that $\operatorname{Im} q = L_B(F,F')$ and $\ker q = L_B(V,V';V(\pi),V'(\sigma'))$. So q factorizes through the linear continuous isomorphism $\hat{q}: \hat{L}_B(F,F') \to \hat{L}_B(F,F')$, where $\hat{L}_B(F,F') = L_B(V,V';F,F')/\ker q$ is a B-space modeled in $L_B(V,V')$. Then $\hat{q}$ identifies canonically $L_B(F,F')$ with the B-space $\hat{L}_B(F,F')$. We observe that the map $\hat{q}$ preserves the composition. Moreover, $\hat{L}_B(F,F')$ is Z_2-graded as B-space and $\hat{q}$ is even.

We study now L_B-isomorphisms. Let $q(\overline{T}) = T \in L_B(F,F')$ be a linear isomorphism such that $\overline{T}$ is a B-isomorphism. Then it follows that $\overline{T}^{-1} \in L_B(V',V;F',F)$ and $T^{-1} = q(\overline{T}^{-1}) \in L_B(F',F)$.

Definition 4.8: $T \in L_B(F,F')$ is a nongraded L_B-isomorphism iff it is a linear isomorphism and there exists a L_B-isomorphism $\overline{T} \in L_B(V,V')$ such that $q(\overline{T}) = T$. If T is even we call it simply an L_B-isomorphism. We denote the set L_B-isomorphisms by $GL_B(F,F')$.

Theorem 4.9 (dimension theorem): Let $F = V(\pi,\sigma)$, $F' = V'(\pi',\sigma')$ be B-spaces and $T \in L_B(F,F')$ a nongraded L_B-isomorphism. Then $\pi = \pi'$ and $\sigma = \sigma'$ up to a reordering of the basis of V'.

Proof: By Theorem 4.5 $k = m + n = m' + n' = k'$. We prove now the theorem by induction over k. For $k = 1$, since $\overline{T} = (a)$ is invertible $r(a) \neq 0$, so that $\varnothing \in H * H'$, $\varnothing \in K * K'$. Therefore, $M \in H'$ $\forall$ $M \in H$ and $L \in K'$ $\forall$ $L \in K$, i.e., $H \subset H'$ and $K \subset K'$. Conversely as $r(a^{-1}) \neq 0$, we conclude $H' \subset H$ and $K' \subset K$. So $H = H'$ and $K = K'$, which completes the proof for $k = 1$. Now we proceed by induction. We assume that the theorem is true for $k = 1$ and prove it for k. Taking the matrix representations $\overline{T} = (a^i_j)$ and $T^{-1} = (b^i_j)$ we have $\Sigma^k_{j=1} a^k_j b^j_k = 1$, so we can choose j such that $r(a^k_j) \neq 0$ and $r(b^k_j) \neq 0$. Now reordering the basis of V' we take $j = n$. Next, from $r(a^k_n) \neq 0$ and $r(b^k_n) \neq 0$, we have $H_n = H'_n$ and $K_n = K'_n$ as before. Now we take out the k th coordinate obtaining two reduced B-spaces $F_0 = V_0(\pi_0,\sigma_0)$ and $F'_0 = V'_0(\pi'_0,\sigma'_0)$, where π_0, π'_0, σ_0, σ'_0 coincide with the first k-1 dimensions of π, π', σ, σ', respectively. Then $\overline{U} = (a^i_j)^{k-1}_{i,j=1}$ is a L_B-isomorphism from V_0 onto V'_0 because $r(b^k_k) \neq 0$ and $\overline{U} \in L_B(V_0,V'_0;F_0,F'_0)$. So there exists $U = q(\overline{U})$ which is a nongraded L_B-isomorphism from F_0 onto F'_0. Therefore, by the induction hypothesis $\pi_0 = \pi'_0$, and $\sigma_0 = \sigma'_0$ up to a recording of the basis of V'_0, and the theorem follows. Q.E.D.

In the case of T being even, the reordering of the basis of V' proceeds independently for the even and odd parts of the basis.

As $L_B(F)$ is a B-space we can define on it many different bodies. However, we will need the concept of body operator, as in Lemma 4.3, rather than that of the body itself, for the transition maps of fiber bundles. For a given body c of F we consider the subspace c-$L_B(F) = \{T \in L_B(F) \mid c(uT) = 0 \text{ if } c(u) = 0\}$. Then for each $T \in c$-$L_B(F)$ there exists a unique $T_c \in L(F_c)$ making commutative the following diagram:

$$\begin{tikzcd} F \arrow[r, "T"] \arrow[d, "c"'] & F' \arrow[d, "c'"] \\ F_c \arrow[r, "T_c"] & F'_c \end{tikzcd} \tag{11}$$

We call T_c the c-body operator of T and T is c-bodied. Of course, there exists always a body on $L_B(F)$ such that its restriction to c-$L_B(F)$ produces T_c from T. We observe that if c is a maximal body then c-$L_B(F) = L_B(F)$. Similar considerations can be taken for $L_B(F,F')$.

Given a B-space F modeled on V not all the bases of V make sense for F, but only those obtained from the canonical one $e_1,\ldots,e_{m+n}$ through an element of $GL_B(F)$. So $GL_B(F)$ can be considered as the set of reference frames on F. We remark that when F and/or F' are not free B-modules the space of B-linear maps from F into F' (in the usual sense) is bigger in general than the space $L_B(F,F')$.

To end up with the discussion about L_B-operators between B-spaces, we define the dual B-space to a B-space F as $F^* = \hat{L}_B(F,B)$. It is easy to see that F is not reflexive in general ($F \subset F^{**}$). However, F^* is always reflexive $F^{***} = F^*$.

C. L_B-operators between superspaces

Superspaces are a particular case of B-spaces and all the results of Sec. 4B hold for superspaces. However, we shall give here some specific features which will be of interest later on.

Given two superspaces S, S' modeled, respectively, on V and V' we have that $\hat{L}_B(S,S') \subset \hat{L}_B(S,V')$. Indeed $L_B(V,V';S,S') \subset L_B(V,V';S,V') = L_B(V,V')$ and

$$L_B(S,S') = \frac{L_B(V,V';S,S')}{L_B(V,V';S,\{0\})},$$

$$L_B(S,V') = \frac{L_B(V,V')}{L_B(V,V';S,\{0\})}.$$

The dimensions of $L_B(V,V';S\{0\})$ are given by $P_i = P_{H_i}$ with $H_i = \{L \in F(J) | L \supset K_i\}$ iff K_i is finite and $i \geqslant m+1$, and $p_i = 0$ otherwise. In particular $L_B(V,V';S\{0\}) = \{0\}$ iff S is free. As a consequence if $\overline{T}_1 = (a_i^j)$ and $\overline{T}_2 = (b_i^j)$ are two extensions of $T \in L_B(S,V')$, i.e., $q(\overline{T}_1) = q(\overline{T}_2) = T$, then $a_i^j - b_i^j \in \dot{\beta}_{K_i}$, where $\dot{\beta}_{K_i} = \{\text{multiples of } \beta_{K_i}\}$ if K_i is finite and $a_i^j = b_i^j$ otherwise.

Concerning the dual space S^* of a superspace, we have that S^* is a B-module (free iff S is free) and $S^{**} = \overline{S} = \{\Sigma_{i=1}^{m+n} a^i e_i \in V | a^i \beta_{K_i} = 0\}$ is also a B-module. In fact, $S^* = V^*$ and $\overline{S} = V$ when S is free. On the other hand, if $q(\overline{T}_1) = q(\overline{T}_2) = T \in L_B(S,V')$, it holds that $\overline{T}_1|_{\overline{S}} = \overline{T}_2|_{\overline{S}}$. In fact, $\overline{S}$ can be characterized as the maximal subset of V such that all extensions of operators in $L_B(S,V')$ coincide.

For each element $T \in L_B(S,S')$ we have different extensions $\overline{T} \in L_B(V,V';S,S')$ and different matrix expressions of T. The matrix elements of $\overline{T} = (a_i^j)$ are then of the form $a_i^j = b_i^j + c_i^j$ with $b_i^j \in B_{K_j'}^{r_i + r_j}$, $c_j^i \in \dot{\beta}_i$ if K_i is finite for $r_i = 1$, and $c_i^j = 0$ otherwise. Moreover, if $K_i \not\subset K_j'$, $b_i^j \in \dot{\beta}_{K_i \cap K_j}$ if K_i is finite and card $(K_i - K_j') = 1$ for $r_i = 0$, and $b_i^j = 0$ otherwise.

In what follows we shall exclude the pathological case card $(K_i - K_j') = 1$ with K_i finite and $r_i = 0$.

We observe that if S is free, $L_B(S,S')$ is a superspace modeled on $L_B(V,V')$. Finally, we remember that a superspace $S^{m,n}$ has a canonical body map $r: S^{m,n} \to \mathbb{R}^m$ which is maximal. Therefore r', $r - \hat{L}_B(S,S') = \hat{L}_B(S,S')$ and each $T \in L_B(S)$ has a $T_{r,r'} \in L(\mathbb{R}^m)$ associated with it such that

$$\begin{tikzcd} S^{m,n} \arrow[r, "T"] \arrow[d, "r"'] & S^{m',n'} \arrow[d, "r'"] \\ R^m \arrow[r, "T_{r,r'}"] & R^{m'} \end{tikzcd} \tag{12}$$

On the other hand, we have a maximal body map from $\hat{L}_B(S,S')$ onto $L(R^{m+n}, R^{m'+n'})$ given by $R(T) = \overline{T}_R$, where $q(\overline{T}) = T$. $R(T)$ is well defined since $\overline{T}_{1_R} = \overline{T}_{2_R}$ if $q(\overline{T}_1) = q(\overline{T}_2)$.

5. SUPERTENSOR ALGEBRA

In this section we shall study multi-L_B maps and we shall construct supertensor algebra over a superspace.

Definition 5.1: Let V,V' be B-modules. A B-multilinear operator from V^k into V' is a multilinear map $T \in L^k(V,V')$ such that $(u_1,\ldots,au_i,\ldots,u_k)T = (-1)^{r\Sigma_{j=1}^{i-1} r_j} a(u_1,\ldots,u_i,\ldots,u_k)T$ for each homogeneous $a \in B^r$ and $u_i \in V_i^{r_i}$.

The set of B-multilinear operators will be denoted by $L_B^k(V,\overset{k}{\ldots},V,V')$ and it is a B-module. The Z_2 grading of $L_B^k(V,\overset{k}{\ldots},V,V')$ is given by the rule $T \in L_B^k(V,\overset{k}{\ldots},V,V')^r$ if

$$\left(V \times \overset{k}{\cdots} \times V\right)^s T \subset V'^{r+s},$$

where

$$(V \times \cdots \times V)^s = \bigcup_{s_1 + \cdots + s_k = s} V^{s_1} \times \overset{k}{\cdots} \times V^{s_k}.$$

We remark that the above rule does not define a grading for $V \times \overset{k}{\cdots} \times V$. A tensor representation for T is $(V_1,\ldots,V_k)T = \Sigma \epsilon_{i_1\ldots i_k} a_1^{i_1} \cdots a_k^{i_k} b^j_{i_1\ldots i_k} e_j'$, where the sign $\epsilon_{i_1\ldots i_k} = (-)\Sigma_{l=2}^k \Sigma_{h=1}^{l-1} |a_l^{i_l}| r_{i_k}$, and $v_i = \Sigma a_i^j e_j \in V$, $b^j_{i_1\ldots i_k} \in B$. $T \in L^k(V,V')$ is graded (skew-) symmetric if for

$v_i \in V^{r_i}$, $1 \leqslant i \leqslant k$,

$$\begin{aligned}(v_1,\ldots,v_{i+1},v_i,\ldots,v_k)T &= (-1)^{r_i r_{i+1} + r}(v_1,\ldots,v_i,v_{i+1},\ldots,v_k)T \\ &= (-1)^{r_i r_{i+1} + r}(v_1,\ldots,v_i,v_{i+1},\ldots,v_k)T\end{aligned}$$

with $r = 0$ $(r = 1)$. In particular, if T is graded skew-symmetric then $(e_{i_1},\ldots,e_{i_k})T = 0$ provided that some odd (even) e_i appears twice.

We define multi-L_B operators in the same way as L_B-operators.

Definition 5.2: Let $F = V(\pi,\sigma)$ and $F' = V(\pi',\sigma')$ be B-spaces; then $T \in L^k(F \times F \times \cdots \times F, F')$ is multi-L_B (or $T \in L_b^k(F \times F \times \cdots \times F, F')$ iff there exist T' and $\overline{T}$ such that

$$\begin{tikzcd} V \times V \times \cdots \times V \arrow[r, "\overline{T}"] & V' \\ V(\pi) \times \cdots \times V(\pi) \arrow[u, "i"] \arrow[d, "p \times \cdots \times p"'] & V'(\pi') \arrow[u, "i"'] \arrow[d, "p'"] \\ F \times \cdots \times F \arrow[r] & F' \end{tikzcd} \tag{13}$$

$L_B^k(F \times \cdots \times F, F')$ can be canonically identified with a B-space $\hat{L}_B^k(F \times \cdots \times F, F')$ in a similar way as for L_B-maps.

Now we restrict the discussion to superspaces. In particular, we are interested in the form of the tensor representations of multi-L_B operators in order to deal with partial derivatives. When the superspace S is not free there exist many extensions $\overline{T} \in L_B^k(V,V')$ for a given $T \in L_B^k(S,V')$. For two extensions $\overline{T}_1 = (a^j_{i_1,\ldots,i_k})$, $\overline{T}_2 = (b^j_{i_1,\ldots,i_k})$ we have that if $K_{i_{j_1}},\ldots,K_{i_{j_s}}$ are finite, and

$$r_{i_{j_1}} = r_{i_{j_2}} = \cdots = r_{i_{j_s}} = 1, \quad a^j_{i_1,\dots,i_k} - b^{\,j}_{i_1,\dots,i_k} \in E,$$

where E is the annihilator of $A = B^1_{K_{i_{j_1}}} B^1_{K_{i_{j_2}}} \cdots B^1_{K_{i_{j_s}}}$ (see Lemma 2.6) and $a^j_{i_1,\dots,i_k} = b^{\,j}_{i_1,\dots,i_k}$ otherwise. We observe that a (skew-) symmetric operator $T \in L^k_B(S,V')$ has always a graded (skew-) symmetric extension $\overline{T} \in L^k_B(V,V')$.

We discuss now tensor algebra and exterior algebra over a superspace. First we consider the case of a free superspace. Then $S^* = V^*$ and $\overline{S} = V$. So it reduces to the case of B-modules. Given V_1 and V_2 B-modules the tensor product $V_1 \otimes_B V_2$ is well defined because V_1 and V_2 can be simultaneously considered as right and left modules. We can endow $V_1 \otimes_B V_2$ with the structure of a B-module because B is graded commutative,

$$a(v_1 \otimes_B v_2) = av_1 \otimes_B v_2 = (-1)^{r' r_1} v_1 \otimes_B av_2,$$

where $a \in B^{r'}, v_1 \in V_1^{r_1}, v_2 \otimes V_2$.

The grading of $V_1 \otimes_B V_2$ is given by

$$(V_1 \otimes_B V_2) = \underset{r_1+r_2}{\oplus} = r(V_1^{r_1} \oplus_B V_2^{r_2}). \tag{14}$$

To simplify the notation the suffix B will be dropped from now on. The usual properties of commutativity and associativity of the tensor product hold here, and the same is true for the universal property (with respect to B-multilinearity).

Next we consider the contravariant tensor algebra $T(V)$ of a B-module V. As usual $T(V) = \oplus_{k=0}^{\infty} T^k(V)$, where $T^0(V) = B$, and $T^k(V) = \overset{k}{V \otimes \cdots \otimes V}$ because of the grading we need some modifications in the definition of symmetric and skew-symmetric tensors. Let S_k be the permutation group and S, A two homomorphisms $S,A: S_k \to GL_B(T^k(V))$,

$$S(\tau)(X_1 \otimes \cdots \otimes X_k) = (-1)^{r_i r_{i+1}} X_1 \otimes \cdots \otimes X_{i+1} \otimes X_i \otimes \cdots \otimes X_k, \tag{15a}$$

$$A(\tau)(X_1 \otimes \cdots \otimes X_k) = (-1)^{r_i r_{i+1}+1} X_1 \otimes \cdots \otimes X_{i+1} \otimes X_i \otimes \cdots \otimes X_k, \tag{15b}$$

where τ is the transposition $(i,i+1)$, $X_i \in V^{r_i}$, $X_{i+1} \in V^{r_{i+1}}$. Then the symmetrization S and the alternation A are defined on $T(V)$ as

$$S(R) = \frac{1}{k!} \sum_{\sigma \in S_k} S(\sigma)(R), \qquad R \in T^k(V). \qquad A(R) = \frac{1}{k!} \sum_{\sigma \in S_k} A(\sigma)(R), \tag{16}$$

S and A are even idempotent linear operators on $T(V)$. $S(V) = \operatorname{Im} S$ is the space of graded-symmetric tensors and $A(V) = \operatorname{Im} A$ is the space of graded skew-symmetric tensors.

The exterior product in $A(V)$ can be defined by

$$R \Lambda R' = \frac{(k+k')!}{k!\,k'!} A(R \otimes R'), \quad R \in A^k(V), \quad R' \in A^{k'}(V). \tag{17}$$

The isomorphisms $S(V) \approx S(V^0) \otimes A(V^1)$ and $A(V) \approx A(V^0) \otimes S(V^1)$ can be proved as consequences of the universal factorization property of the tensor product.

Given a basis $\{e_i\}_{i=1}^{m+n}$ for V with $e_i \in V^0$ for $i \leqslant m$ and $e_i \in V^1$ for $i > m$ we define for each set

$$(k_1,\dots,k_n) \in \mathrm{N}^n, e(k_1,\dots,k_n) = \overset{k_1}{e_{m+1} \Lambda \cdots \Lambda e_{m+1}} \Lambda \overset{k_2}{e_{m+2} \cdots \Lambda e_{m+2}} \Lambda \cdots \Lambda \overset{k_n}{e_{m+n} \cdots \Lambda e_{m+n}}, \quad e(0,\dots,0) = 1.$$

Note that the defined elements are not zero because we are in $S(V^1)$. Then a basis for $A(V)$ is given by $\{e_{i_1} \wedge e_{i_r} \wedge e_{(k_1,\dots,k_n)}\}$ with $1 \leqslant i_1 < \cdots i_r \leqslant m$ and $(k_1,\dots,k_n) \in \mathrm{N}^n$.

The covariant tensors are defined obviously by $T_r(V) = T^r(V^*)$ and the mixed tensors by $T^r_s(V) = T^r(V) \otimes T_s(V)$. Then, we can define the mixed tensor algebra $\mathcal{T}(V) = \oplus_{r,s=0}^{\infty} T^r_s(V)$. We observe that the B-module $T^r_s(V)$ is isomorphic to a B-module of B-multilinear maps $T^r_s(V) \approx L^r_B(V; T^s(V))$.

Concerning body maps for tensors, we choose the body of $V(=\overline{S})$ as $r = R \cdot P^0$. For each k a canonical body map r from $T^k(V)$ on $T^k(V_r)$ is defined by $r(V_1 \otimes \cdots \otimes V_k) = r(V_1) \otimes \cdots \otimes r(V_k)$. In a similar way body maps for $A(V)$ and $\mathcal{T}(V)$ are defined. We observe that r is an algebra homomorphism from $\mathcal{T}(V)$ onto $\mathcal{T}(V_r)$.

For S not free we have the following situation. Since $\overline{S}$ is a (not-free) B-module $\overline{S} \otimes_B \overline{S}$ is determined by B-linearity and has the universal property

$$\begin{tikzcd} \overline{S} \times \overline{S} \arrow[r] \arrow[dr, "T"'] & \overline{S} \times_B \overline{S} \arrow[d, "U"] \\ & V' \end{tikzcd} \tag{18}$$

i.e., for T B-bilinear there exist U B-linear making commutative the above diagram. However, we are interested in the L_B-maps rather than in B-linear maps. As the last ones form, in general, a wider class than the former ones, we have that the space $\overline{S} \otimes_{L_B} \overline{S}$ satisfying the universal property for L_B-maps is bigger than $\overline{S} \otimes_B \overline{S}$. In fact $\overline{S} \otimes_{L_B} \overline{S} \approx \hat{L}^2_B(S^*, S^*, B) \not\approx \overline{S} \otimes_B \overline{S}$. We remark that $\overline{S} \otimes_{L_B} \overline{S}$ has also the universal property for L^2_B operators defined on $S \times S$. So $S \otimes_{L_B} S = \overline{S} \otimes_{L_B} \overline{S}$ and we shall denote it shortly by $S \otimes S$ or $\overline{S} \otimes \overline{S}$. Notice that first-order tensors are then $\overline{S}$ and not S. Finally, these considerations can be straightforwardly extended to higher tensor products of S and S^*.

6. CONCLUSIONS

In this paper we have introduced the concept of generalized superspace. The starting point for it is the structure of Grassmann–Banach algebra which includes, as main examples, the exterior algebra of a finite-dimensional vector space and infinite-dimensional Grassmann algebras of Rogers.[4] Our definition contrasts with Banach–Grassmann algebras of Jadczyk and Pilch[5] which do not include finite-dimensional cases.

The generalization of superspace consists in considering the possibility of different Grassmann algebras for each coordinate. This allows for a richer field of applications in physics.[6] The "dimension" of a generalized superspace is defined as a sequence of sets such that each one determines the generators spanning the Grassmann algebra to which each coordinate belongs.

B-linear maps over graded B-modules and L_B-maps over superspaces have been discussed. It happens that the set of L_B-maps between superspaces is not a superspace in general. To solve this problem we have introduced the category of B-spaces which is wider than that of superspaces and where morphisms are L_B-maps. Also multi-L_B maps and tensor products of superspaces are discussed in the framework of B-space category.

In the next paper of this series, we shall study analysis on generalized superspaces.

[1]A Salam and J. Strathdee, Fortschr. Phys. **26**, 57 (1978).

[2]F. A. Berezin and D. A. Leĭtes, Sov. Math. Dokl. **16**, 1218 (1975); D. A. Leĭtes, Usp. Mat. Nauk. **35**, 1 (1980) [Russian Math. Surveys **35**, 1 (1980)].

[3]B. Kostant, in *Differential Geometrical Methods in Mathematical Physics, Bonn 1975* (Springer-Verlag, Berlin, 1977).

[4]A specific realization of the algebra $A(X)$ as the space of cross sections of a bundle with base X and fiber Λ^q has been given in J. Dell and L. Smolin, Comm. Math. Phys. **66**, 197 (1979).

[5]A. Rogers, J. Math. Phys. **21**, 1352 (1980).

[6]A similar approach with trivial topology in the anticommuting sector is due to B. de Witt (unpublished).

[7]A. Hadczyk and K. Pilch, Comm. Math. Phys. **78**, 373 (1981).

[8]M. M. Day, *Normed Linear Spaces* (Springer-Verlag, Berlin, 1973).

Generalized supermanifolds. II. Analysis on superspaces

J. Hoyos
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Matemáticas, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

M. Quirós
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain

J. Ramírez Mittelbrunn and F. J. de Urríes
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Física, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

(Received 26 November 1982; accepted for publication 30 September 1983)

The analysis on generalized superspaces is studied. G-differentiability of functions and indetermination of partial G-derivatives are treated in detail. A canonical expansion for supersmooth functions and a particular choice of partial G-derivatives are given, and their properties studied. Existence of (nonanalytic) G^{∞}-functions is proved.

PACS numbers: 02.40. + m, 11.30.Pb, 02.30. + g

1. INTRODUCTION

In this paper we discuss the analysis on generalized superspaces following the program outlined in Ref. 1. Analysis on homogeneous superspaces has been recently considered by Rogers[2] and Jadczyk and Pilch.[3] However, some important questions are left out in these papers, while others are not solved in a fully satisfactory way. In particular, the domain of validity of the extension theorem for G^{∞}-functions has not been well determined in Ref. 2. This problem has been solved by Jadczyk and Pilch, through the concept of G-connectedness. However, their approach only applies to infinite-dimensional Grassmann algebras. On the other hand, the superfield expansion in terms of ordinary C^{∞}-functions is not satisfactorily treated in Ref. 3 due to the difficulties inherent to work with infinite-dimensional Grassmann algebras. To be more precise, the subclass of C^{∞} functions which can appear in a superfield expansion is not clear. We prove that this class is large enough to guarantee the existence of partitions of unity in ρ-supermanifolds (Paper III). For finite-dimensional Grassmann algebras the superfield expansion has been given by Rogers, but it lacks uniqueness. All these problems, together with the new aspects introduced by the use of generalized superspaces, have been solved in this paper. In particular, a rule to fix a canonical superfield expansion is given in the case of nonfree superspaces.

In Sec. 2 the concept of G-differentiability is introduced. Spaces of G^{k}-functions are studied and special attention is devoted to the inverse and implicit mapping theorem. The extension theorem, and some of its consequences, is developed in Sec. 3. A particularly interesting result of this section is that the analyticity is preserved by the extension. The Z-operator mapping C^{∞} into G^{∞}-functions is defined and G^{∞}-functions are expanded in terms of families of C^{∞}-functions. While this expansion is certainly not unique for nonfree superspaces, we have fixed a particular canonical expansion. Moreover, from the canonical expansion, partial G-derivatives can be uniquely determined. For the case of even coordinates taking values in infinite-dimensional Grassmann subalgebras, a Z-operator is defined, mapping C^{∞}-functions with nonvanishing convergence radius for its Taylor series at each point, into G^{∞}-functions. In particular C^{ω}-functions are mapped onto G^{ω}-functions. However, the existence of a superfield expansion for G^{∞}-functions which are not G^{ω} allows one to show the existence of "bell" functions on superspaces, which in turn is crucial for having partitions of unity on ρ-supermanifolds.

2. G-DIFFERENTIABILITY

In this section we shall proceed to develop further the concept of G-differentiability, first introduced by Rogers,[2] extending it to functions defined on generalized superspaces.

We first recall that $f: U \subset E \to F$, E,F Banach spaces and U an open set, is said to be C^{k} if it is k times continuously differentiable. Then $D^{k}f(x) \in L^{k}(E, F)$ denotes the k th differential of f and is a symmetric k-multilinear operator.[4]

Definition 2.1: Let $F = V(\pi,\sigma)$ and $F' = V'(\pi',\sigma')$ be B-spaces and $U \subset F$ an open set. (a) f is G^{0} if it is C^{0}. (b) f is G^{1} iff it is C^{1} and there exists a continuous map $\overline{Df}: U \to L_{B}(V,V';F,F')$ such that $Df(x) = q(\overline{Df(x)})$ $\forall x \in U$. (c) f is G^{k} if it is G^{1} and $\overline{Df}$ is G^{k-1}. (d) f is G^{∞} if it is G^{k} $\forall k$. (e) f is G^{ω} (G-analytic) if it is analytic (C^{ω}) and for each $\mathbf{x}_0 \in U$, there exists $U_0 \subset U$ open neighborhood of $\mathbf{x}_0$ such that for all $\mathbf{x} \in U_0$, $\mathbf{x} = (x^{i})_{i=1}^{m+n}$

$$f(\mathbf{x}) = \sum_{i_1,\dots,i_{m+n}=0}^{\infty} (x^{1} - x_0^{1})^{i_1} \cdots (x^{m+n} - x_0^{m+n})^{i_{m+n}} \times v'_{i_1,\dots,i_{m+n}},$$

where $v'_{i_1,\dots,i_{m+n}} \in V'$, and the series converges uniformly on $\overline{B_r(x_0)} \subset U$. As immediate consequences of Definition 2.1, if f is G^{k}-differentiable, then Df is G^{k-1}-differentiable and the composition of two G^{k}-differentiable functions, f and g, is G^{k}-differentiable.

A particularly interesting case is $F' = B$. A function $f: U \to V'$ defines the family $f^{j}: U \to B$, $1 \leqslant j \leqslant m' + n'$ by

$f(\mathbf{x}) = \Sigma_{j=1}^{m'+n'} f^j(\mathbf{x})e'_j$, $\mathbf{x} \in U$. Then f is G^k-differentiable iff f^j, $1 \leqslant j \leqslant m' + n'$, is G^k-differentiable.

Let us now see how the partial G-derivatives of a function appear. For $f: U \to F'$, $D^k f(\mathbf{x})$, $\mathbf{x} \in U$, is a multi-L_B operator over S. A tensor representation of $D^k f(\mathbf{x})$ can be expressed as

$$(h_1,...,h_k) D^k f(\mathbf{x}) = \sum_{j=1}^{m'+n'} \left(\sum_{1<i_e<m+n} h_1^{i_1} \cdots h_k^{i_k} G_{i_k} \cdots G_{i_1} f^j(\mathbf{x}) \right) e'_j \tag{1}$$

which defines the partial G-derivatives $G_{i_k} ... G_{i_1} f^j(\mathbf{x})$. However, for $F = S$ nonfree (see I) superspaces, they are not completely determined for odd coordinates. On the other hand, if f is G^l the existence of partial derivatives $G_{i_k} \cdots G_{i_1} f$ that are G^{l-k} functions is guaranteed by the iterative form of Definition 2.1(a). For a G-differentiable function $f: u \to S'$, if $K_i \not\subset K'_j$, then f^j does not depend on the coordinate x^i. In fact for a G-differentiable function $f: U \to F'$, the dimensions of $L_B(S, F')$ might enforce $G f^j(x) = 0$, $x \in U$.

Next, let $G^k(U^j, F') \subset C^k(C, F')$ be the set of G^k-functions from U into F', $k = 0,1,..., \infty, \omega$. In the particular case $F' = B$ we simply denote this set by $G^k(U)$.

Proposition 2.2:(i) $G^k(U)$ is a Z_2-graded commutative algebra and a Z_2-graded B-module. (ii) $G^k(U;V')$ is a Z_2-graded $G^k(U)$-module. In particular it is also a Z_2-graded B-module, and a vector subspace of $C^k(U;V')$.

The grading of $G^k(U;V')$ is given by $f \in G^k(U;V')^r$ iff $\mathrm{Im}(f) \subset V'^r$. Moreover, using the canonical projections $p^r: V' \to V'^r$ and the property $p^r \cdot f \in G^k(U;V')^r$, the decomposition $f = p^0 \cdot f + p^1 \cdot f$ holds.

The product of functions is defined as usual, $(fg)(\mathbf{x}) = f(\mathbf{x}) g(\mathbf{x})$, where $f \in G^k(U)$ and $g \in G^k(U)$ [yielding the structure of algebra for $G^k(U)$], *or* $g \in G^k(U;V')$ [yielding the structure of $G^k(U)$-module for $G^k(U;V')$]. The commutativity of $G^k(U)$ comes from that of B.

We end these comments with some remarks concerning the grading of partial G-derivatives of G^k-functions. Let $r_i = 0$ if $1 \leqslant i \leqslant m$, $r_i = 1$ if $m + 1 \leqslant i \leqslant m + n$, $f, g \in G^k(U)$. Then: (i) If $f \in G^k(U)^r$, then $G_i f \in G^{k-1}(U)^{r+r_i}$; (ii) $G_i(f+g) = G_i f + G_i g$; (iii) if $f \in G^k(U)^r$, then $h_i G_i(fg) = h_i((G_i f)g + (-1)^{rr_i} f(G_i g))$; (iv) if $a \in B^r$, $h_i G_i(af) = h_i((-1)^{rr_i} a(G_i f))$.

Let $f \in G^k(U, F')$, $U' \subset F'$ be an open set with $f(U) \subset U'$ and $g \in G^k(U', F'')$. Then the composition $g \cdot f \in G^k(U, F'')$ and the total differential reads (chain rule)

$$D(g \cdot f)(\mathbf{x}) = Df(\mathbf{x}) \cdot Dg(f(\mathbf{x})), \tag{2}$$

where total differential operators are right L_B-operators as in (1). This leads to the chain rule for partial G-derivatives,

$$h_i G_i(g \cdot f)(\mathbf{x}) = h_i \sum_{j=1}^{m'+n'} G_i f^j(\mathbf{x}) G_j g(f(\mathbf{x})). \tag{3}$$

For a $F = S$ nonfree superspace the indetermination of partial G-derivatives affects Eq. (3) in the following way: the possible indetermination of $G_j g$ has no effect; if $m + 1 \leqslant i \leqslant m + n$ and K_i is finite, the right-hand side of (3) gives a possible determination of $G_i(g \cdot f)$; in other cases formula (3) is exact without the h_i's.

Next, we shall generalize the inverse mapping theorem and its consequences. The approach and results contained in Ref. 5 are borrowed. To fill the gap from C^∞- to G^∞-differentiability we need the following lemma.

We use the notation $L_B(F)$ insted of $L_B(F;F)$ and denote by $GL_B(F)$ the isomorphisms in $L_B(F)$.

Lemma 2.3:(i) Let $T \in L^k(F;F')$. Then: (i) T is $G^1 \Rightarrow T \in L_B^k(F;F') \Rightarrow T$ is G^∞. (ii) Let $\Psi: L_B(F) \times L_B(F) \to L_B(L_B(F))$ be defined by $C(A, B)\Psi = -A \cdot C \cdot B$. Then Ψ is L_B-bilinear, hence G^∞. (iii) $GL_B(F)$ is an open set of $L_B(F)$. (iv) Let $I: GL_B(F) \to GL_B(F)$ be the inversion map $i(T) = T^{-1}$. Then i is G^∞ and $Di(T) = (T^{-1}, T^{-1})\Psi$.

Proof: (i) Comes from the definition of G-differentiability and $DT(x) = T$ for linear mappings. Ψ is L_B-bilinear since there exists an obvious extension to $L_B(V)$ given by the very definition of Ψ. Property (iii) comes from the fact that $GL(V)$ is an open set of $L(V)$. Finally, it is known that i is C^∞ and $Di(T) = (T^{-1}, T^{-1})\Psi$. So from (ii) and proceeding by induction, property (iv) follows. Q.E.D.

We remark that Lemma 2.3 is still true even when S is a nonfree superspace. In this case $L_B(S)$ is not a superspace but a B-space. However, with the natural extension of the notion of G^∞-function to B-spaces, $i(T) = T^{-1}$ is yet a G^∞-function. So the proof of Theorem 2.4 is valid also for nonfree superspaces.

Theorem 2.4 (Inverse mapping theorem): Let $f: U \to F'$ be G^k, $k > 1$, and $\mathbf{x}_0 \in U$ such that $Df(\mathbf{x})$ is a L_B-isomorphism. Then, there exists a neighborhood of $\mathbf{x}_0$, $U_0 \subset U$ such that $f(U_0)$ is an open set of S' and $f: U_0 \to f(U_0)$ is a G^k-diffeomorphism (i.e., one-to-one and f^{-1} is G^k).

Proof: From the ordinary inverse mapping theorem we only need to prove that f^{-1} is G^k. We can identify F' with F by means of the L_B-isomorphism $Df(\mathbf{x}_0)$. We shall proceed by induction. First, f^{-1} is G^1 since it is C^1 and for all $\mathbf{y} \in f(U_0)$, $Df^{-1}(\mathbf{y}) = (Df(f^{-1}(\mathbf{y})))^{-1}$ is L_B-linear (it is the inverse of a L_B-linear map). Now, if f^{-1} is G^l, $l \leqslant k - 1$, writing $Df^{-1} = i \cdot Df \cdot f^{-1}$ we have that Df^{-1} is G^l, since Df is G^{k-1} and i is G^∞ by Lemma 2.3. Hence f^{-1} is G^{l+1}. This concludes the induction procedure. Q.E.D.

Prior to some corollaries of Theorem 2.4 we shall give some definitions, enlarging the concept of closed subspace which splits to the case of superspaces.

Let S and S_1 be superspaces. We say that S_1 splits S if there exists a superspace S_2 and a L_B-isomorphism $T_1 \in L_B(S;S_1 \times S_2)$.

A one-to-one B-linear operator $T \in L_B(S;S')$ is said to split if Im T is closed in S', S splits S', and there exists a L_B-isomorphism $T_1 \in L_B(S';S \times S_2)$ such that $T \cdot T_1 \in L_B(S, S \times O)$ is a L_B-isomorphism.

Corollary 2.5: Let $f: U \to S'$ be a G^k-function, $k \geqslant 1$, $f(\mathbf{x}_0) = 0$ and $\mathbf{x}_0 \in U$. (i) If $Df(\mathbf{x}_0)$ is a one-to-one L_B-linear operator which splits, then there exists $U' \subset S'$, open set, and a function $g: U' \to S \times S_2$ which is a G^k-diffeomorphism onto $g(U') \subset S \times S_2$, open set, such that there exists a neighborhood U_0 of $\mathbf{x}_0$, $U_0 \subset U$, and $g \cdot f(U_0) \subset S \times O$ is an open set in $S \times O$ and $g \cdot f$ is a G^k-diffeomorphism onto its image. (ii) If $Df(\mathbf{x}_0)$ is onto S' and $S_1 = \ker Df(\mathbf{x}_0)$ splits $S\,(S \cong S_1 \times S_2)$, then there exist open subsets $U_1 \subset S_1$ and $U_2 \subset S_2$ and $\mathbf{x}_0 \in U_0 \subset U$ and $h: U_1 \times U_2 \to U_0$, G^k-diffeomorphism onto U_0, such that $f \cdot h$ is a projection $U_1 \times U_2 \to S'$ (i.e., can be expressed as the composition of the projection $U_1 \times U_2 \to U_2$ followed by a G^k-diffeomorphism from U_2 onto an open subset of S').

Corollary 2.6 (Implicit function theorem): Let S_1, S_2, and S' be superspaces, $U_1 \subset S_1$, $U_2 \subset S_2$ open subsets, and $f: U_1 \times U_2 \to S'$ a G^k-function such that $f(\mathbf{x}_0, \mathbf{y}_0) = 0$, $(\mathbf{x}_0, \mathbf{y}_0) \in U_1 \times U_2$, and $D(f \cdot j_{\mathbf{x}_0})(y_0)$ a L_B-isomorphism, where $j_{\mathbf{x}_0}: U_2 \to U_1 \times U_2$ is the inclusion map $j_{\mathbf{x}_0}(\mathbf{y}) = (\mathbf{x}_0, \mathbf{y})$. Then, there exists an open neighborhood of $\mathbf{x}_0$, $U_0 \subset U_1$ and a unique G^k-map g: $U_0 \to U_2$ such that $g(\mathbf{x}_0) = \mathbf{y}_0$ and $f(\mathbf{x}, g(\mathbf{x})) = 0$ for all $\mathbf{x} \in U_0$.

The proof of analogous corollaries given in Ref. 5 applies here since the inverse mapping theorem holds.

3. EXTENSION OF G^∞-FUNCTIONS

We deal in this section with the extension of G^∞-functions and their natural domain of definition. We shall see that G^∞-functions behave along the soul as analytic functions. The extension theorem was first stated by Rogers,[2] without fixing its domain of validity. Later, Jadczyk and Pilch[3] found the domain in which the theorem holds and generalized it to infinite-dimensional Grassmann algebras.

Definition 3.1: Let S be a superspace and $U \subset S$ a subset. The r-saturation of U is $\widetilde{U} = r^{-1}(r(U))$. U is said to be G-connected (G-convex) if for all $\mathbf{x} \in U$, $\{\tilde{\mathbf{x}}\} \cap U$ is connected (convex).

Clearly, if U is open, $\widetilde{U}$ is also open.

Theorem 3.2: Let S be a superspace, $U \subset S$ a G-connected open set and $f \in G^\infty(U;F')$ where F' is a B-space with body c. Then: (i) If $\mathbf{x},\mathbf{y} \in U$ and $r(\mathbf{x}) = r(\mathbf{y})$, then $cf(\mathbf{x}) = xf(\mathbf{y})$; (ii) there exists a unique extension $\tilde{f} \in G^\infty(\widetilde{U};F')$; (iii) $c(\tilde{f}(\widetilde{U})) = c(f(U))$; (iv) moreover, if $f \in G^\omega(U;F')$, then $\tilde{f} \in G^\omega(\widetilde{U};F')$.

Proof: It is enough to consider the case of $f \in G^\infty(U)$, and $c = r$. The proof of Proposition 5.4 of Ref. 3 applies here to (i), (ii), and (iii) without any modification.

Let us suppose now that f is G^ω. Then f is G^∞ and hence there exists, by (ii), a unique $\tilde{f}$ on $\widetilde{U}$ that is G^∞. We now prove that $\tilde{f}$ is G^ω. Let $\mathbf{x}_0 \in \widetilde{U}$ and $\mathbf{h}$, with $r(\mathbf{h}) = 0$ nilpotent and such that $\mathbf{x}_0 = \mathbf{y}_0 + \mathbf{h}$ with $\mathbf{y}_0 \in U$.

Let $\rho(\mathbf{y}_0) > 0$ be the convergence radius for f in $\mathbf{y}_0$. For $\|\mathbf{x} - \mathbf{x}_0\| \leqslant \rho < \rho(\mathbf{y}_0)$, $\mathbf{x} = \mathbf{y} + \mathbf{h}$ with $\|\mathbf{y} - \mathbf{y}_0\| \leqslant \rho$. Because $\mathbf{h}$ is nilpotent, $f(\mathbf{x}) = \Sigma_{p=0}^{k}\,(1/p!) h^p D^p \tilde{f}(\mathbf{y})$ and $D^q \tilde{f}(\mathbf{x}_0) = \Sigma_{p=0}^{k} (1/p!) h^p D^{p+q} \tilde{f}(\mathbf{y}_0)$. Since f is analytic in $\mathbf{y}_0$, $D^p f(\mathbf{y}) = \Sigma_{q=0}^{\infty} (1/q!)(\mathbf{y} - \mathbf{y}_0)^q D^{p+q} f(\mathbf{y}_0)$. Using that $\tilde{f}$ is an extension of f, and $\mathbf{x} - \mathbf{x}_0 = \mathbf{y} - \mathbf{y}_0$, we have

$$\begin{aligned}
\tilde{f}(\mathbf{x}) &= \sum_{p=0}^{k} \frac{1}{p!} h^p \left(\sum_{q=0}^{\infty} \frac{1}{q!} (\mathbf{y} - \mathbf{y}_0)^q D^{p+q} f(\mathbf{y}_0) \right) \\
&= \sum_{q=0}^{\infty} \frac{1}{q!} (\mathbf{x} - \mathbf{x}_0)^q \left(\sum_{p=1}^{k} \frac{1}{p!} h^p D^{p+q} f(\mathbf{y}_0) \right) \\
&= \sum_{q=0}^{\infty} \frac{1}{q!} (\mathbf{x} - \mathbf{x}_0)^q D^q \tilde{f}(\mathbf{x}_0)
\end{aligned} \tag{4}$$

that proves that $\tilde{f}$ is analytic at $\mathbf{x}_0$ and $\rho(\mathbf{x}_0) \geqslant \rho(\mathbf{y}_0)$. Q.E.D.

Corollary 3.3: Let $U \subset S$ be a G-connected open set and $U_0 \subset U$ an open set. Let f, $g \in G^\infty(U;F')$ such that $f = g$ in U_0. Then $f = g$ in $\widetilde{U}_0 \subset U$.

Proof: It is enough to prove that if $\mathbf{x} \in U_0$, then $f = g$ in $\{\tilde{\mathbf{x}}\} \cap U$. Let $\mathbf{x} \in U_0$ and $B_r(\mathbf{x}) \subset U_0$ an open ball, then $B_r(\mathbf{x})$ is G-connected and $f = g$ in $B_r(\mathbf{x})$; by Theorem 3.2 there exists a unique $h \in G^\infty(\widetilde{B}_r(\mathbf{x});F')$ extension of $f|_{B_r(\mathbf{x})}$. Therefore, by the unicity, $f = g$ in $\widetilde{B}_r(\mathbf{x}) \subset U$. Q.E.D.

Proposition 3.4: Let $U \subset S$ and $U' \subset S'$ be G-connected open sets and $f \in G^\infty(U;S')$, $g \in G^\infty(U';F'')$ with $f(U) \subset U'$. Then $\tilde{f}(\widetilde{U}) \subset \widetilde{U}'$ and $(g \cdot f)^{\tilde{}} = \tilde{g} \cdot \tilde{f}$.

Proof: $\tilde{f}(\widetilde{U}) \subset \widetilde{U}'$ by Theorem 3.2 (iii). Since $(g \cdot f)^{\tilde{}} = \tilde{g} \cdot \tilde{f}$ holds in U, by Corollary 3.3 it holds in $\widetilde{U}$. Q.E.D.

Corollary 3.5: Let $U \subset S$ be a G-connected open set and $f \in G^\infty(U;S')$ be a G^∞-diffeomorphism onto $f(U)$, a G-connected open subset of S'. Then $\tilde{f} \in G^\infty(\widetilde{U};S')$ is a G^∞-diffeomorphism onto $f(U)^{\tilde{}}$.

Proof: It is a straightforward consequence of Proposition 3.4 and the fact that the extension of the identity is the identity map. Q.E.D.

We conclude this section with a particular construction which will be crucial in our definition of (bodied) supermanifold.

Proposition 3.6: Let S, S' be superspaces and $U \subset S$ a G-connected open subset. Let $f \in G^\infty(U;F')$. Then there exists a unique function $f_c \in C^\infty(r(U);F'_c)$ such that $f_c \cdot r = c \cdot f$. This can be expressed by the commutative diagram

$$\begin{tikzcd}
U \arrow[r, "f"] \arrow[d, "r"'] & F' \arrow[d, "c"] \\
f(U) \arrow[r, "f_c"] & F'_c
\end{tikzcd} \tag{5}$$

Proof: Without lost of generality we can assume $U = \widetilde{U}$ and $f = \tilde{f}$. Let $i: r(U) \to U$ the inclusion $i(\mathbf{x}) = (\mathbf{x}, \mathbf{0})$. Then we define $f_c = c \cdot f \cdot i$ which is C^∞ because it is a composition of C^∞-functions and satisfies $f_c \cdot r = c \cdot f$ by Theorem 3.2 (i). To see the unicity, suppose f'_c such that $f'_c \cdot r = c \cdot f$. Then $f'_c \cdot r = f_c \cdot r$ so that $f'_c = f_c$ because r is onto.

It is not true in general that the image of a G-connected open set by a G^∞-diffeomorphism is G-connected. This is because the map induced in the bodies could fail to be injective. The following proposition deals with this problem.

Proposition 3.7: Let U be a G-connected open subset of S and f a G^∞-diffeomorphism from U onto $f(U)$, open in S'. (i) If $V \subset r(U)$ is open and $f_r|_V$ is injective, then $f(r^{-1}(V) \cap U)$ is G-connected. (ii) If $Z \subset f(U)$ is G-connected open, then $f_r|_{rf^{-1}(Z)}$ is injective. (iii) For all $\mathbf{x} \in U$ $\exists\, \alpha > 0$ such that $f(\widetilde{B}_\alpha(\mathbf{x}) \cap U)$ is G-connected.

Proof: (i) Let $\mathbf{x},\mathbf{y} \in r^{-1}(V) \subset U$ such that $rf(\mathbf{x}) = rf(\mathbf{y})$. Then $f_r\, r(\mathbf{x}) = rf(\mathbf{x}) = rf(\mathbf{y}) = f_r\, r(\mathbf{y})$ implies $r(\mathbf{x}) = r(\mathbf{y})$. So there is a continuous path α: $[0,1] \to r^{-1}(\mathbf{x})$ joining $\mathbf{x}$ and $\mathbf{y}$, since U is G-connected. By 3.2 $f \cdot \alpha$ is a continuous path in $r^{-1}(f(\mathbf{x}))$ joining $f(\mathbf{x})$ and $f(\mathbf{y})$. So $f(r^{-1}(V) \cap U)$ is G-connected. (ii) Let $\mathbf{x},\mathbf{y} \in f^{-1}(Z)$ such that $f_r\, r(\mathbf{x}) = f_r\, r(\mathbf{y})$. Then $rf(\mathbf{x}) = rf(\mathbf{y})$. By 3.2 $r(\mathbf{x}) = r(\mathbf{y})$, since Z is G-connected and $f^{-1}|_Z$ is G^∞. So $f_r|_{r(f^{-1}(Z))}$ is injective. (iii) Given $\mathbf{x} \in U$

we take $\beta > 0$ such that $B_\rho(f(\mathbf{x})) \subset f(U)$ and $\alpha > 0$ such that $B_\alpha(\mathbf{x}) \subset f^{-1}(B_\rho(f(\mathbf{x})))$. Then (iii) follows from (i) and (ii). Q.E.D.

Part (iii) of Proposition 3.7 means that Corollary 3.5 only applies locally for G^∞-diffeomorphisms.

4. EXPANSION OF G^∞-FUNCTIONS

Every G^∞-function can be expanded as a polynomial in the odd coordinates whose coefficients are functions defined over the even coordinates. By imposing some additional conditions these coefficients can be determined by functions of real variables. The purpose of this section is to discuss this expansion. Moreover, as a consequence, partial G-derivatives will be determined.

A. The Z-operator

Given a superspace $S^{m,n}$ we are going to construct a linear operator Z applying C^∞-functions defined on $r(S)$ into G^∞-functions defined on S.

Without losing generality we can restrict ourselves to the case of $f \in G^\infty(U)$ and U being r-saturated. First we consider the case of $S^{m,0}$. We define a linear operator which applies $f \to \hat{f} = f \cdot i$ from $G^\infty(U)$ into $C^\infty(r(U), B)$, when i is the inclusion $i{:}r(U) \to U$. Then using the chain rule we obtain $\partial\hat{f}(y)/\partial y^i = G_i f(i(\mathbf{y}))$ that leads to $(\partial^{|\alpha|}\hat{f}/\partial y^{1\alpha_1} \cdots \partial y^{m\alpha_m})(\mathbf{y}) = G^{|\alpha|}_{\alpha_1\cdots\alpha_m} f(i(\mathbf{y}))$, where $|\alpha| = \alpha_1 + \cdots + \alpha_m$ and $G^{|\alpha|}_{\alpha_1\cdots\alpha_m} = G_1^{\alpha_1} \cdots G_m^{\alpha_m}$.

Lemma 4.1: Let $f \in G^\infty(U)$, $\mathbf{x}$, $\mathbf{x}+\mathbf{h} \in U$ with $r(U) = 0$.

(a) If $\cup_{i=1}^m K_i$ is a finite set, there exists a finite integer $\overline{m}$ [depending on $\mathscr{K} = (K_i)$] such that

$$f(\mathbf{x}+\mathbf{h}) = \sum_{i_1,\dots,i_{m+n}}^{\overline{m}} \frac{1}{i_1! \cdots i_{m+n}!} \times (h^1)^{i_1} \dots (h^{m+n})^{i_{m+n}} G^{i_{m+n}}_{m+n} \cdots G^{i_1}_1 f(\mathbf{x}). \tag{6}$$

(b) For general dimension, if f has a nonvanishing convergence radius for its Taylor series at x and h is nilpotent (i.e., h_i nilpotent for $i = 1,\dots,m$), then

$$f(\mathbf{x}+\mathbf{h}) = \sum_{i_1,\dots,i_{m+n}} \frac{1}{i_1! \cdots i_{m+n}!} \times (h^1)^{i_1} \cdots (h^{m+n})^{i_{m+n}} G^{i_{m+n}}_{m+n} \cdots G^{i_1}_1 f(\mathbf{x}). \tag{7}$$

Proof: (a) Since U is r-saturated, the segment joining $\mathbf{x}$ and $\mathbf{x}+\mathbf{h}$ is contained in U. Now if $\mathbf{h}$ is nilpotent, the Taylor series cuts off, and by Taylor's theorem it is equal to $f(\mathbf{x}+\mathbf{h})$. On the other hand, if $\cup_{i=1}^m K_i$ is finite there exists $\overline{m}$ such that $\cup_{i=1}^m F(K_i) \subset F(J_{\overline{m}})$. So $\mathbf{h}^\alpha = 0$ if $|\alpha| > \overline{m}$ and (a) follows. Take now $\rho > 0$ to be the convergence radius of the Taylor series for f at $\mathbf{x}$. Then in the closed ball B_ρ, $= \{\mathbf{h} \mid \|\mathbf{h}\| < \rho' < \rho\}$ the series converges absolutely and uniformly. Therefore it defines a continuous function on $B_{\rho'}$, which coincides with $f(\mathbf{x}+\mathbf{h})$ for $\mathbf{h}$ nilpotent and hence for all $\mathbf{h}$ by Proposition I.2.4.

Next for a given $\mathbf{h}$, by Proposition I.2.5 there exists $K = K(h)$ such that $\|\mathbf{h}^\alpha\| \leqslant K\theta^{|\alpha|}$. So the Taylor series converges absolutely for all $\mathbf{h}$ (nonuniformly). However, by decomposing $\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2$ with $\|h_1\| < \rho'$ and h_2 nilpotent, it follows then that the Taylor expansion coincides with $f(\mathbf{x}+\mathbf{h})$ by a similar argument as in the proof of Theorem 3.2. Q.E.D.

Corollary 4.2: The linear operator $\Lambda f \to \hat{f}$ from $G^\infty(U)$ into $C^\infty(r(U);B)$ is injective.

Proof: We show that f is determined by $\hat{f}$. By (b) of Lemma 4.1 and the preceding remarks,

$$f(i(\mathbf{x})+\mathbf{h}) = \sum_{\alpha_1,\dots,\alpha_m}^{\overline{m}} \frac{(h^1)^1 \cdots (h^m)^m}{\alpha_1! \cdots (\alpha_m)!} G^{|\alpha|}_{\alpha_1\cdots\alpha_m} f(i(\mathbf{x}))$$
$$= \sum_{\alpha_1,\dots,\alpha_m}^{\overline{m}} \frac{(h^1)^{\alpha_1} \cdots (h^m)^{\alpha_m}}{\alpha_1! \cdots \alpha_m!} \frac{\partial^{|\alpha|}\hat{f}}{\partial y^{1\alpha_1} \cdots \partial y^{m\alpha_m}}(\hat{\mathbf{y}}). \tag{8}$$

Q.E.D.

Now since "Λ" is injective we have the inverse linear operator Z from $\operatorname{Im}\Lambda = \widehat{G}^\infty(U) \subset C^\infty(r(U);B)$ onto $G^\infty(U)$. Then it is easy to see that the operator Z commutes with derivatives, i.e.,

$$G_i(Z(g)) = Z\left(\frac{\partial g}{\partial x^i}\right). \tag{9}$$

A characterization of $\widehat{G}^\infty(U)$ is not easy to give in general, but as we show in the next theorem $\widehat{G}^\infty(U)$ is wide enough to quarantee the existence of nonanalytic C^∞-functions which in turn are needed to construct partitions of unity. Prior to stating the next theorem we define $C^\infty_*(r(U);B)$ as the subset of functions $f \in C^\infty(r(U);B)$ such that their radii of convergence for their Taylor series are nonvanishing at each point $\mathbf{x} \in r(U)$.

Theorem 4.3: (a) For $g \in C^\infty_*(r(U);B)$, $g \in \widehat{G}^\infty(U)$ and

$$Z(g)(i(\mathbf{x})+\mathbf{h}) = \sum_{\alpha=0}^{\infty} \frac{1}{\alpha!} \mathbf{h}^\alpha \frac{\partial^{|\alpha|} g}{\partial x^\alpha}(x) \tag{10}$$

for $\mathbf{x} \in r(U)$, $r(\mathbf{h}) = 0$ and $i(\mathbf{x}) + \mathbf{h} \in U$. (b) Moreover, if $\cup K_i$ is finite then $\widehat{G}^\infty(U) = C^\infty(r(U);B)$ and (10) holds.

Proof: (a) As $g \in C^\infty_*(r(U);B)$ we can define the function $f(i(\mathbf{x})+\mathbf{h}) = \Sigma_{\alpha=0}^\infty (1/\alpha!)\mathbf{h}^\alpha(\partial^{|\alpha|}g/\partial x^\alpha)(\mathbf{x})$ in a G-connected open neighborhood of $i(r(U))$. Obviously the function so defined is G^∞ and therefore, Theorem 3.2 can be extended to a G^∞-function f defined on U. Also $\hat{f} = g$, and $g \in \widehat{G}^\infty(U)$. Moreover, by Lemma 4.1(b) (10) holds. (b) When $\cup K_i$ is finite, the series (10) cuts off for some integer $\overline{m}$, and it follows that $\widehat{G}^\infty(U) = C^\infty(r(U);B)$.

Next we construct "bell" functions on superspaces.

Corollary 4.4: Let S be a superspace and $U \subset S$ a G-connected open subset. Let $U_0 \subset U_1 \subset U$ be open subsets such that $\overline{r(U_0)} \subset r(U_1)$ and is compact. Then, there exists $g \in G^\infty(U)$ such that $g = 1$ on $\widetilde{U}_0 \cap U$ and $g = 0$ on $U - \widetilde{U}_1$.

Proof: By ordinary differential analysis, there exists $f \in C^\infty(r(U);\mathbb{R})$, $\mathbb{R} \subset B$, such that $f = 1$ on $r(U_0)$, $f = 0$ on $r(U) - r(U_1)$ (in fact $0 \leqslant f \leqslant 1$) and f has nonvanishing convergence radius for its Taylor series at each point so $f \in C^\infty_*(r(U);B)$, and taking $g = Z(f)$ we obtain the properties required by the proposition.

We give below some properties of the Z operator.[6]

Proposition 4.5: (i) Let S be a superspace, $U \subset S$ a G-connected open subset, and V' a B-module. If $f \in C^\infty(r(U);B)$, $g \in C^\infty(r(U);V')$, then $Z(fg) = Z(f)Z(g)$.

(ii) Let S, S' be superspaces, $U \subset S$, $U' \subset S'$ G-connected open subsets, and V'' a B-module. If $f \in C^\infty(r(U);R^{m'})$ such that $Z(f) \in G^\infty(U;S')$ and $f(r(U)) \subset r(u')$, and $g \in C^\infty(r(U');V'')$, then $Z(g \cdot f) = Z(g) \cdot Z(f)$.

The proof is standard in the theory of expansion of functions.

In general for a superspace $S^{m,n}$ we have that the map $G^\infty(U) \xrightarrow{\wedge} \hat{G}^\infty(U) \subset C^\infty(r(U);B)$ is not injective. However, Λ is injective when restricted to the subset $G_e^\infty(U) = \{f \in C^\infty(U) | f \cdot e = f\}$ where $e: U \to U$ is the projection onto the "even" part of superspace, i.e., $e(\Sigma_{i=1}^{m+n} a^i e_i) = \Sigma_{i=1}^m a^i e_i$. Then Z is defined by $Z = (\hat{\ }|_{G_e^\infty(U)})^{-1}$. Obviously $Z(f)^\wedge = f$ for $f \in \hat{G}^\infty(U)$. On the other hand, for $f \in G^\infty(U)$, $f_e = f \cdot e \in G_e^\infty$ and $\hat{f}_e = \hat{f}$. Hence $Z(\hat{f}) = f_e$, and $G_i Z(\hat{f}) = 0$ for $i = m+1,...,m+n$. It is immediate to see that all statements made before for the Z operator in the case of a superspace $S^{m,0}$ are still valid in the general case $S^{m,n}$.

B. Expansion of G^∞-functions

In this subsection we give the expansion of G^∞-functions in terms of C^∞-functions ("superfield expansion").

We introduce the family of maps $\Pi_L: S^{m,n} \to B$ where $L \subset \{1,..., n\} = J_n$ defined by $\Pi_\varnothing(x) = 1$ and $\Pi_L(\Sigma_{i=1}^{m+n} a^i e_i) = a^{m+i_1} a^{m+i_2} \cdots a^{m+i_k}$ for $L = \{i_1,...,i_k\}$ $(i_1 < \cdots < i_k)$.

Theorem 4.6: $f \in G^\infty(U)$ iff there exists a family $f_L \in \hat{G}^\infty(U)$, $L \in F(J_n)$ such that

$$f(x) = \sum_{L \in F(J_n)} \Pi_L(x) Z(f_L)(x). \tag{11}$$

Proof: First since Π_L and $Z(f_L)$ are G^∞, $f(x) = \Sigma \Pi_L(x) Z(f_L)(x)$ is G^∞. To prove the converse we introduce some notation; $\mathbf{x} = \Sigma_{i=1}^{m+n} a^i e_i = (\mathbf{y},\boldsymbol{\theta}) \in (B^0)^m \times (B^1)^n$, $y^i = a^i$ for $i = 1,...,m$ and $\theta^i = a^{j+m}$ for $j = 1,...,n$. Now given $f(x) \in G^\infty(U)$ we write the Taylor expansion around $e(\mathbf{x}) = (\mathbf{y},\mathbf{0})$,

$$f(\mathbf{y},\boldsymbol{\theta}) = f(\mathbf{y},\mathbf{0}) + \sum_{1 \leqslant i_1 < \cdots < i_k \leqslant n} \theta^{i_1} \cdots \theta^{i_k} G_{i_k+m} \cdots G_{i_1+m} f(\mathbf{y},\mathbf{0}) \tag{12}$$

which proves the theorem by observing that $\pi_L(\mathbf{y},\boldsymbol{\theta}) = \theta^{i_1} \cdots \theta^{i_k}$ and that $G_L f(\mathbf{y},\mathbf{0}) = G_L f \cdot e(\mathbf{y},\boldsymbol{\theta}) = Z(G_L f)(\mathbf{y},\boldsymbol{\theta})$. So $f_L = G_L f$. Q.E.D.

The relationship between superfields (in physics) and G^∞-functions is given by (11). This expansion corresponds to the expression of a superfield in forms of ordinary fields. It is worthwhile to observe that different families of functions f_L may correspond to the same G^∞-function f, unless S is free.

Having chosen a family $\{f_L\}$ corresponding to a G^∞-function f, partial G-derivatives of f can be fixed as follows:

$$\begin{aligned} G_i f(\mathbf{x}) &= \sum_l \Pi_L(\mathbf{x}) Z\left(\frac{\partial}{\partial x^i} f_L\right)(\mathbf{x}), \quad 1 \leqslant i \leqslant m, \\ G_i f(\mathbf{x}) &= \sum_L \Pi_L(i)(\mathbf{x}) Z(f_L)(\mathbf{x}), \quad m+1 \leqslant i \leqslant m+n, \end{aligned} \tag{13}$$

where $\Pi_L(i) = 0$ if $i \notin L$, $\Pi_L(i_r) = (-1)^{r+1} \Pi_{L-\{i_r\}}(\mathbf{x})$ for $L = \{i_1,...,i_r,...,i_k\}$ ordered.

Partial G-derivatives of higher order can be obtained by iteration.

We observe that the partial G-derivatives so fixed are G^∞-functions. Equations (13) provide the so-called induced expansion (from that of f) for partial G-derivatives. From the usual Schwartz lemma and (13) we get the Schwartz lemma for G^∞-functions as

$$G_i G_j f(\mathbf{x}) = (-1)^{r_i r_j} G_j G_i f(\mathbf{x}). \tag{14}$$

We have shown in (13) how to fix partial G-derivatives of a G^∞-function f once a family $\{f_L\}$ is given. However, there still remains the problem of how to choose such a family in the case where S is nonfree. The choice in Proposition 2.11 of Ref. 2 is unsatisfactory since the projection used to define f_L does not eliminate the ambiguity in the odd partial G-derivatives of order greater than one. Next, we give a prescription to do that.

Let us remind the reader that $q_L = I - p_L$, p_L being the continuous projection onto the annihilator of Im Π_L. (See Lemma I.2.6.)

Lemma 4.7: (i) Let $g \in C^\infty(r(U);B)$ and $L \in F(J_n)$; then for all $\mathbf{x} \in U$, $\Pi_L(\mathbf{x}) Z(g)(\mathbf{x}) = \Pi_L(\mathbf{x}) Z(q_L \cdot g)(\mathbf{x})$. (ii) If f_L, $g_L \in C^\infty(r(U);B)$, $L \in F(J_n)$, and $\Sigma_L \Pi_L Z_L(f_L) = \Sigma_L \Pi_L Z_L(g_L)$, then $q_L \cdot f_L = q_L \cdot g_L$ for all L.

Proof: (i) Since Z is L_B-linear, it is enough to prove that $\Pi_L Z(p_L \cdot g) = 0$, i.e., that $Z(p_L \cdot g)(\mathbf{x}) \in E_L$, the annihilator of Im Π_L, $\mathbf{x} \in U$. From (10) and E_L being an ideal (Lemma I.2.6) this follows. In fact, the partial derivatives of $p_L \cdot g$ are E_L-valued because $p_L \cdot g$ is E_L-valued. (ii) It is enough to prove that $\Sigma_L \Pi_L Z(f_L) = 0$ implies $q_L \cdot f_L = 0$ for all $L \in F(J_n)$. Let us suppose we have proved that $q_L \cdot Z(f_L) = 0$; then $q_L \cdot f_L = q_L \cdot Z(f_L) \cdot i = 0$. $q_L \cdot Z(f_L) = 0$ can be expressed as $\Pi_L Z(f_L) = 0$. We prove the last equality by induction over card (L). If $L = \varnothing$ and $\mathbf{x} = (\mathbf{y},\mathbf{0})$, then $\Pi_\varnothing(\mathbf{x}) = 1$ and $\Pi_L(\mathbf{x}) = 0$, $L \neq \varnothing$. Then $\Sigma_L \Pi_L(\mathbf{x}) Z(f_L)(\mathbf{x}) = Z(f_\varnothing)(\mathbf{x}) = 0$. For any $\mathbf{x} = (\mathbf{y},\mathbf{z})$, $Z(f_\varnothing)(\mathbf{x}) = Z(f_\varnothing)(\mathbf{y},\mathbf{0}) = 0$. We now assume that $\Pi_L Z(f_L) = 0$ for card $(L) \leqslant k-1$, $k \geqslant 1$. Let L be fixed with card $(L) = k$ and $\mathbf{x} = (\mathbf{y},\mathbf{z})$, $z^i = 0$ if $i \notin L$. Then $\Pi_{L'}(\mathbf{x}) = 0$ if $L' \neq L$ and card $(L') \geqslant k$. Using the induction hypothesis

$$\sum_{L'} \Pi_{L'} Z(f_{L'}) = \sum_{\mathrm{card}\,(L') \geqslant k} \Pi_{L'} Z(f_{L'}),$$

hence

$$0 = \sum_{\mathrm{card}\,(L') \geqslant k} \Pi_{L'}(\mathbf{x}) Z(f_{L'})(\mathbf{x}) = \Pi_L(\mathbf{x}) Z(f_L)(\mathbf{x}).$$

For a general $\mathbf{x} = (\mathbf{y},\mathbf{z}) = (\mathbf{y},\mathbf{z}_L) + (\mathbf{0},\mathbf{z}'_L)$, $z_L^i = z^i$, $i \in L$, and $z_L^i = 0$, $i \notin L$. Then $\Pi_L(x) = \Pi_L(\mathbf{y},\mathbf{z}_L)$ and $Z(f_L)(\mathbf{x}) = Z(f_L)(\mathbf{y},\mathbf{z}_L)$. This proves the induction.

Theorem 4.8: There is a rule to assign to each G^∞-function a unique expansion, called the canonical expansion, such that the induced expansion for partial G-derivatives is canonical.

Proof: The rule to obtain the canonical expansion of f is given by the family $\{q_L \cdot f_L\}$ if the family $\{f_L\}$ is any expansion of f. This rule is well defined because of Lemma 4.3.

To prove that the induced expansion is canonical, we first observe that $q_L \cdot f_L = f_L$ iff $\{f_L\}$ is canonical. For $1 < i < m$, $q_L((\partial/\partial x^i)q_L \cdot f_L) = (\partial/\partial x^i)q_L \cdot f_L$ because $(\partial/\partial x^i)q_L \cdot f_L$ has its image in Im q_L. Now, if $\{f_L\}$ is canonical, this proves that $\{(\partial/\partial x^i)f_L\}$, $1 \leqslant i \leqslant m$, is canonical. For $m+1 \leqslant i \leqslant m+n$, we have $(G_i f)_{L-\{i\}} = f_L$, $i \in L$ in the induced expansion, and $q_{L-\{i\}} \cdot q_L = q_L$, $i \in L$ proves that $\{(G_i f)_{L-\{i\}}\}_{i \in L}$ is canonical if $\{f_L\}$ is canonical.

Although the results of this section have been derived for G^∞-functions taking values in B, they are also valid for G^∞-functions taking values in a B-space. We observe that $\widehat{G}^\infty(U;F)$ is a proper subalgebra of $C^\infty(U;F)$ and the functions $f \in \widehat{C}^\infty(U;F)$ and their derivatives can be subjected to many restrictions to have their image into F. Moreover, for the canonical expansion of $f \in G^\infty(U;F)$ some f_L must necessarily vanish (depending of the B-space dimension of F).

5. CONCLUSION

In this paper the analysis on generalized superspaces has been discussed. First of all, the concept of G-differentiability has been established. The successive differentials of G^∞-function are L_B-multilinear operators involving extension problems (see I) and the lack of uniqueness of partial G-derivatives. We have solved this problem by means of the canonical expansion of G^∞-functions introduced in Sec. 4. The extension theorem has been revisited for generalized superspaces. Some new important properties of the extension are proved: composition of functions is stable, extension of a diffeomorphism is again a diffeomorphism, and analyticity is preserved by the extension. A linear operator Z has been constructed which allows one to produce G^∞-functions out of C^∞-functions. In general not every C^∞-function has an image in G^∞ through Z, but bell functions can be mapped on nonanalytic G^∞-functions. Nevertheless, Z maps analytic functions on analytic functions.

The next paper of the series (III) will be devoted to the study of supermanifolds.

[1]J. Hoyos, M. Quirós, J. Ramírez Mittelbrunn, and F. J. de Urríes, J. Math. Phys. **25**, 833 (1984), hereafter referred to as I.

[2]A. Rogers, J. Math. Phys. **21**, 1352 (1980).

[3]A. Jadczyk and K. Pilch, Commun. Math. Phys. **78**, 373 (1981).

[4]J. Dieudonné, *Foundations of Modern Analysis* (Academic, New York, 1960).

[5]S. Lang, *Differential Manifolds* (Addison-Wesley, Reading, MA, 1972).

[6]L. Bonora, P. Pasti, and M. Tonin, "Supermanifolds and BRS Transformations," Preprint IFPD 48/81, April 1981.

Generalized supermanifolds. III. ρ-supermanifolds

J. Hoyos
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Matemáticas, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

M. Quirós
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain

J. Ramírez Mittelbrunn and F. J. de Urríes
Instituto de Estructura de la Materia, Serrano 119, Madrid-6, Spain and Departamento de Física, Universidad de Alcalá de Henares, Alcalá de Henares, Madrid, Spain

(Received 26 November 1982; accepted for publication 30 September 1983)

ρ-supermanifolds are characterized as supermanifolds modeled in a superspace S having a well-behaved body. Tangent, cotangent, and tensor bundles to ρ-supermanifolds are discussed. The elements of the theory of Lie supergroups and principal superfiber bundles as ρ-supermanifolds are given.

PACS numbers: 02.40. + m, 11.30.Pb

1. INTRODUCTION

This is the third paper of a series in which we develop step-by-step a theory of generalized supermanifolds, i.e., manifolds containing commuting and anticommuting coordinates. Our aim is to give a mathematical setup capable to serve as framework for a rigorous study of geometrical problems appearing in supersymmetric theories.

We shall refer here to the two preceding papers as I and II.[1] In Paper I a model for generalized superspaces has been constructed using as a building block a Grassmann–Banach algebra B. Also L_B-operators between superspaces and their matrix representations have been studied there. In Paper II the analysis on superspaces was developed, centering the discussion around the concept of G^∞-functions and their properties.

The present paper is devoted to the study of supermanifolds modeled in generalized superspaces.

We start in Sec. 2 with a concept of supermanifold which is a direct generalization of that of Ref. 2. However, the existence of a well-behaved body for the supermanifolds impose restrictions on them. As a result of the analysis of these restrictions we arrive at the concept of ρ-supermanifolds. In Sec. 3 we give precise definitions of ρ-supermanifolds and of G^∞-functions defined on them. Ordinary C^∞-manifolds can be considered as a particular case of ρ-supermanifolds, for which the ρ-supermanifold and its body coincide. The existence of partitions of unity for ρ-supermanifolds is established in Sec. 4. Section 5 is devoted to the study of G^∞-vector fields and G^∞-derivations on supermanifolds. We find that G^∞-vector fields are a proper subset of G^∞-derivations and we reserve the name of supervector fields for G^∞-derivations. Then, apart from the tangent bundle, whose G^∞-sections are the G^∞-vector fields, we give room to the G^∞-derivations as sections of a "supertangent" fiber bundle in Sec. 6. In this way we define a class of vector bundles, which we call supervector bundles or L_B-bundles, characterized by the fact of having L_B-transition functions. In Sec. 7 we deal with supertensor fields and differential superforms on ρ-supermanifolds. Finally in Sec. 8 Lie supergroups and principal superfiber bundles are defined and their main properties within the framework of ρ-supermanifolds outlined.

In some parts of this paper we restrict our study to the free superspaces (see I). This restriction simplifies the discussion considerably and, on the other hand, is the most relevant case in physics, because nonvanishing products of an arbitrary number of anticommuting fields are required.

2. DISCUSSING THE CONCEPT OF SUPERMANIFOLD

G^∞-supermanifolds were first introduced by Rogers.[2] We start from that definition of G^∞-supermanifold, focusing our attention on having an operative body for the supermanifold.

Definition 2.1: Let S be a superspace and E a Banach C^∞-manifold[3] modeled on S. (a) A G^∞-atlas on E is a C^∞-atlas $\{(U\alpha,\Psi\alpha)\}\alpha\in\Lambda$ on E, such that for each pair $\alpha,\beta\in\Lambda$ $\psi_{\alpha\beta}=\psi_\beta\cdot\psi_\alpha^{-1}$ is G^∞ from $\psi_\alpha(U_\alpha\cap U_\beta)$ on to $\psi_\beta(U_\alpha\cap U_\beta)$. (b) A G^∞-structure on E is a maximal G^∞-atlas on E. (c) A G^∞-supermanifold E is a Banach C^∞-manifold endowed with a G^∞-structure. If S is free we call E a free G^∞-supermanifold.

As we saw in the previous papers I and II of this series, a major feature of the superspace is the existence of a body map $r:S^{m,n}\to R^m$. In the following we analyze the sense in which it is possible to also define a body $\rho: E\to E_\rho$ for the supermanifold with similar properties to r. This analysis will lead us, in turn, to reformulate the concept of supermanifold.

Let us study to what extent the body of a supermanifold can be defined through an equivalence relation. This question was earlier discussed by Rogers[2] following the work of de Witt.[4] There the equivalence relation $x\sim y$ iff $\exists$ (U,ψ) G^∞-chart such that $x,y\in U$ and $r\psi(x)=r\psi(y)$ is proposed. This relation is, however, of no use in Haussdorff manifolds (with which we shall deal) because all points become related into a single class. To avoid the triviality of the former relation we could introduce restrictions on the allowed charts. In view of the natural domain of definition of G^∞-functions discussed in II, the suited condition would seem to be $\psi(U)$ cG-connected (connected $+$ G-connected). Unfortunately this condition destroys the transitivity of the relation. Let us illustrate it with an example.

Example 2.2: Let us take the superspace $S^{1,1} = R \times B^1$ where B^1 is the odd part of the Grassmann algebra generated by the two anticommuting generators β_1 and β_2. Let us consider the two open subsets A and B of $S^{1,1}$ defined as

$$A = \{(x, y_1 B_1 + y_2\beta_2) | x \in \mathbb{R}, (0 < y_1 < 1 \text{ and } 0 < y_2 < 3) \text{ or } (1 \leqslant y_1 < 2 \text{ and } (0 < y_2 < 1 \text{ or } 2 < y_2 < 3))\} \tag{1}$$

and

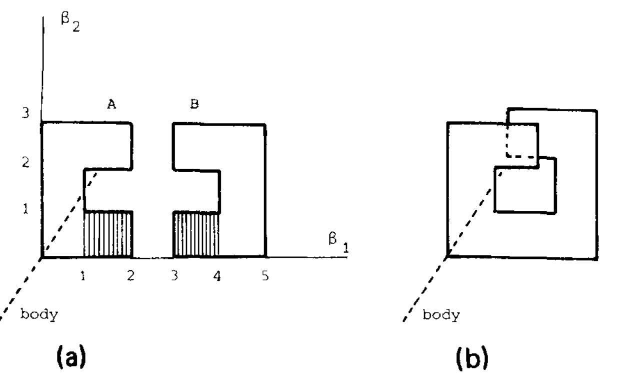
$$B = \{(x, y_1\beta_1 + y_2\beta_2) | x \in \mathbb{R}, (3 < y_1 \leqslant 4 \text{ and } (0 < y_2 < 1 \text{ or } 2 < y_2 < 3)) \text{ or } (4 < y_1 < 5 \text{ and } 0 < y_2 < 3)\}. \tag{2}$$

Next, we take in $A \cup B$ the equivalence relation

$$(x, y_1\beta_1 + y_2\beta_2)\mathscr{R}(x', y_1'\beta_1 + y_2'\beta_2)$$
$$\Leftrightarrow \begin{cases} x = x', y_1 = y_1', y_2 = y_2', \text{ or} \\ x = x', 1 < y_2 < 2, 0 < y_2 < 1, y_1' = y_1 + 2, y_2' = y_2, \text{ or} \\ x = x', 3 < y_1 < 4, 0 < y_2 < 1, y_1 = y_1' + 2, y_2' = y_2. \end{cases} \tag{3}$$

Now we consider the quotient space $E = A \cup B / \mathscr{R}$ [see Fig. 1(a) and 1(b)] and the canonical projection π: $A \cup B \to E$. Then $\pi(A) = \overline{A}$ and $\pi(B) = \overline{B}$ are open subsets of E, and π/A, π/B are homeomorphisms. We now endow E with a G^∞-atlas consisting of two charts $(\overline{A}, \psi)$ and $(\overline{B}, \phi)$, with $\psi = \pi/A^{-1}$ and $\psi = t \cdot \pi/B^{-1}$, where t is the translation on $S^{1,1}$ given by $t(x, y_1\beta_1 + y_2\beta_2) = (x, (y_1 - 2)\beta_1 + y_2\beta_2)$. Then $\psi(\overline{A}) = A$ and $\phi(\overline{B}) = t(B)$ are open cG-connected sets, $\psi(\overline{A} \cap \overline{B}) = \phi(\overline{A} \cap \overline{B}) = \{(x, y_1\beta_1 + y_2\beta_2) | x \in \mathbb{R}, 1 < y_1 < 2, 0 < y_2 < 1\}$ and $\psi \cdot \phi^{-1} = \phi \cdot \psi^{-1} = I_C$. So we have a G^∞-supermanifold structure on E.

Let us now see that the relation $\sim$ fails to be transitive. Take $u = (0, \frac{3}{2}\beta_1 + \frac{5}{2}\beta_2) \in A$, $v = (0, \frac{7}{2}\beta_1 + \frac{5}{2}\beta_2) \in B$, $w = (0, \frac{3}{2}\beta_1 + \frac{3}{2}\beta_2) \in A$, and $\pi(u) = \bar{u}$, $\pi(v) = \bar{v}$, $\pi(w) = \overline{w}$. Then $\bar{u} \sim \overline{w}$ through $(\overline{A}, \psi)$ and $\overline{w} \sim \bar{v}$ through $(\overline{B}, \phi)$. However, there is no G^∞-chart (U, δ) with $\delta(U)$ cG-connected with $\bar{u}, \bar{v} \in U$. In fact, would such a chart exist, the transition functions $\delta\psi^{-1}$ and $\delta\phi^{-1}$, being G^∞, should have the form



(a) (b)

FIG. 1. (a) A section of the subset $A \cup B$ with constant body. (b) The manifold $E = A \cup B / R$. The points in the two shadowed regions of A and B have been identified.

$$[\delta \cdot \psi^{-1}](x, y_1\beta_1 + y_2\beta_2) = (a(x), b(x) + (y_1\beta_1 + y_2\beta_2)c(x)),$$
$$[\delta \cdot \phi^{-1}](x', y_1'\beta_1 + y_2'\beta_2) = (a'(x'), b'(x') + (y_1'\beta_1 + y_2'\beta_2)c'(x')).$$

Then, since $\delta(U)$ is cG-connected we have $U \cap \overline{A} \cap \overline{B} \neq \varnothing$. As $\phi \cdot \psi^{-1} = I_C$, we arrive at $\delta \cdot \psi^{-1} = \delta \cdot \phi^{-1}$ on $\psi(U \cap \overline{A} \cap \overline{B})$. Hence $a = a'$, $b = b'$, $c = c'$. As a consequence the contradiction $\delta(\bar{u}) = \delta(\bar{v})$ with $\bar{u} \neq \bar{v}$ (δ must be injective) results. Therefore the chart (U, δ) does not exist and the relation $\sim$ fails to be transitive.

Although in the former example the relation $\sim$ fails to be an equivalence, it is clear that a body can be naturally defined by $\rho(x, y_1\beta_1 + y_2\beta_2) = X$. On the other hand the relation $\sim$ is reflexive and symmetric, and can be extended to an equivalence relation $\approx$ on E by means of finite chains: $x \approx y$ iff $\exists\, x_1, \ldots, x_n$ such that $x \sim x_1, x_1 \sim x_2, \ldots, x_n \sim y$. Then in Example 2.2 $E/\approx\, = \mathbb{R}$ which is the manifold that we would look at naturally as the body of E. In general we can define the body of E as $E\rho = E/\approx$ with the quotient topology. The body map is then the canonical projection of $\approx$. We observe that if $x \sim y$, a path joining x and y with constant body does exist; therefore the same is true for $x \approx y$. Then the body of E is determined by a given G^∞-atlas. However, this constructive procedure would be useless to find explicitly the body in a particular supermanifold. The main reason is that the coordinates in a given chart (U, ψ) of two related points $x \approx y$ could have a different body, i.e., $r\psi(x) \neq r\psi(y)$. We give now an example to illustrate this point.

Example 2.3: Let us take the superspace $S^{1,1} = \mathbb{R} \times B^1$, where B^1 is the odd part of the Grassmann algebra generated by one generator β. We consider the band A in $\mathbb{R}^2$, $A = \{(x, y) \in \mathbb{R}^2 | x \in \mathbb{R} \text{ and } 0 \leqslant y \leqslant 1\}$, and make a cylinder E of it by identifying $(x,0)$ with $(x,1)$ for $x \in \mathbb{R}$. We proceed to give a G^∞-atlas to E.

Let $U = \{[(x, y)] \in E | x \in \mathbb{R}, 0 \leqslant y < \frac{1}{3} \text{ or } \frac{2}{3} < y \leqslant 1\}$ and $V = \{[(x, y)] \in E | x \in \mathbb{R}, 0 < y < 1\}$. $[(x,y)]$ is the equivalence class of $(x,y) \in A$ in E, and U, V are open connected sets in E.

Let ψ: $U \to S^{1,1}$ given by

$$\psi([(x,y)]) = \begin{cases} (x, y\beta), x \in \mathbb{R}, \frac{2}{3} < y \leqslant 1, \\ (x, (y+1)\beta), x \in \mathbb{R}, 0 \leqslant y < \frac{1}{3} \end{cases} \tag{4}$$

and ϕ: $V \to S^{1,1}$ given by

$$\phi([(x,y)]) = \begin{cases} (x, y\beta), x \in \mathbb{R}, 0 < y \leqslant \frac{1}{3}, \\ (x + y - \frac{1}{3}, y\beta), x \in \mathbb{R}, \frac{1}{3} \leqslant y \leqslant \frac{2}{3}, \\ (x + \frac{1}{3}, y\beta), x \in \mathbb{R}, \frac{2}{3} \leqslant y < 1. \end{cases} \tag{5}$$

Then $\psi(U)$ and $\phi(V)$ are cG-connected open subsets of $S^{1,1}$. Moreover, $\psi \cdot \phi^{-1}$: $\phi(U \cap V) \to \psi(U \cap V)$ is G^∞ and so is its inverse $\phi \cdot \psi^{-1}$.

Let us now take $[(x, \frac{5}{6})] \in E$. Then $[(x, \frac{5}{6})] \sim [(x + \frac{1}{3}, \frac{1}{6})]$ through V. But $[(x + \frac{1}{3}, \frac{1}{6})] \sim [(x + \frac{1}{3}, \frac{5}{6})]$ through U.

Therefore $[(x, \frac{5}{6})] \approx [(x + \frac{1}{3}, \frac{5}{6})]$. Using the same argument and induction we see that $[(x, y)] \approx [(x + n/3, y)]$ $\forall n \in Z$. So we have $r\psi[(x,y)] = x \neq x + n/3 = r\psi[(x + n/3, y)]$, $n \neq 0$, but $\rho([(x, y)]) = \rho([(x + n/3, y)])$.

Nevertheless, in this example it is easy to see what the body and the lines of constant body are. The body $E/\approx\, = S^1 = \mathbb{R}/Z$ and the lines of constant body are drawn in Fig. 2.

We remark that $E \neq S^1 \times B^1$ as a manifold, since the natural bijection between them cannot be continuous, because U is connected in E but its image in $S^1 \times B^1$ is not.

We now go further in the discussion of the body defined by $\approx$. We shall see in the next example that this body could fail to be a manifold.

Example 2.4: Let $E = \mathbb{R}/Z \times \mathbb{R}/Z$ be the torus. A G^∞-atlas for E is given by

$$
\begin{aligned}
U_1 &= \{[(x,y)] \mid 0 < x < 1, 0 < y < 1\},\\
U_2 &= \{[(x,y)] \mid 0 < x < 1, \tfrac{1}{2} < y < \tfrac{3}{2}\},\\
U_3 &= \{[(x,y)] \mid \tfrac{1}{2} < x < \tfrac{3}{2}, 0 < y < 1\},\\
U_4 &= \{[(x,y)] \mid \tfrac{1}{2} < x < \tfrac{3}{2}, \tfrac{1}{2} < y < \tfrac{3}{2}\},
\end{aligned} \tag{6}
$$

and

$$\psi_i([x,y]) = (y - \alpha x, x\beta), \quad i = 1, 2, 3, 4, \tag{7}$$

where α is an irrational number $\alpha \in \mathbb{R} - Q$. The transition functions are always translations. Hence $\{(U_i, \psi_i), 1 \leqslant i \leqslant 4\}$ defines a G^∞-supermanifold structure on E. Now each line $\{(x,y) \in \mathbb{R}^2 | y = \alpha x + b\} = L_b \subset \mathbb{R}^2$, with $b \in \mathbb{R}$, corresponds to a line of constant body on E. Then the body of E is $\mathbb{R}/\alpha Z + Z$, which is not a manifold with the quotient topology.

In view of these examples, we should look for conditions ensuring the existence of a well-behaved body. A natural condition to impose is

Condition (*a*): Existence of a manifold E_ρ, a G^∞-atlas with cG-charts $\{(U,\psi)\}$ on E, and a C^∞-atlas $\{(V, \phi)\}$ on E, such that $\rho(U) = V$ and $\phi \cdot \rho = r\psi$.

Condition (a) is essentially the same as that imposed by Marchetti and Percacci[5] in their definition of supermanifold with body. Unfortunately, it is not strong enough to guarantee the existence of nontrivial G^∞-functions on supermanifolds. The reason is that compactness along the soul is allowed, and this in turn excludes G^∞-functions nonconstant along the soul. We illustrate this point with another example:

Example 2.5: Let us take the cylinder $E = \mathbb{R} \times \mathbb{R}/Z$. We endow E with a G^∞-supermanifold structure through the atlas

$$
\begin{aligned}
U_1 &= \{[(x,y)] \mid x \in \mathbb{R}, 0 < y < 1\},\\
U_2 &= \{[(x,y)] \mid x \in \mathbb{R}, \tfrac{1}{2} < y < \tfrac{3}{2}\},\\
\psi_i([x,y]) &= (x, y\beta) \in S^{1,1},
\end{aligned} \tag{8}
$$

where $S^{1,1}$ is the same superspace as in Example 2.4.

The body of E is $E\rho = \rho(U_1) = \rho(U_2) \approx \mathbb{R}$. Then it is immediate that the isomorphism $\phi: E_\rho \to \mathbb{R}$ defines a manifold structure in E_ρ and satisfies $r \cdot \psi_i = \phi \cdot \rho$.

We now take a G^∞-function $f: E \to B$, given in coordinates by $f \cdot \psi_i^{-1} = f_i(x,y\beta) = a_i(x) + y\beta b_i(x)$. For $(x,y\beta)$ with $\frac{1}{2} < y < 1$ we obtain $f_1(x,y\beta) = f_2(x,y\beta)$; therefore $a_1(x) = a_2(x) = a(x)$ and $b_1(x) = b_2(x) = b(x)$. On the other hand, for $(x, y\beta)$ with $0 < y < \frac{1}{2}$ we obtain $f_1(x,y\beta) = f_2(x,(y+1)\beta)$; therefore $a(x) + y\beta b(x) = a(x) + (y+1)\beta b(x)$. So $\beta b(x) = 0$ and $f_i(x,y\beta) = a(x)$. The last means that f depends only on the body, i.e., $f = a \cdot \rho$ where $a: E\rho \to B$ is C^∞. In general, this phenomenon will always appear when we have $\rho^{-1}(x)$ compact. In this case G^∞-functions are polynomials and periodic along $\rho^{-1}(x)$, so they must be constants on $\rho^{-1}(x)$ for a fixed $x \in E\rho$. So they depend only on x.

From these considerations we conclude that a further requirement is needed in order to have G^∞-supermanifolds with body which can support nontrivial G^∞-functions. To Condition (a) the following must be added.

Condition (*b*): The atlas $\{(U,\phi)\}$ and $\{(V,\phi)\}$ of Condition (a) also satisfy $\rho^{-1}(V) = U$.

In the next section we begin giving a formalized definition of ρ-supermanifold which meets all requirements discussed above.

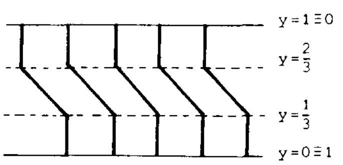


FIG. 2. A line of constant body on a supermanifold constructued over a cylinder.

3. ρ-SUPERMANIFOLDS. G^∞-FUNCTIONS

As pointed out at the end of the last section, we begin with the definition of ρ-supermanifold (supermanifold with an operative body).

Definition 3.1: Let E be a C^∞-manifold modeled on a superspace $S_1^{m,n}$, E_ρ a C^∞-manifold modeled on $\rho(S) = \mathbb{R}^m$, and ρ a C^∞-map from E onto E_ρ. (a) A ρ-atlas on the triple (E,E_ρ,ρ) is a pair of a G^∞-atlas $\{(U_\alpha, \psi_\alpha)\}$ on E and a C^∞-atlas $\{(U_{\rho\alpha}, \psi_{\rho\alpha})\}$ on E_ρ such that (i) $\psi_\alpha(U_\alpha)$ is cG-connected, (ii) $U_\alpha = \rho^{-1}(U_{\rho\alpha})$, (iii) $\psi_{\alpha\rho} \cdot \rho = r \cdot \psi_\alpha$ (where r is the body map on $S^{m,n}$). (b) A G^∞-ρ-structure on the triple (E,E_ρ,ρ) is a maximal ρ-atlas. (c) A ρ-supermanifold is a triple (E,E_ρ,ρ) with a G^∞-ρ-structure on the triple.

The manifold E_ρ is homeomorphic to the body $E/\approx$. This allows one to endow the body $E/\approx$ with a structure of C^∞-manifold. In what follows we shall not distinguish anymore between E_ρ and $E/\approx$. Superspace is an example of ρ-supermanifold. An open subset $U \subset E$ is said to be G-connected if for each ρ-chart (U_1,ψ_1) of E, $\psi_1(U_1 \cap U)$ is G-connected. A G-connected subset of a ρ-supermanifold is itself a ρ-supermanifold. The direct product of two ρ-supermanifolds is again a ρ-supermanifold with the obvious ρ-atlas. Example 2.3 is a ρ-supermanifold, i.e., it admits a ρ-atlas compatible with the atlas given there.

Given a ρ-supermanifold E, we shall define its ρ-saturation $\widetilde{E}$ which is itself a ρ-supermanifold with the same body. Moreover, E will be an open submanifold of $\widetilde{E}$. We proceed to construct $\widetilde{E}$.

Let $\{(U_\alpha,\psi_\alpha)\}\alpha \in \Lambda$ be a ρ-atlas of E. For each $\alpha \in \Lambda$, let $(\psi_\alpha(U_\alpha)^\sim)$ be the ρ-saturation of $\psi_\alpha(U_\alpha)$ and $E' = U'_{\alpha\in\Lambda}(\psi_\alpha(U_\alpha)^\sim)$ the disjoint union. We define an equivalence relation in E' by $x\,R\,y$ iff $\tilde{\psi}_{\beta\alpha}(x) = y$, $x \in (\psi_\alpha(U_\alpha)^\sim)$, $y \in (\psi_\beta(U_\beta)^\sim)$. R is an equivalence relation by the results of Paper II, Sec. 3. Let $\widetilde{E} = E'/R$. It is well known that $\widetilde{E}$ is a manifold. We now define $\tilde{\rho}:\tilde{E}\to E_\rho$ by $\tilde{\rho}([x]) = \psi_{\rho\alpha}^{-1}\cdot r(x)$ if $x\in(\psi_\alpha(U_\alpha)^\sim)$; $\tilde{\rho}$ is clearly a well-defined C^∞-map from $\tilde{E}$ onto E_ρ. A ρ-atlas for $\tilde{E}$ is constructed as follows: let $\tilde{U}_\alpha = [\psi_\alpha(U_\alpha)^\sim]$ and $\tilde{\psi}_\alpha:\tilde{U}_\alpha\to S$ be the map $\tilde{\psi}_\alpha([x]) = x$ for $x\in\psi_\alpha(U_\alpha)^\sim$. Finally, we give the inclusion j from E into $\tilde{E}$: $j(x) = [\psi_\alpha(x)]$ if $x\in U_\alpha$. By identifying E with its image we have $U_\alpha\subset\tilde{U}_\alpha$, $\psi_\alpha = \tilde{\psi}_\alpha|_{U_\alpha}$ and $\rho = \tilde{\rho}/E$.

Let us remark that $\tilde{E}$ is a de Witt[4] supermanifold with a stronger topology. As observed in Ref. 2 the fiber of $\tilde{E}$ is a vector space. However, $\tilde{E}$ is not a vector bundle because the transition functions fail to be linear in the fibers.

Next we define G^∞-functions on ρ-supermanifolds. Let E_1, E_2 be ρ-supermanifolds and V a B-module. Let $U\subset E_1$ be an open subset and f: $U\to V$; f is G^∞ iff for each ρ-chart (U_1, ψ_1) of E_1, $f\cdot\psi_1^{-1}$ is a G^∞-function from $\psi_1(U_1\cap U)$ into V. Similarly, f: $U\to E_2$ is G^∞ iff for each ρ-chart (U_1,ψ_1) of E_1 and (U_2,ψ_2) of E_2, $\psi_2\cdot f\cdot\psi_1^{-1}$ is a G^∞-function from $\psi_1(U_1\cap U)$ into $\psi_2(U_2)$.

Let $G^\infty(U,V)$ be the set of G^∞-functions from U into V. In the particular case $V=B$, we simply denote it by $G^\infty(U)$. Proposition 2.2 of Paper II applies here.

Since each G^∞-function defined on E can be uniquely extended to $\tilde{E}$, we shall deal in the following only with saturated ρ-supermanifolds and drop the notation $\sim$. Given a subset U of E, the saturation $\tilde{U}$ of U is defined as $\tilde{U} = \rho^{-1}(\rho(U))$.

As in Proposition II. 3.6 for superspaces; if $U\subset E$ is a G-connected open set and f is a G^∞-function from U into E_2, there exists a unique C^∞-function f_ρ from $\rho(U)\equiv U_\rho\subset E_{1\rho}$ into $E_{2\rho}$ such that $f_\rho\cdot\rho = r\cdot f$.

Let us finally observe that if E is a ρ-supermanifold and (U,ψ) a ρ-chart of E, then U is G-connected, ψ is a G^∞-function, and ψ_ρ is the induced map from U_ρ into $r(S)$.

4. PARTITIONS OF UNITY

We deal in this section with the existence of "bell" functions on ρ-supermanifolds and the problem of constructing partitions of unity.

Lemma 4.1: Let E be a ρ-supermanifold, $U\subset E$ an open set, and $p\in U$. Then, there exists an open neighborhood U_0 of p, $U_0\subset U$, and $g\in G^\infty(E)$ such that $g=1$ on $\tilde{U}_0$, $g=0$ on $E-\tilde{U}$ and $0\leqslant r\cdot g\leqslant 1$.

Proof: Let (V,ϕ) be a ρ-chart with $p\in V$. Take $W=\phi(\tilde{U}\cap V)$. Since r is an open map $r(W)$ is open in $\mathbb{R}^m$, and $r\cdot\phi(p)\in r(W)$. Then there exist two relatively compact open sets Z_0, Z such that $r\cdot\phi(p)\in Z_0\subset\bar{Z}_0\subset Z_1\subset\bar{Z}_1\subset r(W)$. As a consequence $\phi(p)\in U'_0\subset\bar{U}'_0\subset U'_1\subset\bar{U}'_1\subset W$ where $U'_0 = r^{-1}(Z_0)$, $U' = r^{-1}(Z_1)$. Now by Corollary II. 4.4 applied to $U'_0\subset U'_1\subset\phi(V)$ there exists $h\in G^\infty(\phi(V))$ such that $h=1$ on U'_0 and $h=0$ on $\phi(V)-U'_1$. We can also assume that $0\leqslant r\cdot h\leqslant 1$ [choosing $h = Z(r\cdot h)$].

Next, we define $U_0 = \phi^{-1}(U'_0)\cap U$ and f: $E\to B$ by $g(q) = h\cdot\phi(q)$ for $q\in V$ and $g(q) = 0$ for $q\notin V$. Then g is a well-defined G^∞-function satisfying the conditions of the lemma, because $\phi^{-1}(\bar{U}'_1)\subset\tilde{U}\cap V\subset V$. Q.E.D.

A similar result also holds replacing the point p by a set K, with $\rho(K)$ relatively compact.

Corollary 4.2: Let E, U be as in Lemma 4.1. Let $K\subset U$ be such that $\rho(K)$ is compact. Then, there exists a neighborhood U_0 of K, $U_0\subset U$ and $g\in G^\infty(E)$ such that $g=1$ on $\tilde{U}_0$, $g=0$ on $E-\tilde{U}$, and $0\leqslant g_\rho\leqslant 1$.

Proof: For each $p\in K$, there exists $p\in U_p\subset U$ and $g_p\in G^\infty(E)$ such that $g_p = 1$ on $\tilde{U}_p$, $g_p = 0$ on $E-\tilde{U}$, and $0\leqslant r\cdot g_p\leqslant 1$. We can assume that $\tilde{U}_p = U_p$. We have $\rho(K)\subset\cup_{p\in K}\rho(U_p)$ and $\rho(U_p)$ open. Since $\rho(K)$ is compact, there exist $p_1,\ldots,p_n\in K$ such that $\rho(K)\subset\cup_{i=1}^n\rho(U_{p_i})$. Let $U_0 = \rho^{-1}(\cup_{i=1}^n\rho(U_{p_i}))\cap U = \cup_{i=1}^n U_{p_i}$ and $g\in G^\infty(E)$ be $g = 1-(1-g_{p_1})(1-g_{p_2})\cdots(1-g_{p_n})$. Trivially $g=1$ on $\tilde{U}_0$, $g=0$ on $E-\tilde{U}$, and $0\leqslant q_\rho\leqslant 1$. Q.E.D.

Theorem 4.3 (partitions of unity): Let E_ρ be paracompact and let $\{U_\alpha\}$ be an open covering of E. Then, there exists a partition of unity subordinate to $\tilde{U}_\alpha$, i.e., there exists a family $\{\psi_\alpha\}$, $\psi_\alpha\in G^\infty(E)$ such that (i) $\mathrm{Supp}(\psi_\alpha)\subset\tilde{U}_\alpha$ (ii) $\{\mathrm{Supp}(\psi_\alpha)\}$ is a locally finite covering of E, (iii) $(\psi_\alpha)_\rho(x)\geqslant 0$ $\forall x\in E_\rho$, and (iv) $\Sigma_\alpha\psi_\alpha(y) = 1, y\in E$.

The proof uses the existence of partitions of unity in E_ρ, Lemma 4.1, and the fact that Z is a linear operator.

5. SUPERVECTOR FIELDS

We shall discuss in this section vector fields on ρ-supermanifolds. It will emerge from the discussion that G^∞-vector fields do not coincide with G^∞-derivations, and two tangent spaces must be considered. First, we shall endow the usual tangent space $T(E)$ over a ρ-supermanifold E with a ρ-structure. Let $T(E_\rho)$ be the tangent space of E_ρ (as Banach manifold), and ρ_*: $T(E)\to T(E_\rho)$ the derivative of ρ. Then, for $\{(U_\alpha,\psi_\alpha)\}$ a ρ-atlas of E, $\{T(U_\alpha),\psi_{\alpha *}\}$ is a ρ-atlas for the triple $(T(E), T(E_\rho), \rho_*)$. This defines a ρ-supermanifold structure which does not depend on the specific ρ-atlas chosen for E.

As $T(E)$ is a ρ-supermanifold, it makes sense to consider G^∞-vector fields on E. Moreover, if X: $E\to T(E)$ is a G^∞-vector field, it is immediate that X can be considered as a derivation on $G^\infty(E)$. However, G^∞-vector fields do not exhaust all derivations on $G^\infty(E)$.

Definition 5.1: (i) A supervector field X on E is a derivation of $G^\infty(E)$; i.e., an element $X\in L_B(G^\infty(E))$ such that

$$X(fg) = (Xf)g + (-1)^{rs}f(Xg)$$

with $X\in L_B(G^\infty(E))^r$, $f\in G^\infty(E)^s$, $g\in G^\infty(E)$. (ii) A G^∞-vector field is a G^∞-section of $T(E)$.

Let us observe that a G^∞-vector field is image preserving, i.e., for each $K\subset J$ and $f\in G^\infty(E)$, $(\mathrm{Im}\,f)\subset B'_k$ $\Rightarrow\mathrm{Im}(XG)\subset B'_k$ for XG^∞-vector field.

As a consequence of Corollary 4.2 we have that if $U\subset E$ is open and $f\in G^\infty(E)$ vanishes on U, then $Xf=0$ for every supervector field X.

Then, the field X in a chart (U,ψ) will be expressed by $X = \Sigma_{i=1}^{m+n} a^i\,\partial/\partial x^i$, where $\partial f/\partial x^i = G_i(f\cdot\psi^{-1})$ and a^i: $\omega\to B$ is a G^∞-function (G_i is the canonical derivative introduced in Paper II, Sec. 4). We observe that $\partial/\partial x^i$ is not a derivation if x^i is odd and K_i is finite. However, we can write X in local coordinates and get a characterization for supervector and G^∞-vector fields.

Proposition 5.2: Let (U,ψ) be a ρ-chart of E. Then, (i) X is a supervector field on U iff $X = \Sigma_{i=1}^{m+n} a^i \partial/\partial x^i$, where $a^i \in G^\infty(U)$ and $\beta_{K_i} a^i = 0$, when $m+1 \leqslant i \leqslant m+n$ and K_i is finite. (ii) X is a G^∞-vector field on U iff X is a supervector field and $X = \Sigma_{i=1}^{m+n} a^i \partial/\partial x^i$ with $\operatorname{Im} a^i \subset B_{K_i}^{r_i}$, $r_i = 0$ for $1 \leqslant i \leqslant m$ and $r_i = 1$ for $m+1 \leqslant i \leqslant m+n$.

To prove Proposition 5.2 we need the following lemma.

Lemma 5.3: Let S be a superspace and $U \subset S$ a G-convex subset. Let $f \in G^\infty(U)$ and $\mathbf{x}_0$ U. Then, there exist $g_i \in G^\infty(U)$, $1 \leqslant i \leqslant m+n$, such that

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{m+n} (x^i - x_0^i) g_i(x)$$

and

$$g_i(\mathbf{x}_0) = G_i f(\mathbf{x}_0).$$

Proof: Taking $g_i(\mathbf{x}) = \int_0^1 G_i f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0))\,dt$, the proposition follows from the Fundamental Theorem of Calculus applied to the function $t \to f(\mathbf{x}_0 + t(\mathbf{x} - \mathbf{x}_0))$, where $G_i f$ is the canonical derivative. Q.E.D.

Proof of Proposition 5.2: By Lemma 5.3, if X is a derivation, $X = \Sigma_{i=1}^{m+n} a^i \partial/\partial x^i$ with $a^i = X(x^i) \in G^\infty(U)$. On the other hand, if K_i, $m+1 \leqslant i \leqslant m+n$, is finite, then $x^i \beta_{K_i} = 0$. Hence $0 = X(x^i \beta_{K_i}) = X(x^i) \beta_{K_i}$. Conversely as we noted before, $\partial/\partial x^i$ is a derivation unless K_i, $m+1 \leqslant i \leqslant m+n$ is finite. In that case $\partial(fg)/\partial x^i - (\partial f/\partial x^i) g - (-1)^{|f|} f(\partial g/\partial x^i)$ is a function multiple of β_{K_i}. Therefore, if $\beta_{K_i} a^i = 0$, $\Sigma_{i=1}^{m+n} a^i \partial/\partial x^i$ is a derivation.

For X a G^∞-vector field and (U, ψ) a ρ-chart, as we said before the expression in coordinates is $X = \Sigma_{i=1}^{m+n} a^i \partial/\partial x^i$, where $a = (a^i, \ldots, a^{m+n}) \in G^\infty(U, S)$. Hence (ii) holds. Q.E.D.

We shall denote the set of supervector fields, and G^∞-vector fields by $\overline{\chi}(E)$ and $\chi(E)$, respectively. We remark that the only supervector fields with associated (local) uniparametric group of transformations are the G^∞-vector fields, because each vector tangent to a curve on E is necessarily in $T(E)$.

We observe that the space of supervector fields $\overline{\chi}(E)$ is a $G^\infty(E)$-module. It is also a Lie superalgebra (over B) with the bracket $[X, Y] = XY - (-1)^{|X||Y|} YX$. However, $\chi(E)$ is not a $G^\infty(E)$-module, and the bracket gives it a structure of Lie algebra over $\mathbb{R}$.

We consider now how given a G^∞-derivation x on $G^\infty(E)$, it is possible to associate to X a derivation X_ρ (the body of X) on $C^\infty(E_\rho)$. First, in the case of a G^∞-vector field $X: E \to T(E)$ we can define a (unique) C^∞-section $X_\rho: E \to T(E_\rho)$ which provides a C^∞ derivation on E_ρ as the one satisfying $\rho_* \cdot X = X_\rho \cdot \rho$.

Taking ρ-coordinates $(x^1, \ldots, x^{m+n})$ in E and $(y^1, \ldots, y^m)$ in E_ρ with $y^i = x^i_r$ for $1 \leqslant i \leqslant m$, the local expressions of X and X_ρ in these coordinates are $X = \Sigma_{i=1}^{m+n} a^i \partial/\partial x^i$ and $X_\rho = \Sigma_{i=1}^m a^i_\rho \partial/\partial x^i$. For a general supervector field we use the last expression to define X_ρ in a ρ-chart. This defines also X_ρ globally. In fact, if we take another ρ-chart with coordinates $(\bar{x}^i)_1^{m+n}$ and $(\bar{y})_1^m$, we have $X(\bar{x}^j) = \Sigma_{i=1}^{m+n} X(x^i) \partial \bar{x}^j/\partial x^i$ and $X(\bar{x}^j)_\rho = \Sigma_{i=1}^{m+n} X(x^i)_\rho (\partial \bar{x}^j/\partial x^i)_\rho$. Now taking into account that $(\partial \bar{x}^j/\partial x^i)(P) \in L_B(S)$ and that for each function between two supermanifolds $f: E \to E'$, $(f_\rho)_* = (f_*)_\rho$, we obtain for $1 \leqslant j \leqslant m$

$$\left(\frac{\partial \bar{x}^j}{\partial x^i}\right) = \begin{cases} \dfrac{\partial \bar{y}^j}{\partial y^i}, & 1 \leqslant i \leqslant m, \\ 0, & m+1 \leqslant i \leqslant m+n \end{cases}, \tag{9}$$

so $X(\bar{x}^j)_\rho = \Sigma_{i=1}^m X(x^i)_\rho \partial \bar{y}^j/\partial y^i$.

Next we discuss the dual of $\overline{\chi}(E)$. $\chi^*(E) = \{\omega: \overline{\chi}(E) \to G^\infty(E)$, with ω a $G^\infty(E)$-linear mapping$\}$. By Lemma 4.1 if $X \in \overline{\chi}(E)$ and $X = 0$ on U open in E, then $X_\omega = 0$ on U. This allows one to construct local expressions for the forms in $\chi^*(E)$. Take the chart (U, ψ). We consider first E free. Then, $\partial/\partial x^i \in \overline{\chi}(E)$ and $\Sigma_{i=1}^{m+n} (a^i \partial/\partial x^i)\omega = \Sigma_{i=1}^{m+n} a^i (\partial\omega/\partial x^i) = \Sigma_{i=1}^{m+n} a^i \omega_i$. In this case, we define dx^i by $\partial(dx^i)/\partial x^j = \delta^i_j$, so that $\omega = \Sigma_{i=1}^{m+n} dx^i\, \omega_i$. However, if some K_i $(m+1 \leqslant i \leqslant m+n)$ is finite, $\partial/\partial x^i \notin \chi(E)$ and $(\partial/\partial x^i)\omega$ makes no sense. To avoid this problem we consider the subset $\Omega^1(U) \subset \chi^*(U)$ of maps of the form $\Sigma_{i=1}^{m+n} dx^i\, \omega_i$ with $\omega^i \in G^\infty(U)$, defined by $\Sigma_{j=1}^{m+n} (a^j \partial/\partial x^j) \Sigma_{i=1}^{m+n} dx^i\, \omega_i = \Sigma_{i=1}^{m+n} a^i \omega_i$. We observe that the coefficients ω^i are not uniquely determined because $dx^i\, \omega_i'$ gives the same linear mapping on $\overline{\chi}(E)$, provided $\omega_i' - \omega_i \in \dot{\beta}_{K_i}$ for K_i odd and finite. Now define $\Omega^1(E)$ as the subset of elements ω of $\chi^*(E)$, such that $\omega|_U \in \Omega^1(U)$ for each ρ-chart (U, ψ) of E. We remark that when E is free, $\Omega^1(E) = \chi^*(E)$.

In a similar way to the case of supervector fields the body of a 1-superform ω on E is defined as the 1-form ω_ρ on E_ρ given in coordinates by $\omega_\rho = \Sigma_{i=1}^m \omega_{\rho i}\, dy^i$, with $\omega = \Sigma_{i=1}^{m+n} \omega_i\, \mathrm{d}x^i$, where again x^i and y^i are corresponding coordinates in a ρ-chart.

6. SUPERVECTOR BUNDLES

We construct in this section L_B-bundles whose sections are supervector fields and 1-superforms. As shown in the preceding section, we have to distinguish between supervector and G^∞-vector fields. Therefore, we must introduce two different fibers and construct two tangent bundles. Also a third fiber is needed to construct the cotangent bundle. In view of Proposition 5.2 and the definition of $\Omega(E)$, the fibers to be considered are the superspace S^{m+n}, in which E is modeled and $\overline{S} = \{\Sigma_{i=1}^{m+n} a_i e_i \in V \mid a_i \beta_{K_i} = 0$ for $m+1 \leqslant i \leqslant m+n\}$, and $S^* = L_B(S, B)$. (See Paper I.)

Definition 6.1: An L_B-bundle with base a ρ-supermanifold E and fiber a B-space F is a vector bundle $M(E, F, \pi)$,[3] such that there exists a trivializing covering $\{U_i, \tau_i\}$ satisfying the condition that for each $x \in U_i$, the transition functions $x \to \psi_{ij}(x) = \tau_{ix} \cdot \tau_{jx}^{-1}$ from $U_i \cap U_j$ into $c - L_B(F)^0$ are G^∞ $\forall i, j$. From now on we shall refer also to L_B-bundles as supervector bundles.

Given an L_B-bundle M and a body F_c for F, we can associate to them a real vector bundle M_c called the body of M, with base E_ρ, fiber F_c, and transition functions $(\psi_c)_{ij}$ determined by the diagram

$$\begin{tikzcd} U_i \cap U_j \arrow[r, "\psi"] \arrow[d, "\rho"] & c - GL_B(F)^0 \arrow[d, "\bar{c}"] \\ \rho(U_i) \cap \rho(U_j) \arrow[r, "\psi_c"] & GL(F_c). \end{tikzcd} \tag{10}$$

Now given a ρ-supermanifold E modeled on a superspace S, two tangent L_B-bundles with fibers S and $\overline{S}$, respectively, can be considered. We designate them by $T(E)$ and $\overline{T}(E)$. Given a ρ-atlas $\{U_i, \psi_i\}$ of E, the transition functions of the tangent bundles are given by ψ_{ij*}. Taking the body r for S and $\overline{S}$ where $r = t \cdot R$ and t is the projection over the first m coordinates, the body bundle (associated to r) of the two tangent bundles $T(E)$ and $\overline{T}(E)$ is the usual tangent bundle $T(E_\rho)$. Also a cotangent bundle $T^*(E)$ with fibers S^* is constructed with transition functions ψ^*_{ij}. The body of $T^*(E)$ associated to the body of $T^*(E)$ associated to the body r of S^* is the usual cotangent bundle $T^*(E_\rho)$.

The spaces of supervector fields $\chi(E)$, $\overline{\chi}(E)$, and of 1-superforms $\Omega(E)$, are the spaces of G^∞-sections of $T(E)$, $\overline{T}(E)$, and $T^*(E)$, respectively. In general it is not possible to construct an extension of the bundle $T(E)$ to an L_B-bundle with fiber V. The reason is that the transition functions $\psi_{\alpha\beta}(x)_*$ have many extensions from S to V (unless S is free) and the choice of one extension satisfying the cocycle condition is not possible. When S is free, $\psi_{\alpha\beta}(x)_*$ has a unique extension from S to V.

7. SUPERTENSOR FIELDS AND DIFFERENTIAL SUPERFORMS

In this section we shall restrict our discussion to the case of supertensors defined on free ρ-supermanifolds. In this case $\overline{S} = V$ and the study becomes less involved. The discussion of tensor fields for nonfree ρ-supermanifolds can be done similarly, but taking into account the peculiarities of tensor algebra on nonfree superspaces studied in Paper I.

Now we construct tensor bundles over a ρ-supermanifold as L_B-bundles. Given a ρ-supermanifold E we define the tensor bundle $T^e_k(E)$ as an L_B-bundle with base E, fiber $T^e_k(V)$, and transition functions induced by ψ_{ij*}. These transition functions obviously have their image in $r - L_B(T^e_k(V))$. Associated with the body r of $T^1_k(V)$ we have a body $T^1_k(E)_r$ of $T^1_k(E)$ which coincides with $T^1_k(E_\rho)$. This provides an *a posteriori* justification for choosing r as the body map for $T^1_k(V)$.

Supertensor fields of type $(1, k)$ are G^∞-sections of $T^1_k(E)$. A supertensor field T^1_k induces an ordinary tensor field $(T^1_k)_r$ on E_ρ called the body of T^1_k.

Differential superforms appear as usual as a particular case of supertensor fields. Differential superforms of degree p are the G^∞-sections of the L_B-bundle $\Lambda_p(E)$. We denote the space of differential superforms by $\Omega(E)$.

The exterior product in $\Omega(E)$ is defined through the exterior product in the fibers. The exterior differential d is defined as an even linear map $d: \Omega_p(e) \to \Omega_{p+1}(E)$ satisfying (i) $d(v \wedge w) = dv \wedge w + (-1)^p v \wedge dw, v \in \Omega_p(E)$, (ii) $d^2 = 0$, (iii) $df = \Sigma^{m+n}_{i=1} dx^i \partial f/\partial x^i$ in coordinates. Let us observe that d is L_B-linear and that the Grassmann degree does not enter in the sign $(-1)^p$ in (i) because d is even. For a p-form $\omega = f\, dx^{i_1} \wedge \cdots \wedge dx^{i_p}$ in coordinates $d^\omega = df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p}$, where $i_1 \leqslant \cdots \leqslant i_p$ and the sign $=$ is only possible if $r_{i_j} = 1$.

Differential superforms are obviously graded-skew-symmetric L_B-multilinear maps on supervector fields. As an example we give the exterior differential of a Λ-superform and the exterior product of two Λ-superforms.

$$
\begin{aligned}
(X_1, X_2) d\omega_1 &= X_1(X_2\omega_1) - (-1)^{(r_1+s_1)s_2} \\
&\quad \times X_2(X_1\omega_1) - ([X_1, X_2])\omega_1, \\
(X_1, X_2)\omega_1 \wedge \omega_2 &= (-1)^{r_2 s_1}(X_1\omega_1)(X_2\omega_2) \\
&\quad + (-1)^{1 + r_1(r_2 + s_1)}(X_2\omega_1)(X_1\omega_2),
\end{aligned}
\tag{11}
$$

where $X_i \in \overline{\chi}(E)^{r_i}$ and $\omega_i \in \Omega_1(E)^{s_i}$.

For a G^∞-function between two ρ-supermanifolds $f: M \to N$ the usual properties $f^*(\omega_1 \wedge \omega_2) = f^*\omega_1 \wedge f^*\omega_2$ and $f^* d\omega = df^*\omega$ holds.

8. SUPERGROUPS AND PRINCIPAL SUPERFIBER BUNDLES

We describe in this section how supergroups and principal superfiber bundles are included within the framework of ρ-supermanifolds. Throughout all the section we shall restrict discussion to the case of free ρ-supermanifolds.

A. Lie supergroups and lie superalgebras

Definition 8.1: A Lie supergroup G is a ρ-supermanifold with a group structure given by a composition law such that for $a, b, \in G$ the map $(a, b) \to ab^{-1}$ is G^∞.

As a consequence of Definition 8.1, the body G_ρ of G is a Lie group. The composition law for G_ρ is induced from that of G by $\rho(a)\rho(b) = \rho$ (ab). So the body map ρ is a group homomorphism $\rho: G \to G_\rho$.

Left and right translations L_a and R_a are G^∞-diffeomorphisms. Then, the space of left-invariant supervector fields $\overline{\mathscr{V}}$ is a Lie superalgebra over B, i.e., the Lie bracket is graded and L_B-bilinear. On the other hand, the space of left-invariant G^∞-vector fields is a B-Lie algebra, i.e., the Lie bracket involves only commutators and is L_B-bilinear whenever the multiplication by elements of B does not get out of $\mathscr{V}$.

There exists an L_B-isomorphism from $\overline{\mathscr{V}}$ on to $\overline{T}_e(G)$ that takes $\mathscr{V}$ onto $T_e(G)$. Then the B-Lie algebra $T_e(G)$ and the B-Lie superalgebra $\overline{T}_e(G)$ are B-spaces whose body is $\mathscr{V}_\rho$ which is the Lie algebra of G_ρ.

On the other hand, another body can be defined for $\overline{\mathscr{V}}$ in the following way. In the basis $\{X_1, \dots, X_{m+n}\}$ of $\overline{\mathscr{V}}$ the Lie superalgebra is determined by the structure constants $[X_i, X_j] = \pm c^k_{ij} X_k$. Take now $V = V^0 \oplus V^1$ to be a real graded vector space with $\dim V^0 = m$ and $\dim V^1 = n$.

Take a basis $\{Y_1, \dots, Y_{m+n}\}$ with $Y_1, \dots, Y_m \in V^0$ and $Y_{m+1}, \dots, Y_{m+n} \in V^1$. We define $R: \overline{\mathscr{V}} \to V$ by $R(\Sigma^{m+n}_{i=1} a^i X_i) = \Sigma^{m+n}_{i+1} r(a^i) Y_i$. Then V is furnished with a real Lie superalgebra structure by the bracket $[Y_i, Y_j] = r(c^k_{ij}) Y_k$. This provides an ordinary real graded Lie algebra[6] associated to the group G which we shall call $\mathscr{V}_R$. In addition we have that $\mathscr{V}_\rho$ is isomorphic to $\mathscr{V}^0_R$.

The elements of $\mathscr{V}$ give rise to real monoparametric Lie subgroups of G while for elements $X \in \overline{\mathscr{V}}$ such that $X \notin \mathscr{V}$, an integral curve does not exist. Hence, the exponential maps into G, but it is not defined on the whole $\overline{\mathscr{V}}$.

B. Representations of Lie supergroups and superalgebras

Definition 8.2: A matrix representation D of dimension (m, n) of a supergroup G is a G^∞-homomorphism D: $G \to GL_B(V)^0$, where V is a B-module of dimension (m,n).

From this definition it is immediate that for each free superspace modeled on V, $GL_B(S)$ admits a representation on V.

Given a matrix representation

$$\begin{pmatrix} D_{11} D_{12} \\ D_{21} D_{22} \end{pmatrix}(g)$$

of a supergroup G, $D_{11}(g)$ and $D_{22}(g)$ are $m\times m$ and $n\times n$ dimensional, respectively, and both have entries in B^0. On the other hand $D_{12}(g)$ and $D_{21}(g)$ are, respectively, $m\times n$ and $n\times m$ dimensional and both have entries in B^1.

Since D is G^∞,

$$\rho(g) = \rho(h) \Rightarrow D(g)_R = D(h)_R = \begin{pmatrix} R(D_{11}(g)) & 0 \\ 0 & R(D_{22}(g)) \end{pmatrix}.$$

So a representation D of G induces a reducible real $(m+n)$-dimensional representation of $\mathscr{V}_\rho$, $D_R: G_\rho \to GL(\mathbb{R}^{m+n})$ given by $D(\rho(g)) = D(g)_R$. Then D_R reduces to $D_R = D_\rho + D'_\rho$, where $D_\rho: G_\rho \to GL(\mathbb{R}^m)$, D'_ρ: $G_\rho \to GL(\mathbb{R}^n)$, and $D_\rho(\rho(g)) = R(D_{11}(g))$, $D'_\rho(\rho(g)) = R(D_{22}(g))$.

We remark that the matrix supergroups of Rittenberg and Scheunert[7] are examples of the G^∞-supergroups defined in 8.1, whose body is a direct product $G_1\times G_2$.

Given a representation D of a supergroup G we have, associated to it, representations for the "Lie" algebras $\overline{\mathscr{V}}$, $\mathscr{V}$, $\mathscr{V}_R$, and $\mathscr{V}_\rho$ associated with G, straightforwardly.

C. Principal superfiber bundles

Definition 8.3: A principal superfiber bundle P is a principal fiber bundle $P(M,G,\pi)$ such that P and M are ρ-supermanifolds, G is a Lie supergroup and π, the transition functions, and the trivializations are G^∞-functions.[8]

As a consequence of the definition, P_ρ is a principal fiber bundle $P_\rho(M_\rho,G_\rho,\pi_\rho)$. The L_B-bundles defined in Sec. 6 with base E and fiber a B-space F, are associated to a principal superfiber bundle $P(E,GL_B(F))$.

Associated to an element $A \in \mathscr{V}$, we have a fundamental G^∞-vector field A^* constructed by means of the monoparametric subgroup of A. To deal with connections and curvature forms, it is convenient to extend the fundamental fields to elements of $\overline{\mathscr{V}}$. To do that, we look at $A^*_u = \sigma_{u*e}(A)$ where $\sigma_u: G \to P$ is given by $\sigma_u(g) = ug$. The last expression for A^*_u makes sense also for elements of $\overline{\mathscr{V}}$, and $A^*_u \in \overline{T}_u(P)$.

A superconnection form on a principal superfiber bundle P is defined as an element ω of $\Omega^1_{\overline{\mathscr{V}}}(P)^0$, i.e., an even $\overline{\mathscr{V}}$-valued superform on P satisfying the conditions $R^*_a\omega = \mathrm{ad}(a^{-1})\,\omega\ \forall\, a \in G$ and $A^*\omega = A\ \forall a \in \overline{\mathscr{V}}$.

The vertical part of $\overline{T}_u(P)$ is determined, as usual, as the tangent part to the fiber $\pi^{-1}(U)$ on U, $(\overline{T}(\pi^{-1}(U)) = \overline{\mathscr{V}}_u$ and σ_{u*e} gives a B-isomorphism from $\overline{\mathscr{V}}$ into $\overline{\mathscr{V}}_u$. The horizontal part for ω is $h(\overline{T}_u(P)) = \mathrm{Ker}\,\omega_u$, and $\overline{T}_u(P) = h\overline{T}_u(P) + \overline{G}_u$. So h is a well-defined G^∞-projection.

The covariant differential D is defined by applying the exterior differential to the horizontal part of the supervector fields. So the curvature $\Omega(X,Y) = D\omega(X,Y) = d\omega(hX, hY)$ is a G^∞-$\overline{\mathscr{V}}$-valued two-superform. The structure equation $\Omega = d\omega + [\omega, \omega]$ and the Bianchi identity $d\Omega = 0$ hold.

9. CONCLUSION

We have constructed in this paper the basic objects super manifolds, supergroups, supervector bundles, and principal superfiber bundles necessary to provide a mathematical model suited to propose and solve geometric problems in spaces with commuting and anticommuting coordinates. We hope that the present study will be of worth in discussing the geometrical structure of supersymmetric gauge theories and supergravity. In fact, using this model a geometrical interpretation for gauge theories with Faddeev–Popov fields has been proposed using a principal superfiber bundle structure.[9]

The central role in our construction is played by the so-called ρ-supermanifolds which are nothing but C^∞-Banach manifolds[3] with a superimposed Grassmann structure and related to an ordinary C^∞-manifold via the projection ρ. This point of view allows one to translate most constructions of ordinary differential geometry into differential supergeometry almost directly, in contrast to Kostant's approach.[10]

One of the results of this work has been precisely to characterize the class of ρ-supermanifolds or supermanifolds with a well-behaved body, which are very similar to de Witt's proposal,[4] but for the fact that ρ-supermanifolds are Haussdorff and allow nontrivial topologies in the anticommuting sector. On the other hand, the existence of the Grassmann structure and of the projection ρ on a finite-dimensional real C^∞-manifold allow one to handle ρ-supermanifolds as if they were finite dimensional in many cases. In particular, the existence of partitions of unity can be proven.

Concerning tangent bundles to ρ-supermanifolds we have found that apart from the ordinary tangent bundle, a supertangent bundle can be defined in a natural way which is a B-module. The supertangent space in a point turns out to be reflexive and permits one to construct easily differential superforms and supertensors on ρ-supermanifolds.

Lie supergroups appear naturally in our model and generalize the matrix supergroups of Ref. 7. Also a precise relation emerges between the "Lie algebra" of a supergroup understood as the set of infinitesimal operations and the real Lie superalgebras treated in the literature.[6] In addition some relations between the representations of a supergroup and those of its body have been found. Finally the elements of the theory of principal superfiber bundles and associated L_B-bundles have been established.

[1]J. Hoyos, M. Quirós, J. Ramirez Mittelbrunn, and F. J. de Urríes, "Generalized Supermanifolds I," IEM preprint, July 1981; "Generalized Supermanifolds II," IEM preprint, August 1981.
[2]A. Rogers, J. Math. Phys. 21, 1352 (1980).
[3]S. Lang, *Differential Manifolds*, (Addison Wesley, Reading, MA, 1972).
[4]B. S. de Witt, "Supermanifolds" (unpublished).

[5]P. A. Marchetti and R. Percacci, "Categorical properties of supermanifolds," Preprint ISAS Trieste 3/81.
[6]M. Scheunert, *The Theory of Lie Superalgebras, Lecture Notes in Mathematics*, Vol. 716 (Springer-Verlag, Berlin, 1979).
[7]I. Rittenberg and M. Scheunert, J. Math. Phys. **19**, 709 (1978).
[8]S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Interscience, New York, 1963), Vol. I.
[9]J. Hoyos, M. Quirós, J. Ramirez Mittelbrunn, and F. J. de Urríes, "Superfiber bundle structure of gauge theories with Faddeev–Popov fields," J. Math. Phys. **23**, 1504 (1982); J. Hoyos, M. Quirós, J. Ramírez Mittelbrunn, and F. J. de Urríes, "Superfield Formalism of Gauge Theories with Ghost and Gauge Fixing Terms from Dynamics on Superspace," Nucl. Phys. B **218**, 159 (1983).
[10]B. Kostant in *Differential Geometrical Methods in Mathematical Physics*, Bonn, 1975 (Springer, Berlin, 1977).

Symmetries and stable periodic orbits for one-dimensional maps

R. Vilela Mendes
CFMC, Instituto Nacional de Investigação Científica Av. Gama Pinto, 2, 1699 Lisboa Codex, Portugal

(Received 6 May 1983; accepted for publication 10 June 1983)

Global $\phi^{-1} \circ f \circ \phi = f$ and infinitesimal $\alpha \circ f = Df(\alpha)$ symmetries are considered for dynamical maps f. The general solution of $\alpha \circ f(x) = f'(x)\alpha(x)$; $x \in [a,b] \subset \mathbb{R}$ is constructed, and for S-unimodal maps, the existence of a nontrivial continuous solution is shown to be equivalent to the existence of a stable periodic orbit.

PACS numbers: 02.40. + m, 02.20. + b, 03.20. + i

I. SYMMETRIES IN DISCRETE DYNAMICAL SYSTEMS

Symmetries play an important role in the study of dynamical systems defined by flows of vector fields. For Hamiltonian systems that remain invariant under the action of a group, the moment map construction[1] unifies the derivation of constants of motion. These, in turn, allow the elimination of a number of variables and the construction of a reduced phase space for the system.[2] Also, in bifurcation problems, symmetry considerations are particularly useful in the derivation of the branching equations.[3]

Although various genericity results already exist for equivariant diffeomorphisms,[4,5] symmetries do not seem to have played as important a role in the study of discrete dynamical dynamics as they have for flows.

By a *discrete dynamical system*, we mean a pair (M,f) where M is a differentiable manifold and f a smooth map. Mostly we will think of M as an open set in Euclidean space R^n. A discrete dynamical system will have a *symmetry* if f is equivariant for the action of a diffeomorphism ϕ, i.e.,

$$\phi^{-1} \circ f \circ \phi = f. \tag{1.1}$$

Now let f be equivariant for the action of a local one-parameter group ϕ_t. Let $\alpha = d\phi_t/dt\,|_{t=0}$ be the corresponding Lie algebra element. Then, differentiation of (1.1) yields

$$\alpha \circ f = Df(\alpha) \tag{1.2}$$

$\left[\text{in local Euclidean coordinates } \alpha^i(f(x)) = \alpha^j(x)\dfrac{\partial}{\partial x^j} f^i(x)\right]$.

Notice that Eq. (1.2) may, in principle, have a solution even when there is no corresponding one-parameter group of diffeomorphisms. When (1.2) has a nontrivial continuous solution, the map f will be said to have an *infinitesimal symmetry*.

As in the "continuous-time" systems, one may, in some cases, construct constants of motion from the knowledge of the symmetries. The following result is due to J. T. Duarte.[6]

Lemma: Let f be a $R^n \to R^n$ map of constant Jacobian. If f has a symmetry ϕ, then det $D\phi$ is constant on the orbits of f.

Proof: From $\phi \circ f = f \circ \phi$, differentiating, and computing the determinant, one obtains

$$\det D\phi\,(f(x)) \det Df(x) = \det Df(\phi(x)) \det D\phi\,(x).$$

Because det Df = const, the result follows:

$$\det D\phi\,(f(x)) = \det D\phi\,(x). \qquad \blacksquare$$

Notice that the same result holds if, instead of $\phi^{-1} \circ f \circ \phi = f$, one has $\phi^{-1} \circ f \circ \phi = f \circ h$ with det Df = const and det $Dh = 1$.

The main problem I will be concerned with in this paper is the relation between the existence of symmetries and the dynamical properties of maps. Here, I will restrict this study to one-dimensional maps. For them, a relation is established between nontrivial symmetries and the existence of stable periodic orbits.

II. SYMMETRIES AND STABLE PERIODIC ORBITS IN ONE-DIMENSIONAL MAPS

Let us examine the question of the existence of infinitesimal symmetries. For this purpose, one seeks the general solution $\alpha(x)$ to the functional equation

$$\alpha \circ f(x) = f'(x)\alpha(x), \tag{2.1}$$

where f is a differentiable function defined in an interval $[a,b]$ with a finite number of critical points in that interval. The method used is inspired by a technique of Kuczma[7] (see, in particular, Theorem 1.3, Chap. I, p. 40). However, the particular form of Eq. (2.1) allows an important simplification in the construction of the solution, namely, one avoids entirely explicit reference to the family f_λ^{-1} of inverses of f. Furthermore, it is also possible to define the solution in all the interval, whereas Theorem 1.3 of Kuczma gives it only in a submodulus set E $(f(E) \subset E)$.

Given a critical point x_i^c, let E_i^c be the set of antecedents of x_i^c plus the critical point itself,

$$E_i^c = \{x: \exists k \geqslant 0,\ f^k(x) = x_i^c\}.$$

Each E_i^c may contain other critical points besides x_i^c. We call *ground-level critical point* one for which the set E_i^c contains no other critical points [i.e., $x \in E_i^c$ and $x \neq x_i^c \Rightarrow f'(x) \neq 0$].

Z-orbit is the equivalence class defined by $x \sim y$ iff $\exists\ m,n: f^m(x) = f^n(y)$, $m,n \geqslant 0$. This should not be confused with the notion of *orbit* of x, i.e., $\{y: \exists\ k \geqslant 0, f^k(x) = y\}$. The set $E = [a,b] - U_i E_i^c$ will play an important role: In $[a,b]$ we consider the classes C_k, $k \geqslant 0$, $x \in C_k$, $k \geqslant 1$ iff $\exists\ j: f^{j+k}(x) = f^j(x)$, j being the smallest integer for which the equality holds, $x \in C_0$ otherwise. $U_{K>1} C_k$ is the set of points that either belong to a periodic orbit or fall into one.

We now define the class of functions which will be used to construct the general solution to the functional equation.

A real-valued function ϕ is said to belong to the class $\psi[A]$ when the following is true.

(a) ϕ is defined in a set $A \subset [a,b]$.

(b) $\phi(x) = 0$ if $x \in C_k$ ($k \neq 0$) and $f^{(i+k)'}(x) \neq f^{i'}(x)$, i.e., the periodic orbit where x falls has a multiplier different from unit ($f^{k'} \neq 1$).

(c) $\phi(x) = 0$ if $x \in E$ and $\exists\ m > 0$, $n \geqslant 0$, and x_i^c:$f^m(x_i^c) = f^n(x)$, i.e., $x \in E$ and is in the Z-orbit of a critical point.

(d) $\phi(x) = 0$ if $x \in U_i E_i^c$ and x is not a ground level critical point or one of its antecedents.

After these preliminaries we can state the following theorem.

Theorem 2.1: Let A be a set which has exactly one point from each Z-orbit contained in $E = [a,b] - U_i E_i^c$ plus all ground level critical points. Then, for every function $\alpha_0(x)$ belonging to the class $\psi[A]$, there exists exactly one function $\alpha(x)$ defined in the interval $[a,b]$ that satisfies the functional Eq. (2.1) and such that $\alpha(x) = \alpha_0(x)$ for $x \in A$. This function is given by the following.

(a) For $x \in E$,

$$\alpha(x) = \alpha_0[a(x)]\, f^{n_x'}[a(x)]/f^{m_x'}(x), \tag{2.2a}$$

where $a(x)$ is the point in A which belongs to the Z-orbit of x, and n_x and m_x are chosen such that

$$f^{n_x}[a(x)] = f^{m_x}(x).$$

(b) If $x \in U_i E_i^c$ and x is not a ground-level critical point or one of its antecedents, then

$$\alpha(x) = 0. \tag{2.2b}$$

(c) If x is an antecedent of a ground-level critical point x_i^c,

$$\alpha(x) = \alpha_0(x_i^c)/f^{m_x'}(x) \tag{2.2c}$$

with $f^{m_x}(x) = x_i^c$.

Proof: The existence of the set A follows from the axiom of choice.

It is easy to check that Eqs. (2.2) are solutions to (2.1) under each of the stated conditions. For example, for (2.2a), because $a(f(x)) = a(x)$,

$$\alpha \circ f(x) = \alpha_0[a(x)]\, \frac{f^{n_x'}[a(x)]}{f^{(m_x-1)'}[f(x)]} = f'(x)\alpha(x).$$

The choice of $\alpha_0(x)$ in the class $\psi[A]$ guarantees that

$$(f^{(j+k)'}(x) - f^{j'}(x))\alpha(x) = 0$$

holds whenever $x \in C_k$ as required by (2.1). Finally, to prove unicity, let $z(x)$ be another function satisfying (2.1) and such that $z(x) = \alpha_0(x)$, $x \in A$. Then, from $f^{n_x}[a(x)] = f^{m_x}(x)$ and $z[a(x)] = \alpha_0[a(x)]$,

$$z(f^{n_x}[a(x)]) = f^{n_x'}[a(x)]z[a(x)] = z(f^{m_x}(x)) = f^{m_x'}(x)z(x),$$

which implies

$$z(x) = \alpha_0[a(x)]\, f^{n_x'}[a(x)]/f^{m_x'}(x) = \alpha(x). \qquad \blacksquare$$

Formulas (2.2), defining a solution to the functional equation (2.1), parametrized by an arbitrary function of class $\psi[A]$, are too general to be of immediate use in the characterization of the dynamics. This arises from the fact that no practical prescription is given to find the set A, and no restrictions are imposed on $\alpha_0(x)$ besides being of class $\psi[A]$. A useful restriction is continuity of the solution. This motivated our definition of an *infinitesimal symmetry* $\alpha(x)$ of f to be a continuous solution to the functional equation (2.1) in $[a,b]$.

For the next result, I will restrict myself to the class of symmetric S-unimodal maps. A one-dimensional map f of $[-1,1]$ into itself is symmetric S-unimodal if

(1) $f(0) = 1$,
(2) $f(x) = f(-x)$,
(3) f is C^3,
(4) The Schwartzian derivative $Sf(x) < 0$, $x \neq 0$,
(5) $f'(0) = 0$ and $f'(x) \neq 0$, $x \neq 0$,
(6) f is strictly increasing in $[-1,0)$.

For this class of maps we prove the following theorem.

Theorem 2.2: For f symmetric S-unimodal, there is a continuous nontrivial solution to the functional equation (2.1) if and only if f has a stable periodic orbit.

Proof: Symmetry of f implies $\alpha(x) = -\alpha(-x)$, and continuity at $x = 0$ leads to $\alpha(0) = 0$. The functional equation then implies $\alpha = 0$ for all the antecedents of the critical point, i.e., $\alpha(x) = 0$, $x \in E^c$. If f is S-unimodal and has no stable periodic orbit, then the set of antecedents of 0 is dense in $[-1,1]$.[8] By continuity, $\alpha = 0$ in the whole interval. Therefore the existence of a stable periodic orbit is a necessary condition for a nontrivial infinitesimal symmetry.

When there is a stable periodic orbit, one proves the existence of continuous solutions by showing that the set A of Theorem 2.1 may be chosen to contain one or two intervals, where $\alpha_0(x)$ is defined as a continuous function.

Let the stable periodic orbit have period n. Then there are n fixed points of f^n, where $|f^{n'}| \leqslant 1$. A choice of semi-open B-intervals is made which is better described by Fig. 1, where

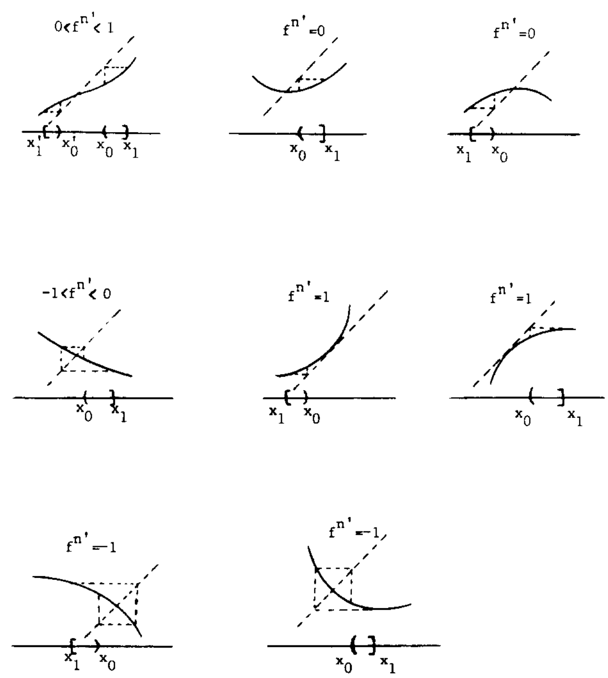


FIG. 1: B-intervals for the proof of Theorem 2.2.

the several possibilities for the behavior of f^n near the fixed points are illustrated.

It should be further specified that all B-intervals are chosen in a small enough neighborhood of one of the fixed points of f^n such that $f^k B \cap B = \emptyset, k = 1,...,n-1$. Also, by construction, $f^n B \cap B = \emptyset$.

Any two points $x,y \in B$, $x \neq y$ belong to distinct Z-orbits. Any point that converges to the stable periodic orbit is either in the Z-orbit of a point in B or in the Z-orbit of the periodic orbit.

Let A be the union of the B-interval(s) and a point ξ of the periodic orbit. For S-unimodal maps, it is known that the Lebesgue measure of those points which do not converge to the stable periodic orbit is zero. Therefore A contains one point from each Z-orbit of almost all points in the interval.

We now define a nonzero function $\alpha_0(x)$ in A such that

(1) $\alpha_0(x)$ is continuous in B,

(2) $\alpha_0(\xi) = 0$,

(3) $\lim_{x \to x_0} \alpha_0(x) = \alpha_0(x_1) = \lim_{x \to x_0'} \alpha_0(x) = \alpha_0(x_1') = 0$,

(4) $\alpha_0(x)$ is of class $\psi[A]$,

(5) $\alpha_0(a(x_c)) = 0$, where $a(x_c) \in A$ is the point of A that is in the Z-orbit of the critical point.

Continuity of the solution $\alpha(x)$ now follows from the continuity of algebraic operations of continuous functions and Eqs. (2.2) of Theorem 2.1. Let y be a point that converges to the stable periodic orbit such that $a(y)$ [i.e., the point in A that belongs to the Z-orbit of y] falls in the interior of B. Then, from Eq. (2.2)

$$\alpha(y) = \alpha_0[a(y)]/f^{m_y'}(y), \tag{2.3}$$

with $a(y) = f^{m_y}(y)$.

Suppose now that for every $a \in B$, there is a point z_a in the neighborhood of y such that $a = f^{m_y}(z_a)$. In this case, by using (2.3) for all points $a \in B$, one constructs a continuous solution in an interval containing y. Furthermore, because of our condition (3) on $\alpha_0(x)$, the solution will vanish at the endpoints of the interval.

If, however, there are some points $a \in B$ which have no antecedent of order m_y in the neighborhood of y, it means that near y, there is an antecedent β of the critical point of order less than m_y. In this case, one uses (2.3) to construct the solution in an interval around y that has β as an endpoint. Again, the solution vanishes at the endpoints because they are either antecedents of the endpoints of B or of a critical point.

Now we pick another point that tends to the stable orbit where the solution is not yet defined and repeat the construction. Repeating the procedure until all points that tend to the stable orbit are exhausted, it follows from unicity and the conditions on $\alpha_0(x)$ that a continuous solution is defined in disjoint intervals vanishing at their endpoints. For the set of points outside the intervals which do not tend to the stable periodic orbit, the solution is defined to be zero. The vanishing of the solution at the endpoints of the intervals insures continuity in the whole interval $[-1,1]$. ∎

Using methods of iteration theory, the existence of continuous solutions to a functional equation was shown to be equivalent to the existence of a stable periodic orbit. This result provides an analytical handle on a dynamical problem because, in some cases, the existence of solutions to a functional equation can be decided purely by methods of analysis.

To close, I will add some remarks concerning directions of future research in this problem.

First, it would be interesting to know whether the relation between existence of symmetries and characteristic properties of the orbits can be extended to higher-dimension maps.

Once the relation between dynamical properties of maps and solutions of functional equations is established, such a relation becomes useful if, by analytical methods, one can decide about the existence or nonexistence of solutions with the required properties.

At least for one-dimensional maps, this could be achieved by writing the solution as a series in orthogonal functions $G(x)$ in $[-1,1]$:

$$\alpha(x) = \sum \alpha_n G_n(x). \tag{2.4}$$

By substitution in the functional equation (2.1), this one is converted into an (infinite) set of algebraic relations. If $\max_{[-1,1]}|G_n(x)|$ is finite, for $\alpha(x)$ to be continuous, it suffices that the solution $\alpha = \{\alpha_n\} \in l^1$.

For $f(x) = 1 - \mu x^2$, for example, if $\{G_n(x)\}$ is the set of Chebyshev polynomials, the proof of aperiodicity of the $\mu = 2$ point becomes a trivial calculation because the only solution is $\alpha_n = 0$.

With other orthogonal sets of appropriate weight, it is, in principle, possible to decide by purely analytical methods on the existence of stable periodic orbits for other μ values.

ACKNOWLEDGMENTS

The author would like to thank J. Taborda Duarte and A. J. B. Lopes Pinto for stimulating discussions.

[1] J. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).
[2] J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[3] D. H. Sattinger, *Group Theoretic Methods in Bifurcation Theory, Lecture Notes in Math. No. 762* (Springer, Berlin, 1970); and references therein.
[4] M. Field, Bull. Am. Math. Soc. **76**, 1314 (1970).
[5] M. Field, Astérisque **40**, 67 (1976).
[6] J. T. Duarte, private communication.
[7] M. Kuczma, *Functional Equations in a Single Variable* (PWN-Polish Scientific Publishers, Warszawa, 1968).
[8] J. Guckenheimer, Commun. Math. Phys. **70**, 133 (1979).

Fermion diffusions

D. Applebaum[a)] and R. L. Hudson
Mathematics Department, University of Nottingham, Nottingham, England

(Received 8 February 1983; accepted for publication 23 June 1983)

We determine the form and solutions of all possible stochastic differential equations driven by fermion Brownian motion for a fermion system of one degree of freedom.

PACS numbers: 02.50.Ey, 05.40. + j, 03.65.Db

1. INTRODUCTION

Much work has been done recently[1–3] on the construction and solution of stochastic differential equations of the form

$$da = dA^{\dagger} F + G\, dA + H\, dt, \tag{1.1}$$

generalizing equations for "classical" diffusions of the type

$$dY = \sigma(Y)\, dQ + m(Y)\, dt, \tag{1.2}$$

where Q is classical Brownian motion.

In Refs. 1 and 2 the process $(a_t;\, t\in\mathbb{R}^+)$ describes the evolution of a boson system of one degree of freedom obeying the commutation relation

$$[a_t, a_t^{\dagger}] = I \tag{1.3}$$

(wherein $a_t^{\dagger}$ is the adjoint of a_t), and the family of pairs $(A_t, A_t^{\dagger})$ represent a quantum-mechanical Wiener process of boson type.[4]

Since the right-hand side of the commutation relation (1.3) is time-independent, the process (a_t) may be regarded as a quantum stochastic process in the sense of Ref. 5.

In this paper, we seek to extend these ideas to fermions. In this connection it should be noted that a theory of stochastic integrals against fermion Brownian motion has recently been developed.[3]

2. STRUCTURE OF FERMION DIFFUSIONS

A fermion system with one degree of freedom is described by a pair of operators $(a^{(0)}, a^{(0)\dagger})$ acting in a Hilbert space $\mathscr{h}_0$ and satisfying the relations

$$\{a^{(0)}, a^{(0)\dagger}\} = I, \quad \{a^{(0)}, a^{(0)}\} = \{a^{(0)\dagger}, a^{(0)\dagger}\} = 0, \tag{2.1}$$

where $\{\ ,\ \}$ is the anticommutator

$$\{A, B\} = AB + BA.$$

An example of such a pair is given by the matrices

$$a^{(0)} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad a^{(0)\dagger} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \tag{2.2}$$

acting in the Hilbert space $\mathbb{C}^2$.

The fermion analog[6] of the von Neumann uniqueness theorem states that any pair of operators satisfying (2.1) is unitarily equivalent to a direct sum of copies of the pair given by (2.2) and in particular, any pair which acts irreducibly is unitarily equivalent to $(a^{(0)}, a^{(0)\dagger})$.

Let $\Gamma(L^2(\mathbb{R}^+))$ be fermion Fock space over $L^2(\mathbb{R}^+)$, and for $f, g\in L^2(\mathbb{R}^+)$ let $a(f)$, $a^{\dagger}(g)$ be annihilation and creation operators (respectively) in $\Gamma(L^2(\mathbb{R}^+))$. By direct analogy with the boson case,[2,4,7] *fermion Brownian motion* is described by the family of pairs $(A_t, A_t^{\dagger})$ where for $t\in\mathbb{R}^+$,

$$A_t = a(\chi[0,t]), \quad A_t^{\dagger} = a^{\dagger}(\chi[0,t]),$$

and expectations are determined in the vacuum state, $\Omega(\mathscr{h}) = (1, 0, 0, \ldots)$. (In Ref. 3 the field $Q_t = A_t + A_t^{\dagger}$ plays the role of Brownian motion.) We consider a quantum stochastic process generated by pairs $(a_t, a_t^{\dagger})$ for $t\in\mathbb{R}^+$, which satisfy the fixed time relations

$$\{a_t, a_t^{\dagger}\} = I, \quad \{a_t, a_t\} = \{a_t^{\dagger}, a_t^{\dagger}\} = 0 \tag{2.3}$$

subject to the differential equation

$$\begin{aligned} da_t &= dA_t^{\dagger} F_t + G_t dA_t + H_t dt, \\ da_t^{\dagger} &= dA_t^{\dagger} G_t^{\dagger} + F_t^{\dagger} dA_t + H_t^{\dagger} dt, \end{aligned} \tag{2.4}$$

where each of the a_t's are operators in

$$\mathscr{h}_0 \otimes (L^2(\mathbb{R}^+)),$$

$\mathscr{h}_0$ being some "initial" Hilbert space. For convenience, we choose $\mathscr{h}_0$ to be $\mathbb{C}^2$ and let

$$a_0 = a^{(0)} \hat{\otimes} I, \quad a_0^{\dagger} = a^{(0)\dagger} \hat{\otimes} I,$$

where the precise meaning of these operators will be given in the next section.

We further select $F^{(0)}$, $G^{(0)}$, and $H^{(0)}$ to be operators in $\mathbb{C}^2$ and we observe that by the fermion uniqueness theorem, a_t and $a_t^{\dagger}$ are unitarily equivalent to a direct sum of copies of $a^{(0)}$ and $a^{(0)\dagger}$. We define F_t, G_t, and H_t to be the pre-images of $F^{(0)}$, $G^{(0)}$, and $H^{(0)}$ under this unitary equivalence of the corresponding direct sum of copies of $F^{(0)}$, $G^{(0)}$, and $H^{(0)}$.

3. $\mathbb{Z}_2$-GRADED HILBERT SPACE

By a $\mathbb{Z}_2$-*graded Hilbert space*, we mean a Hilbert space direct sum

$$\mathscr{h} = \mathscr{h}_+ \oplus \mathscr{h}_-,$$

where $\mathscr{h}_+$, $\mathscr{h}_-$ are called the *even* and *odd* subspaces, respectively.

An operator $T\in B(\mathscr{h})$ is said to be

even if $T\mathscr{h}_+ \subseteq \mathscr{h}_+$ and $T\mathscr{h}_- \subseteq \mathscr{h}_-$

and

odd if $T\mathscr{h}_+ \subseteq \mathscr{h}_-$ and $T\mathscr{h}_- \subseteq \mathscr{h}_+$.

We note that $B(\mathscr{h})$ is a $\mathbb{Z}_2$-graded algebra in the sense of Chevalley.[8] The following rules hold for $\mathbb{Z}_2$ grading the products

[a)] Work carried out while this author was supported by the Science and Engineering Research Council.

of operators of definite parity:

"odd $\times$ odd = even $\times$ even = even,"

"odd $\times$ even = odd."

Let $\mathscr{h}_1$, $\mathscr{h}_2$ be $\mathbb{Z}_2$-graded Hilbert spaces and let T_1, T_2 be operators of definite parity in $B(\mathscr{h}_1)$, $B(\mathscr{h}_2)$, respectively. Then if $f_1 \in \mathscr{h}_1$, $f_2 \in \mathscr{h}_2$ are each either odd or even, we define $T_1 \hat{\otimes} T_2$ by

$$(T_1 \hat{\otimes} T_2)(f_1 \otimes f_2) = \mp T_1 f_1 \otimes T_2 f_2, \tag{3.1}$$

where the $-$ sign is taken only if T_2 and f_1 are both odd. The definition extends to arbitrary T_1, T_2 by bilinearity and to a bounded operator on $\mathscr{h}_1 \otimes \mathscr{h}_2$ by continuous, linear extension.

If $S_1 \hat{\otimes} S_2$ is another such operator, we obtain

$$(S_1 \hat{\otimes} S_2)(T_1 \hat{\otimes} T_2) = \mp S_1 T_1 \hat{\otimes} S_2 T_2, \tag{3.2}$$

where the $-$ sign is taken only if S_2 and T_1 are both odd.

Fermion Fock space can be $\mathbb{Z}_2$ graded as follows:

$$[\Gamma(\mathscr{h})]_+ = \bigoplus_{n=0}^{\infty} \mathscr{h}^{(2n)}, \quad [\Gamma(\mathscr{h})]_- = \bigoplus_{n=0}^{\infty} \mathscr{h}^{(2n+1)},$$

where $\mathscr{h}^{(n)}$ is the antisymmetric tensor product of n copies of $\mathscr{h}$. Annihilation and creation operators are thus seen to be odd.

We also $\mathbb{Z}_2$ grade the Hilbert space $\mathscr{h}_0$ by taking $(\mathscr{h}_0)_\mp$ to be the spans of the vectors $\Omega_\mp$, respectively, where Ω_+ is the vacuum $\begin{pmatrix}1\\0\end{pmatrix}$, and Ω_- the antivacuum $\begin{pmatrix}0\\1\end{pmatrix}$. The operators $a^{(0)}$, $a^{(0)\dagger}$ are clearly odd.

We shall require that the operators a_t, $a_t^\dagger$ of (2.3) are odd throughout their evolution.

We recall that $\Gamma(L^2(\mathbb{R}^+))$ is naturally isomorphic to the tensor product $\Gamma[L^2[0,t[) \otimes \Gamma(L^2[t,\infty[)$. A more correct notation for a_t, $a_t^\dagger$ recognizes them to be operators in $\mathscr{h}_0 \otimes \Gamma(L^2(\mathbb{R}^+))$ of the form

$$a_t^\dagger \hat{\otimes} I_{\Gamma(L^2[t,\infty[)}, \quad a_t^\dagger \hat{\otimes} I_{\Gamma(L^2[t,\infty[)}.$$

Similarly, the differentials dA_t, $dA_t^\dagger$ are correctly written

$$I_{\mathscr{h}_0 \otimes \Gamma(L^2([0,t[))} \hat{\otimes} dA_t, \quad I_{\mathscr{h}_0 \otimes \Lambda(L^2([0,t[))} \hat{\otimes} dA_t^\dagger.$$

Thus, we obtain the conditions

$$\{a_t, dA_t\} = \{a_t, dA_t^\dagger\} = \{a_t^\dagger, dA_t\} = \{a_t^\dagger, dA_t^\dagger\} = 0. \tag{3.3}$$

Furthermore, we see from (2.4) that if a_t is to be odd, then F_t and G_t must be even and H_t odd throughout the evolution.

In the remainder of this paper, we will formally drop the index t from our operators except where it is essential for clarity of understanding.

4. CONSISTENCY EQUATIONS

Before solving (2.4), we obtain conditions on the coefficients F, G, and H. We evaluate products of differentials by bilinear extension of the rules

$$0 = (dA)^2 = (dA^\dagger)^2 = dA^\dagger dA, \quad dA\, dA^\dagger = dt. \tag{4.1}$$

These may be deduced formally from the fact that A_t acts on the total family of vectors

$$\{a^\dagger(f_n)\cdots a^\dagger(f_1)\Omega : n \in \mathbb{N}, f_1, \ldots, f_n \in L^2(\mathbb{R}^+)\} \tag{4.2}$$

as

$$A_t a^\dagger(f_n)\cdots a^\dagger(f_1)\Omega = \sum_{j=1}^{n} (-1)^{n-j} \int_0^t f_j(\tau)d\tau \left(\prod_{\substack{k=1\\k\neq j}}^{n} a^\dagger(f_k)\right)\Omega, \tag{4.3}$$

suggesting the formal relation

$$dA_t a^\dagger(f_n)\cdots a^\dagger(f_1)\Omega = dt \sum_{j=1}^{n} (-1)^{n-j} f_j(t) \left(\prod_{\substack{k=1\\k\neq j}}^{n} a^\dagger(f_k)\right)\Omega, \tag{4.4}$$

whence $(dA_t)^2 = (dA_t^\dagger)^2 = dA_t^\dagger dA_t = 0$, since the matrix elements between vectors of the form (4.2) involve $(dt)^2 = 0$, whereas by the anticommutation relations

$$dA_t\, dA_t^\dagger = -dA_t dA_t^\dagger + dt = dt$$

(see Ref. 1 for the boson case).

Differentiating the relation $\{a, a^\dagger\} = I$, we obtain

$$\begin{aligned} 0 &= d\{a, a^\dagger\} \\ &= \{da, a^\dagger\} + \{a, da^\dagger\} + \{da, da^\dagger\} \\ &= \{dA^\dagger F + G\,dA + H\,dt, a\} \\ &\quad + \{a, dA^\dagger G^\dagger + F^\dagger dA + H^\dagger dt\} \\ &\quad + \{dA^\dagger F + G\,dA + H\,dt, dA^\dagger G \\ &\quad + F^\dagger dA + H^\dagger dt\}, \end{aligned} \tag{4.5}$$

and differentiating $\{a, a\} = 0$ yields

$$\begin{aligned} 0 &= \{da, a\} + \{a, da\} + \{da, da\} \\ &= 2\{dA^\dagger F + G\,dA + H\,dt, a\} \\ &\quad + \{dA^\dagger F + G\,dA + H\,dt, dA^\dagger F \\ &\quad + G\,dA + H\,dt\}. \end{aligned} \tag{4.6}$$

To simplify (4.5) and (4.6), we require the following result:

Lemma 4.1:

(i) $\{X, Z\} = 0 \Rightarrow \{XY, Z\} = X[Y, Z]$,

(ii) $\{Y, Z\} = 0 \Rightarrow \{XY, Z\} = -[X, Z]Y$.

Proof: (i)

$$\begin{aligned} \{XY, Z\} &= XYZ + ZXY \\ &= XYZ - XZY = X[Y, Z]. \end{aligned}$$

(ii) follows similarly. ∎

Application of Lemma 4.1 to (4.5) and (4.6) gives

$$\begin{aligned} 0 &= dA^\dagger([F, a^\dagger] + [G^\dagger, a]) \\ &\quad + ([a, F^\dagger] + [a^\dagger, G])dA \\ &\quad + (F^\dagger F + GG^\dagger + \{H, a^\dagger\} + \{a, H^\dagger\})dt \end{aligned}$$

and

$$0 = dA^\dagger [F, a] - [G, a]dA + (\{H, a\} + GF)dt.$$

Hence we obtain these consistency equations:

$$[F, a^\dagger] + [G^\dagger, a] = 0, \tag{4.7}$$

$$\{H, a^\dagger\} + \{a, H^\dagger\} = -(F^\dagger F + GG^\dagger), \tag{4.8}$$

$$[F, a] = 0, \tag{4.9}$$

$$[G, a] = 0, \tag{4.10}$$

$$\{H, a\} = -GF. \tag{4.11}$$

5. SOLUTIONS OF THE CONSISTENCY EQUATIONS

We find all possible coefficients F, G, and H satisfying the conditions (4.7)–(4.11). We recall that F and G are even and hence their most general form is given by

$$F = \lambda I + \kappa a^\dagger a, \quad \text{where } \lambda, \kappa, \mu, \nu \in \mathbb{C},$$

$$G = \mu I + \nu a^\dagger a.$$

From (4.7) we obtain

$$0 = [\lambda I + \kappa a^\dagger a, a^\dagger] + [\bar{\mu} I + \bar{\nu} a^\dagger a, a]$$
$$= \kappa a^\dagger + \bar{\nu} a \Rightarrow \kappa = \nu = 0,$$

so the only admissible solutions are those in which F and G are complex multiples of the identity.

When $F = G = 0$ we observe that

$$H = i[\mathscr{H}, a], \quad \text{where } \mathscr{H} = \mathscr{H}^*,$$

satisfies (4.8) and (4.11).

This describes the deterministic (Hamiltonian) evolution of the system $\mathscr{h}_0$ in the absence of "noise" effects, as given by the Heisenberg equations of motion:

$$\begin{aligned} da &= H\,dt, \\ da^\dagger &= H^\dagger\,dt, \end{aligned} \tag{5.1}$$

with solutions

$$\begin{aligned} a^t &= e^{i\mathscr{H}t} a_0 e^{-i\mathscr{H}t}, \\ a_t^\dagger &= e^{i\mathscr{H}t} a_0^\dagger e^{-i\mathscr{H}t}. \end{aligned} \tag{5.2}$$

Now suppose that $F, G \neq 0$. Then (4.11) gives

$$\{H, a\} = -\lambda\mu$$
$$\Rightarrow H = -\lambda\mu a^\dagger + \alpha a, \quad \text{where } \alpha \in \mathbb{C}.$$

Substituting this equation into (4.8), we obtain

$$\{-\lambda\mu a^\dagger + \alpha a, a^\dagger\} + \{a, -\bar{\lambda}\bar{\mu} a + \bar{\alpha} a^\dagger\}$$
$$= -(|\lambda|^2 + |\mu|^2) I$$
$$\Rightarrow \alpha + \bar{\alpha} = -(|\lambda|^2 + |\mu|^2).$$

Hence $\operatorname{Re}\alpha = -\frac{1}{2}(|\lambda|^2 + |\mu|^2)$, so the general form of H is given by

$$H = -\lambda\mu a^\dagger - \tfrac{1}{2}(|\lambda|^2 + |\mu|^2) a + i\beta a,$$

where $\beta \in \mathbb{R}$ ($\beta = \operatorname{Im}\alpha$).

We will show that the term $i\beta a$ corresponds to the Hamiltonian aspect of the evolution. Let $H = H' + H_0$ where $H_0 = i[\mathscr{H}, a]$. Now H is odd, so $H_0 = \gamma a + \delta a^\dagger$, $\exists\ \gamma, \delta \in \mathbb{C}$. $\mathscr{H}$ is required to be self-adjoint; thus

$$\mathscr{H} = tI + \omega a + \bar{\omega} a^\dagger + z a^\dagger a,$$

where $t, z \in \mathbb{R}$, $\omega \in \mathbb{C}$. Hence

$$\gamma a + \delta a^\dagger = i[tI + \omega a + \omega a^\dagger + z a^\dagger a, a]$$
$$= i\bar{\omega}(a^\dagger a - a a^\dagger) - iza.$$

$$\Rightarrow \delta = 0, \quad \gamma = -iz.$$

Now, it can be shown that any stochastic differential equation of the type (2.4) containing a Hamiltonian term corresponds to a cocycle perturbation of the same equation without the Hamiltonian term.[9] Hence it suffices to consider the equations

$$da = \lambda\,dA^\dagger + \mu\,dA - (\lambda\mu a^\dagger + \tfrac{1}{2}(|\lambda|^2 + |\mu|^2) a) dt, \tag{5.3}$$

$$da^\dagger = \bar{\mu}\,dA^\dagger + \bar{\lambda}\,dA - (\bar{\lambda}\bar{\mu} a + \tfrac{1}{2}(|\lambda|^2 + |\mu|^2) a^\dagger) dt. \tag{5.4}$$

The probabilistic properties of fermion Brownian motion are invariant under the group U(1), so we can make the gauge transformations

$$A \mapsto e^{-i\theta} A, \quad A^\dagger \mapsto e^{i\theta} A^\dagger.$$

We can also make the gauge transformations

$$a \mapsto e^{-i\phi} a, \quad a^\dagger \mapsto e^{i\phi} a^\dagger,$$

since these leave the relations (2.4) unchanged. Let

$$\theta = \tfrac{1}{2}(\arg\lambda - \arg\mu), \quad \phi = \tfrac{1}{2}(\arg\lambda + \arg\mu).$$

These transform (5.3) and (5.4) to forms where λ and μ are real and nonnegative.

We now introduce the Brownian motion (P_t, Q_t), where

$$P_t = -i(A_t - A_t^\dagger), \quad Q_t = A_t + A_t^\dagger,$$

and the process (p_t, q_t) for which

$$p_t = -i(a_t - a_t^\dagger), \quad q_t = a_t + a_t^\dagger,$$

satisfying the fixed time relations

$$q_t^2 = p_t^2 = I, \quad \{p_t, q_t\} = 0.$$

The addition of (5.3) and (5.4) yields

$$dq = (\lambda + \mu) dQ - \tfrac{1}{2}(\lambda + \mu)^2 q\,dt, \tag{5.5}$$

and subtracting (5.4) from (5.3) gives

$$dp = (\mu - \lambda) dP - \tfrac{1}{2}(\mu - \lambda)^2 p\,dt. \tag{5.6}$$

These have solutions in terms of stochastic integrals as follows:

$$q_t = e^{-(1/2)\rho^2 t} q_0 + \rho \int_0^t e^{-(1/2)\rho^2(t-\tau)} dQ_\tau, \tag{5.7}$$

$$p_t = e^{-(1/2)\theta^2 t} p_0 + \theta \int_0^t e^{-(1/2)\theta^2(t-\tau)} dP_\tau, \tag{5.8}$$

where

$$\rho = \lambda + \mu \quad \text{and} \quad \theta = \mu - \lambda.$$

This pair is the unique solution of Eqs. (2.4) in the case of a single degree of freedom and represents a fermion analog of the classical Ornstein–Uhlenbeck velocity process, which may be compared with the boson Ornstein–Uhlenbeck process[10,11] and is a special case of a fermion process found in Ref. 12.

[1] R. L. Hudson, M. Karandikar, and K. R. Parthasarathy, "Towards a Theory of Non-commutative Semimartingales Adapted to Brownian Motion and a Quantum Itô's Formula," in *Theory and application of random fields*, edited by G. Kallianpur, *Lecture Notes in Control and Information Sciences* (Springer, Berlin, 1983), Vol. 49.

[2] R. L. Hudson and K. R. Parthasarathy, Quantum Diffusions in *Theory and application of random fields*, edited by G. Kallianpur, *Lecture Notes in Control and Information Sciences* (Springer, Berlin, 1983), Vol. 49.

[3] C. Barnett, R. F. Streater, and I. F. Wilde, "The Itô–Clifford Integral," J. Func. Anal. **48**, 172–90 (1982).

[4] A. M. Cockcroft and R. L. Hudson, "Quantum Mechanical Wiener Processes," J. Multivariate Anal. **7**, 107–124 (1977).

[5]L. Accardi, A. Frigerio, and J. T. Lewis, "Quantum Stochastic Processes," Pub. Res. Inst. Math. Sci. **18**, 97–133 (1982).
[6]R. Brauer and H. Weyl, "Spinors in n Dimensions," Am. J. Math. **57**, 425–449 (1935).
[7]R. L. Hudson and R. F. Streater, "Itô's formula is the chain rule with Wick ordering," Phys. Lett. A **86**, 277–79 (1981).
[8]C. Chevalley, *The Construction and Study of Certain Important Algebras*, Publ. Math. Soc. Japan, I (Princeton University, Princeton, NJ, 1955).
[9]R. L. Hudson and K. R. Parthsarathy, "Construction of Quantum Diffusions," in Proceedings of the workshop "Quantum probability and applications to the quantum theory of irreversible processes," Rome, 1982 (to appear in Springer *Lecture Notes in Mathematics*).
[10]R. L. Hudson, P. D. F. Ion, and K. R. Parthsarathy, "Time-orthogonal unitary dilation and non-commutative Feynman–Kac formula," Commun. Math. Phys. **83**, 261–280 (1982).
[11]R. L. Hudson, P. D. F. Ion, and K. R. Parthasarathy, "The Feynman–Kac formula for Boson Wiener processes, " in *Quantum Mechanics in Mathematics, Chemistry and Physics*, edited by K. E. Gustafson (Plenum, New York, 1981), pp. 479–93.
[12]R. F. Streater, "Damped oscillator with quantum noise," J. Phys. A **15**, 1477–85 (1982).

Quantization of spinor fields. IV. Joint Bose–Fermi spectral problems

Piotr Garbaczewski
Institute of Theoretical Physics, University of Wrocław, 50-205 Wrocław, Poland

(Received 21 January 1983; accepted for publication 22 July 1983)

In this continued study of the connection between classical c-number spinor models and their quantized Fermi partners, we elaborate further necessary consequences of the *bosonization*. The genuine (c-number) path integral representation of tr $\exp(-iHt)$ is derived for the Fermi oscillator and simple lattice Fermi models. We find that the underlying Hamiltonian of the Fermi system H_F can be equivalently written as PH_BP, where H_B is the related Bose Hamiltonian, P is an appropriate projection in the state space of the Bose system, and $[P, H_B]_- = 0$, $H_F = PH_BP$. Grassmann algebras are not used. We prove further that both for the massive Thirring model (MT) and the chiral invariant Gross–Neveu model (CGN), the Bethe ansatz eigenstates for the Fermi Hamiltonians are exact eigenstates of the Bose MT and CGN Hamiltonians, so introduced that $H_F = PH_BP$, $[P, H_B]_- = 0$. As a consequence, through studying the c-number path integral representation for tr $\exp(-iH_Ft)$, we establish a class of classical (c-number) spinor solutions of the underlying field equations, which at the same time make stationary both the c-number Bose and c-number Fermi actions.

PACS numbers: 11.10.Cd

1. MOTIVATION

As is well known, for simple scalar field theory models, the U matrix of the scattering theory in the coherent state representation is identical with the conventional Feynman integral

$$\begin{aligned} U(t,t') &= (\varphi(t')|\exp[-iH(t'-t)]|\varphi(t)) \\ &= (\varphi(0)|\exp(iH_0t)\exp(-iH(t'-t)) \\ &\quad\times \exp(-iH_0t)|\varphi(0)) \\ &= \int \left[\frac{D\varphi}{\sqrt{2\pi}}\right]\exp\left[i\int_t^{t'} d^4x\,\mathscr{L}(\dot{\varphi},\varphi)\right]. \end{aligned} \tag{1.1}$$

However, in the case of Fermi fields the analogous formula exists only if the *classical* spinor fields to be used in the path integral are elements of the anticommuting function ring (Grassmann algebra). It created a folklore belief that the genuine classical c-number spinor fields have no meaning for the construction and then understanding of the canonical Fermi field theory. In fact, if one would follow the route (1.1) by using the c-number spinor field Lagrangian and the measure with respect to c-number spinor paths, the received result would certainly not coincide with the result known by other (Grassmann algebra) means to be correct for the respective Fermi model.

However, in the preceding series of papers[1–3] (see also Ref. 4), we have demonstrated that it is not meaningless to talk about a connection of classical c-number spinor fields with their quantized Fermi partners. Some other intuitions about a possible quantum meaning of classical c-number spinor fields can be drawn from Refs. 5–8. Our approach to the quantization of spinor fields problems originates from Klauder's[9] idea to avoid the use of Grassmann algebras in describing the canonically quantized Fermi fields, but is different from that work.

We have found previously that a formulation of the correspondence rule between quantized Fermi and classical c-number spinor fields is in principle possible, provided the Fermi model allows for *bosonization*. Different aspects of this problem were studied before in Refs. 1–3. However, a fundamental problem we were confronted with, the quantum meaning of classical c-number spinor models (with the Fermi–Dirac statistics in mind), has remained unsolved.

In the present paper we give a path integral reconstruction of tr $\exp(-iHt)$ in terms of genuine c-number trajectories for simple Fermi models, to prove that it is at all possible without any reference to Grassmann algebras. We do not, however, conclude (as Klauder did) that Grassmann algebras are an unnecessary addition to the mathematical physics. We believe they are extremely useful for explicit perturbative calculations. They suffer, however, from a serious drawback: they do not carry any true physical content, against the naive (but popular) expectations. In the light of the results presented below, it is possible to recover a new face (physical content) of Fermi models, the face which is completely obscure in the Grassmann algebra formulation.

For lattice Fermi models, the Jordan–Wigner transformation allows us to introduce the equivalent spin $\frac{1}{2}$ lattices. The corresponding Hamiltonians can be considered as projected Bose Hamiltonians (in the Hilbert space of the appropriate Bose system),

$$H_F = PH_BP, \tag{1.2}$$

where $[P, H_B]_- = 0$. One should notice that

$$\begin{aligned} 1 = 1_B = P + (1-P) &\Rightarrow \mathrm{tr}\,\exp(iH_Bt) \\ &= \mathrm{tr}\,\exp(iH_Ft) + \mathrm{tr}\,\exp(-(1-P)H_B(1-P)t), \end{aligned} \tag{1.3}$$

and the only problem is to extract the Fermi contribution to the Bose formula, which is known to be provided by using Grassmann algebra methods to compute tr $\exp(-iH_Ft)$.

If there exists a countable family of projections

$$\sum_k P_k = 1, \quad P_k P_l = \delta_{kl} P_k, \quad [P_k, H_B]_- = 0 \quad \forall\ k, \tag{1.4}$$

then

$$\operatorname{tr} \exp(-iH_B t) = \sum_{k=1}^{\infty} \operatorname{tr} \exp(-iH_F^k t), \tag{1.5}$$
$$H_F^k = P_k H_B P_k,$$

and the respective Bose model can be viewed as a (infinitely) reducible Fermi one or even as a *tower* of possibly distinct Fermi models, each one with its own Hamiltonian H_F^k.

For the (continuous) massive Thirring model, we prove that Bethe ansatz eigenvectors of the Fermi Hamiltonian are the exact eigenvectors (with the same eigenvalues) of the Bose massive Thirring model Hamiltonian, see also Ref. 4, i.e.,

$$H_F^{\mathrm{MT}} = P H_B^{\mathrm{MT}} P, \quad [P, H_B^{\mathrm{MT}}]_- = 0. \tag{1.6}$$

Previously,[4] by using indirect (the inverse spectral transform) methods we have concluded that the mass spectrum of bound states is the same for the Bose and Fermi massive Thirring models. If to combine it with our Bethe ansatz observation, we find supported the conjecture of Ref. 4 that the Bose massive Thirring model can be equivalently rewritten as the (infinitely) reducible Fermi one. Next, we deduce the spin $\frac{1}{2}$ type version of H_F^{MT} for which a path integral representation of $\operatorname{tr}\exp(-iH_F t)$ is constructed in terms of c-number trajectories. We prove that spinor trajectories such that $\sigma_i = \phi_i^* \phi_i \leqslant A < \infty$, $i = 1,2$, A being arbitrary, give exactly the same contributions to both Bose and Fermi path integrals via the c-number massive Thirring model action. Its stationary *points* are the classical spinor solutions of the respective field equations satisfying $\sigma_i \leqslant A < \infty$.

Finally the analogous properties are established for the chiral invariant Gross–Neveu model. Our analysis is confined to 1 + 1 dimensions where explicit solutions for the spectral problems are available. We expect, however, that the Bose–Fermi interplay described here will prove useful in 1 + 3 dimensions as well, see, e.g., also Ref. 2.

2. QUANTUM OSCILLATOR PROBLEM: BOSE VERSUS FERMI

The Bose oscillator

$$\hat{L}_B = ib^*(t)\dot{b}(t) - \omega b^*(t)b(t) = b^*(t)\left(i\frac{d}{dt} - \omega\right)b(t) \tag{2.1}$$

is determined by using the CCR algebra generators (equal time variables omitted)

$$[b, b^*]_- = 1_B, \quad b|0\rangle = 0. \tag{2.2}$$

The Hamiltonian reads

$$\hat{h}_B = \omega b^* b; \tag{2.3}$$

hence the infinitesimal propagator of the model is given by

$$\hat{U}_B(\Delta t) = \exp(-i\Delta t \hat{h}_B) \cong 1 - i\Delta t \hat{h}_B. \tag{2.4}$$

In the coherent state representation, the infinitesimal kernel[10] of $\hat{U}_B(t)$ reads

$$U_B(\Delta t) = \exp(\beta^*\beta - i h_B^{\mathrm{cl}} \Delta t),$$
$$h_B^{\mathrm{cl}} = \langle \beta | \hat{h}_B | \beta \rangle = \omega \beta^* \beta, \tag{2.5}$$
$$|\beta\rangle = \exp(\beta b^* - \beta^* b)|0\rangle,$$

so that through the standard arguments,[11] we arrive at the following (formal continuum t limit) path integral representation of $\operatorname{tr}\exp(-i\hat{h}_B t)$:

$$I_B = \operatorname{tr}\exp(-i\hat{h}_B t) = \int [d\beta][d\beta^*] \exp i \int_0^t \{ i\beta^*(t)\dot{\beta}(t) - \omega\beta^*(t)\beta(t)\} dt = \int [d\beta][d\beta^*] \exp i \int_0^t L_B(t) dt. \tag{2.6}$$

One should realize that I_B is given with the accuracy up to the normalization factor.

Let us define the Fermi oscillator

$$\hat{L}_F = a^*(t)\left(i\frac{d}{dt} - \omega\right)a(t) \tag{2.7}$$

by using the Fock representation of the CAR algebra, which is completely embedded in the CCR algebra as follows:

$$[a, a^*]_+ = 1_F, \quad a^*|0\rangle = |1\rangle = b^*|0\rangle, \quad a|0\rangle = 0,$$
$$a^{*2} = 0 = a^2, \tag{2.8}$$
$$a^* = b^*\!: \exp(-b^*b):, \quad a = :\exp(-b^*b):a,$$
$$1_F = :\exp(-b^*b): + b^*\!: \exp(-b^*b):b.$$

The CAR generators act invariantly on a proper subspace $h_F = 1_F h$ of the Bose oscillator Hilbert space, but nevertheless, allow a trivial extension to the whole of h. Consequently the Fermi propagator $\hat{U}_F(\Delta t) = \exp(-i\Delta t \hat{h}_F)$ can be represented in the Hilbert space of the Bose oscillator, thus allowing one to follow the previous path integration route. We have

$$\hat{U}_F(\Delta t) = \exp(-i\hat{h}_F \Delta t) \cong 1_F - i\Delta t \hat{h}_F = 1_F(1 - i\Delta t \hat{h}_B)1_F$$
$$= :\exp(-b^*b): + b^*\!: \exp(-b^*b):b - i\omega\Delta t b^*\!: \exp(-b^*b):b, \tag{2.9}$$

so that the infinitesimal kernel reads

$$U_F(\Delta t) \cong \langle \beta | 1_F - i\Delta t \hat{h}_F) | \beta \rangle \cdot \exp(\beta^*\beta)$$
$$= 1 + \beta^*\beta - i\omega\Delta t \beta^*\beta = (1 + \beta^*\beta)[1 - i\omega\Delta t \beta^*\beta/(1 + \beta^*\beta)]$$
$$\cong \exp[\ln(1 + \beta^*\beta) - i\omega\Delta t \beta^*\beta/(1 + \beta^*\beta)]. \tag{2.10}$$

Furthermore

$$I_F = \operatorname{tr}\exp(-i\hat{h}_F t) = \int [d\beta][d\beta^*] \exp i\int_0^t \frac{i\beta^*\dot{\beta} - \omega\beta^*\beta}{1+\beta^*\beta}\,dt = \int [d\beta][d\beta^*] \exp i\int_0^t dt\,\frac{L_B(t)}{1+\beta^*(t)\beta(t)}, \tag{2.11}$$

which is a c-number alternative for the usual,[6] Grassmann algebra path integral.

Remark: The semiclassical quantization procedure for the continuous Heisenberg system[12] resulted in the introduction of the spin path integral with respect to genuine (i.e., c-number, non-Grassmann) paths in the phase space of the classical spin system. The genuine c-number path notion is also inherent in the approach of Refs. 13–16 based on the SU(2) phase variables, and making use[14–16] of spin coherent states, these being well known in the many-body physics. The general method of Ref. 16 to construct measures for spin variable path integrals can be immediately adopted to either a single spin $\frac{1}{2}$ or to the many-body spin $\frac{1}{2}$ problem.

Attempts to introduce probabilistic ideas (i.e., probabilistic measures) to the study[17,18] of Fermi fields start from an appealing assumption. Consider the classical harmonic oscillator problem. View its equations of motion as stochastic differential equations and then add information that one is considering a two-level Fermi system instead of the ordinary Bose one. It results in specifying the class of stochastic processes in which solutions of the would-be classical oscillator equations of motion are to be found. The underlying processes are Markov processes with values in Z_2 which demonstrates that, except for the form of the equations of motion, the *classical* paths of the Fermi system are as unrelated to the Bose oscillator paths as the Grassmann algebra *paths* would be. An analogous line is followed in Ref. 19, though in a different Poisson processes framework.

3. SPIN $\frac{1}{2}$ LATTICES, LATTICE FERMIONS, AND c-NUMBER PATH INTEGRALS

It is well known that at least in 1 + 1 dimensions, the lattice Fermi systems can be equivalently described as lattices of spins $\frac{1}{2}$, and conversely. It happens so due to the Jordan–Wigner (JW) transformation, realizing fermions as *strings* of spins $\frac{1}{2}$ in the linear chain. An easy example is here in the Ising model in 1 + 1 dimensions, whose Hamiltonian,

$$H = J\sum_i \sigma_i^x\sigma_{i+1}^x, \tag{3.1}$$

after making the JW transformation, and then Fourier transforming the image Fermi variables, goes over to the one[20] which can be unitarily transformed into

$$\begin{aligned} H &= \sum_i \epsilon_q(\xi_q^*\xi_q - \tfrac{1}{2}),\\ \epsilon_q &= \cosh^{-1}\{\cosh 2(J-J')\\ &\quad + (1-\cos q)\sinh 2J'\sinh 2J\},\\ J' &= \tanh^{-1}\exp(-2J), \end{aligned} \tag{3.2}$$

with

$$\begin{aligned} &[\xi_q, \xi_{q'}^*]_+ = \delta_{qq'}, \quad [\xi_q, \xi_{q'}]_+ = 0,\\ &q = (2\pi/N)p, \quad p = 0, \pm 1\ldots, \pm(N-2)/2, N/2. \end{aligned} \tag{3.3}$$

Since $\xi_q^*\xi_q$ is a particle number operator of the qth mode with eigenvalues 0,1, we can replace it by the equivalent one given in terms of Pauli operators

$$\begin{aligned} &\xi_q^*\xi_q \equiv \sigma_q^+\sigma_q^- = b_q^*{:}\exp(-b_q^*b_q){:}\,b_q,\\ &[b_q, b_{q'}^*]_- = \delta_{qq'}, \quad [b_q, b_{q'}]_- = 0,\\ &b_q|0\rangle = 0 = \xi_q|0\rangle = \sigma_q^-|0\rangle \quad \forall\, q,\\ &b_q^*|0\rangle = \xi_q^*|0\rangle = \sigma_q^+|0\rangle. \end{aligned} \tag{3.4}$$

The operator unit contributing the $\frac{1}{2}$ term to H equals $1_F = \Pi_q 1_q$, with $1_q = {:}\exp(-b_q^*b_q){:} + b_q^*{:}\exp(-b_q^*b_q){:}\,b_q$. The Ising Hamiltonian acquires thus the Bose form

$$\begin{aligned} H &= \sum_q \frac{\epsilon_q}{2}[b_q^*{:}\exp(-b_q^*b_q){:}\,b_q - {:}\exp(-b_q^*b_q){:}]\\ &= \sum_q H_q,\\ &[H_q, H_{q'}]_- = 0. \end{aligned} \tag{3.5}$$

Consequently, the infinitesimal propagator

$$\begin{aligned} \hat{U}_F(\Delta t) &= \exp(-iH\Delta t)\\ &\cong \prod_q (1_q - i\Delta t H_q) = 1_F - i\Delta t H\\ &\cong \prod_q \exp(-iH_q\Delta t),\\ 1_F &= \prod_q 1_q \end{aligned} \tag{3.6}$$

allows for the following infinitesimal kernel:

$$\begin{aligned} U_F(\Delta t) &\cong \Big(\exp\sum_q \beta_q^*\beta_q\Big)\cdot\langle\beta|\prod_q(1_q - i\Delta t H_q)|\beta\rangle\\ &= \prod_q\left\{(1+\beta_q^*\beta_q)\left[1 - \frac{i\Delta t\epsilon_q}{2}\,\frac{\beta_q^*\beta_q - 1}{1+\beta_q^*\beta_q}\right]\right\}\\ &\cong \exp\sum_q\left[\ln(1+\beta_q^*\beta_q) - \frac{i\Delta t\epsilon_q}{2}\,\frac{\beta_q^*\beta_q - 1}{1+\beta_q^*\beta_q}\right], \end{aligned} \tag{3.7}$$

where

$$|\beta\rangle = \exp\sum_q(\beta_q b_q^* - \beta_q^* b_q)|0\rangle, \tag{3.8}$$

and so

$$I_F = \mathrm{tr}\exp(-iHt) = \int [d\beta][d\beta^*] \times \exp i \int_0^t \sum_q \frac{i\beta_q^*\dot\beta_q - (\epsilon_q/2)(\beta_q^*\beta_q - 1)}{1+\beta_q^*\beta_q} dt . \quad (3.9)$$

The lattice Fermi model of Ref. 21,

$$H_F = -i\sum_n \{\psi_n^*\psi_{n+1} - \psi_{n+1}^*\psi_n\},$$
$$[\psi_n, \psi_m]_+ = 0, \quad [\psi_n, \psi_m^*]_+ = \delta_{nm}, \quad (3.10)$$
$$\dot\psi_n = -\tfrac{1}{2}(\psi_{n+1} - \psi_{n-1}),$$

can be equivalently rewritten in terms of lattice spins $\frac{1}{2}$:

$$H_F = -\frac{1}{2}\sum_n \{\sigma_n^+\sigma_{n+1}^- + \sigma_n^-\sigma_{n+1}^+\}. \quad (3.11)$$

However, a straightforward application of the previous *bosonization* recipe to the spin $\frac{1}{2}$ generators $\sigma_m^\pm$ is not efficient at all. It is much more reliable to make first a lattice Fourier transformation of the original Fermi variables,

$$\psi_n = N^{-1/2}\sum_q c_q \exp(iqn),$$
$$q = (2\pi/N)p, \quad p = 0, \pm 1, \ldots, \pm(N-2)/2, \ N/2, \quad (3.12)$$

where the familiar lattice identity (N is large)

$$\frac{1}{N}\sum_q \exp[-iq(n-m)] = \delta_{nm} \quad (3.13)$$

guarantees that both ψ_n, ψ_n^*, and c_q, c_q^* are the CAR algebra generators, and moreover

$$\begin{aligned} H_F &= -\frac{i}{2}\sum_n \frac{1}{N} \times \sum_{p,q}\{e^{-iqn}e^{i(n+1)p}c_q^*c_p - e^{ipn}e^{-iq(n+1)}c_q^*c_p\} \\ &= -\frac{i}{2}\sum_{q,p}\Big(\frac{1}{N}\sum_n e^{-in(q-p)}\Big)[e^{ip}c_q^*c_p - e^{-iq}c_q^*c_p] \\ &= -\frac{i}{2}\sum_q [e^{iq}c_q^*c_q - e^{-iq}c_q^*c_q] = \sum_q (\sin q)c_q^*c_q . \end{aligned} \quad (3.14)$$

Now the previous route applies apparently, and then

$$I_F = \mathrm{tr}\exp(-iH_Ft) = \int [d\beta][d\beta^*] \times \exp i\int_0^t \sum_q \frac{i\beta_q^*\dot\beta_q - (\sin q)\beta_q^*\beta_q}{1+\beta_q^*\beta_q} dt . \quad (3.15)$$

The path integral formulation for simplest Fermi models on the lattice, in terms of genuine c-number paths is thus possible, though certainly not well suited for less trivial examples, like, e.g., the lattice version of the massive Thirring model,[22]

$$\begin{aligned} H_{\mathrm{MT}} = \sum_{n=-N+1}^{N} &\Big\{\frac{iv}{2a}(\psi_n^*\psi_{n+1} - \psi_{n+1}^*\psi_n) \\ &+ (-1)^n\frac{m}{2}(\psi_n^*\psi_{n+1}^* + \psi_{n+1}\psi_n) \\ &- \frac{g}{2a}\Big(\psi_n^*\psi_n - \frac{1}{2}\Big)\Big(\psi_{n+1}^*\psi_{n+1} - \frac{1}{2}\Big)\Big\} - E_0, \end{aligned} \quad (3.16)$$

a being the lattice spacing. By using the JW transformation, one arrives at the equivalent spin $\frac{1}{2}$ xyz model Hamiltonian (with cyclic boundary conditions),

$$\begin{aligned} H_{\mathrm{MT}} = -\frac{1}{2}\sum_n &\Big\{\Big(\frac{v}{2a} + \frac{m}{2}\Big)\sigma_n^x\sigma_{n+1}^x \\ &+ \Big(\frac{v}{2a} - \frac{m}{2}\Big)\sigma_n^y\sigma_{n+1}^y + \frac{g}{2a}\sigma_n^z\sigma_{n+1}^z\Big\} - E_0, \end{aligned} \quad (3.17)$$

which is also

$$\begin{aligned} H_{\mathrm{MT}} &= 1_F H_{\mathrm{MT}}^B 1_F, \\ H_{\mathrm{MT}}^B &= -\frac{1}{2}\sum_n \Big\{\frac{v}{2a}(b_n^*b_{n+1} + b_{n+1}^*b_n) \\ &\quad + \frac{m}{2}(b_n^*b_{n+1}^* + b_nb_{n+1}) \\ &\quad - \frac{g}{2a}\Big(b_n^*b_n - \frac{1}{2}\Big)\Big(b_{n+1}^*b_{n+1} - \frac{1}{2}\Big)\Big\} - E_0, \\ 1_F &= \prod_n [:\exp(-b_n^*b_n): + b_n^*:\exp(-b_n^*b_n):b] \\ &= \prod_n P_n . \end{aligned} \quad (3.18)$$

Here H_B is considered as an operator in the Hilbert space $\mathscr{H}_N = h^{\otimes 2N}$, while H_F in $1_F\mathscr{H}_N = h_F^{\otimes 2N}$, $h_F = Ph$. Because 1_F is a projection operator in $\mathscr{H}_N$, we have

$$\begin{aligned} H_B &= H_F + (1-1_F)H_B1_F + 1_FH_B(1-1_F) \\ &\quad + (1-1_F)H_B(1-1_F), \end{aligned} \quad (3.19)$$

and consequently for any $|\psi\rangle = 1_F|\psi\rangle \in \mathscr{H}_N$ we find

$$H_B|\psi\rangle = H_F|\psi\rangle + (1-1_F)H_B|\psi\rangle . \quad (3.20)$$

Hence the necessary and sufficient condition to fulfill

$$H_B|\psi\rangle = H_F|\psi\rangle, \quad 1_F|\psi\rangle = |\psi\rangle \quad (3.21)$$

is that

$$H_B|\psi\rangle \in 1_F\mathscr{H}_N \quad \text{if } |\psi\rangle \in 1_F\mathscr{H}_N . \quad (3.22)$$

Then if $H_F|\psi\rangle = \epsilon|\psi\rangle$ we get automatically $H_B|\psi\rangle = \epsilon|\psi\rangle$ and conversely (provided $1_F|\psi\rangle = |\psi\rangle$).

Since we work in the Hilbert space $\mathscr{H}_N$, the respective spectra are discrete, and $\mathscr{H}_N$ is spanned by a complete (countable) eigenfunction system of H_B. 1_F is a projection in $\mathscr{H}_N$ and $1_F\mathscr{H}_N$ is spanned by these eigenvectors of H_B which obey $1_F|\psi\rangle = |\psi\rangle$. Notice that if $H_F|\psi\rangle = \epsilon|\psi\rangle$, then necessarily

$$H_B|\psi\rangle = H_B \sum_{(\alpha)} a_\alpha |\psi,\alpha\rangle = \sum_{(\alpha)} a_\alpha \epsilon_\alpha |\psi,\alpha\rangle$$
$$= H_F|\psi\rangle = \epsilon|\psi\rangle = \epsilon \sum_{(\alpha)} a_\alpha |\psi,\alpha\rangle,$$
$$\text{i.e., } \epsilon = \epsilon_\alpha \quad \forall\ \alpha \in (\alpha). \tag{3.23}$$

It means that once any H_B is obtained for which $H_F = 1_F H_B 1_F$ for a given Fermi model H_F, then automatically the eigenfunction system of H_F is a subsystem in this of H_B, and moreover the respective eigenvalues of H_B and H_F do coincide. It then follows that

$$\operatorname{tr}\exp(-iH_Bt) = \operatorname{tr}\exp(-iH_Ft) + \operatorname{tr}\exp[-(1-1_F)H_B(1-1_F)]. \tag{3.24}$$

In the N-particle linear chain, the operator $1_F = \Pi_n P_n$ is composed from projections on two lowest levels in each single site Schrödinger problem. Obviously, there is an infinity of other choices. In particular, if to imagine a sequence

$$\left\{ 1_F^k = \prod_n P_n^k \quad \text{such that} \quad \sum_k 1_F^k = 1,\ 1_F^k \cdot 1_F^l = \delta_{kl} 1_F^k, \right.$$

then

$$\operatorname{tr}\exp(-iH_Bt) = \sum_k \operatorname{tr}\exp(-iH_F^k t),$$
$$H_F^k = 1_F^k H_B 1_F^k, \tag{3.25}$$

provided $[H_B, 1_F^k]_- = 0\ \forall\ k = 1,2,\ldots$. In principle any two states at each site can be used for our construction. However, the need for the last commutation rule may play the role of the constraint capable of removing a nonuniqueness appearing in the construction of 1_F. The Bose system may thus happen to be equivalent either to a reducible Fermi one, or to the whole *tower* of might-be-distinct Fermi systems. A speculation on the possibly infinite *tower* of fundamental, say, leptons generated by the relatively simple Bose system, does not seem to be hopeless.

Let us mention that the expansion of the Bose trace into a sum of Fermi traces resembles the expansions which appear in the path integral quantization of spin systems.[13] Then the Green's function is known to propagate all spins simultaneously. Nevertheless the recovery of the usual Pauli spinors is possible by projecting to a specific spin subspace which is propagated into itself. One deals then with a Bose system, which describes a *particle* allowed to live in several (infinity in fact) spin states.

To compute a particular Fermi trace, we may not refer to the original (Bose) phase-space variables, but instead we can use the conventional Grassmann algebra tools. They prove to be successful, indeed, once remaining in the particular Fermi sector of the Bose model.

4. MASSIVE THIRRING MODEL: FROM FERMIONS TO BOSONS

A spectral problem for the Fermi model,

$$H_F = \int dx[-i(\psi_1^*\partial_x\psi_1 - \psi_2^*\partial_x\psi_2) + m(\psi_1^*\psi_2 + \psi_2^*\psi_1) + 2g\psi_1^*\psi_2^*\psi_2\psi_1], \tag{4.1}$$

has been solved by means of the Bethe ansatz in Ref. 23. One assumes to work with a Fock representation of the CAR algebra,

$$[\psi_i(x), \psi_j^*(y)]_+ = \delta_{ij}\delta(x-y),$$
$$[\psi_i(x), \psi_j(y)]_+ = 0, \quad \psi_i(x)|0\rangle = 0,$$
$$\forall\ i = 1,2,\ x \in R^1. \tag{4.2}$$

Then after introducing

$$|\alpha_1,\ldots,\alpha_n\rangle = \int dx_1 \cdots \int dx_n \chi(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n) \times \prod_{i=1}^n \psi^*(x_i,\alpha_i)|0\rangle,$$
$$\chi(x_1,\alpha) = \exp\left(im\sum_i x_i \sinh\alpha_i\right) \times \prod_{1<i<j<n} [1 + i\lambda(\alpha_i,\alpha_j)\epsilon(x_i - x_j)], \tag{4.3}$$
$$\psi(x,\alpha) = e^{\alpha/2}\psi_1(x) + e^{-\alpha/2}\psi_2(x),$$
$$\lambda(\alpha_i,\alpha_j) = -\tfrac{1}{2}g\tanh\tfrac{1}{2}(\alpha_i - \alpha_j),$$

one demonstrates (see, e.g., the Appendix of Ref. 23) that

$$H_F|\alpha_1,\ldots,\alpha_n\rangle = \left(\sum_i m\cosh\alpha_i\right)|\alpha_1,\ldots,\alpha_n\rangle. \tag{4.4}$$

Let us now assume that the fields entering H_F satisfy not the CAR algebra relations (4.2), but the CCR algebra ones,

$$H_F \to H_B = H_F(\psi \to \phi),$$
$$[\phi_i(x), \phi_j^*(y)]_- = \delta_{ij}\delta(x-y), \quad [\phi_i(x), \phi_j(y)]_- = 0, \tag{4.5}$$

provided the CAR algebra (4.2) is constructed in the Fock representation of the CCR algebra, according to Ref. 24. It means that if to introduce the antisymmetric function of Ref. 24,

$$\sigma^3 = \sigma, \quad x_k \in R^1, \quad i_k = 1,2, \quad \forall\ k = 1,2,\ldots,n,$$
$$\sigma = \sigma(x_1,i_1,\ldots,x_j,i_j,\ldots,x_k,i_k,\ldots,x_n,i_n)$$
$$= -\sigma(\ldots x_k,i_k,\ldots,x_j,i_j,\ldots), \tag{4.6}$$

where $\sigma(x_1,i_1,\ldots,x_n,i_n) = 0,\ \pm 1$ depending on the choice of the (x,i) sequence; then the operator

$$\psi_k(x) = \sum_n \frac{\sqrt{1+n}}{n!} \times \sum_{i_1,\ldots,i_n} \int dy_1 \cdots \int dy_n \sigma(y_1,i_1,\ldots,y_n,i_n)$$
$$\times \sigma(x,k,y_1,i_1,\ldots,y_n,i_n)\cdot\phi_{i_1}^*(y_1)\cdots\phi_{i_n}^*(y_n).$$
$$\times :\exp\left\{-\sum_{i=1}^2 \int dz\phi_i^*(z)\phi_i(z)\right\}:$$
$$\times \phi_k(x)\phi_{i_1}(y_1)\ldots\phi_{i_n}(y_n), \quad x \in R^1, \quad k = 1,2 \tag{4.7}$$

is the CAR algebra generator of (4.2), and together with its adjoint, obeys

$$\phi_i^*(x)|0\rangle = \psi_i^*(x)|0\rangle, \quad \phi_i(x)|0\rangle = 0 = \psi_i(x)|0\rangle \quad \forall\ i, x,$$
$$\psi_{i_1}^*(x_1)\cdots\psi_{i_n}^*(x_n)|0\rangle$$
$$= \sigma(x_1, i_1,\ldots, x_n, i_n)\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n)|0\rangle,$$
$$[\psi_i(x), \psi_j^*(y)]_+ = \delta_{ij}\delta(x-y)1_F, \tag{4.8}$$
with
$$1_F = \sum_n \frac{1}{n!} \sum_{i_1,\ldots,i_n} \int dx_1 \cdots \int dx_n \sigma^2(x_1, i_1,\ldots, x_n, i_n)$$
$$\times\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n) \colon \exp\left\{-\sum_{i=1}^{2}\int dz \phi_i^*(z)\phi_i(z)\right\}\colon$$
$$\times\phi_{i_1}(x_1)\cdots\phi_{i_n}(x_n), \tag{4.9}$$
which is a projection in the Hilbert space of the Bose system selecting the Fermi subspace in it.

Consequently the Bethe ansatz states for the original Fermi model can be generated from the vacuum by using the Bose operators. We have
$$[e^{\alpha_1/2}\psi_1^*(x_1) + e^{-\alpha_1/2}\psi_2^*(x_1)]\cdots$$
$$[e^{\alpha_n/2}\psi_1^*(x_n) + e^{-\alpha_n/2}\psi_2^*(x_n)]|0\rangle$$
$$= \sum_{i_1\cdots i_n}\{\sigma(x_1, i_1,\ldots, x_n, i_n)\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n)|0\rangle$$
$$\times \exp\sum_{k=1}(-1)^{i_k+1}\alpha_k/2\}, \tag{4.10}$$
where $i_k = 1,2$, and obviously
$$\prod_{k=1}^{n}[e^{\alpha_k/2}\psi_1^*(x_k) + e^{-\alpha_k/2}\psi_2^*(x_k)]|0\rangle$$
$$= \sum_{i_1,\ldots,i_n}\left[\exp\sum_{k=1}^{n}(-1)^{i_k+1}\alpha_k/2\right]\prod_{k=1}^{n}\psi_{i_k}^*(x_k)|0\rangle. \tag{4.11}$$
Hence
$$|\alpha_1,\ldots,\alpha_n\rangle$$
$$= \int dx_1\cdots\int dx_n \chi(x_1,\ldots,x_n, \alpha_1,\ldots,\alpha_n)$$
$$\times\sum_{i_1\cdots i_n}\sigma(x_1 i_1,\ldots, x_n, i_n)\left[\exp\sum_k(-1)^{i_k+1}\alpha_k/2\right]$$
$$\times\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n)|0\rangle$$
$$= \int dx_1\cdots\int dx_n\sum_{\{i\}}\eta(x,\alpha,i)\phi_{i_1}^*(x_1)\cdots\phi_{i_1}^*(x_n)|0\rangle, \tag{4.12}$$
with
$$\eta(x,\alpha,i) = \chi(x,\alpha)\sigma(i,x)\exp\sum_{k=1}^{n}(-1)^{i_k+1}\alpha_k/2. \tag{4.13}$$
In Ref. 4 we have proved via the inverse spectral transform method, a close connection between the Bose and Fermi versions of the massive Thirring model. To make this connection more explicit, we shall demonstrate that vectors (4.13) are eigenvectors of the Hamiltonian H_B arising as $H_B = H_F(\psi^* \to \phi^*, \psi \to \phi)$. For the kinetic term of H_B, $\Sigma_{i=1}^2 \phi_i^*(x)\partial_x\phi_i(x)(-1)^{i+1}$ by commuting it through the product of ϕ^*'s, and then integrating by parts we arrive at
$$H_B^{\rm kin}|\alpha_1,\ldots,\alpha_n\rangle$$
$$= \int dx_1\cdots\int dx_n\sum_{i_1\cdots i_k}$$
$$\times\sum_{k=1}^{n}\{(-1)^{i_k+1}[(-i\nabla_k\chi)\sigma$$
$$+\chi(-i\nabla_k\sigma)]\exp\left(\sum_{j=1}^{n}(-1)^{i_j+1}\alpha_j/2\right)$$
$$\times\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n)\}|0\rangle$$
$$= \int dx_1\cdots\int dx_n\sum_{k=1}^{n}(-i\nabla_k\chi)\psi^*(\alpha_1, x_1)\cdots$$
$$\times(e^{\alpha_k/2}\psi_1^*(x_k) - e^{-\alpha_k/2}\psi_2^*(x_k))\cdots\psi^*(\alpha_n, x_n)|0\rangle$$
$$= \int dx_1\cdots\int dx_n\chi(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n)\sum_{i_1\cdots i_n}$$
$$\times\sum_{k=1}^{n}(-1)^{i_k+1}(-i\nabla_k\sigma)$$
$$\times\left(\exp\sum_{j=1}^{n}(-1)^{i_j+1}\alpha_j/2\right)\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n)|0\rangle$$
$$= H_F^{\rm kin}|\alpha_1,\ldots,\alpha_n\rangle + |A\rangle. \tag{4.14}$$
An exact form of $H_F^{\rm kin}|\alpha_1,\ldots\alpha_n\rangle$ was given in Ref. 23, and we shall concentrate on the spurious term $|A\rangle$ in $H_B^{\rm kin}|\alpha_1,\ldots,\alpha_n\rangle$. By virtue of (4.6), we have formally
$$(-i\nabla_k)\sigma(x_1 i_1,\ldots, x_n, i_n)$$
$$= (-i\nabla_k)\sigma^{2n+1}(x_1 i_1,\ldots, x_n, i_n)$$
$$= (2n+1)\sigma^{2n}(-i\nabla_k)\sigma = (2n+1)\sigma^2(-i\nabla_k)\sigma$$
$$= (-i\nabla_k)\sigma, \quad n = 1,2,\ldots, \tag{4.15}$$
which holds true for all integers and all possible choices of sequences $\{(x,i)\}$ while inserted in (4.14). Hence the identity (4.15) can be satisfied if and only if (up to a set of measure zero)
$$(-i\nabla_k)\sigma(x_1 i_1,\ldots, x_n, i_n) \equiv 0, \quad \forall\ k = 1,2,\ldots, n. \tag{4.16}$$
It however means that $|A\rangle = 0$ which proves the property
$$H_B^{\rm kin}|\alpha_1,\ldots,\alpha_n\rangle = H_F^{\rm kin}|\alpha_1,\ldots,\alpha_n\rangle. \tag{4.17}$$
Remark: As an example of $\sigma = \sigma^3$ used in the above, one can take
$$\sigma(x_1, i_1,\ldots, x_n, i_n) = \prod_{1 \leqslant j<k \leqslant n} p_{jk},$$
$$p_{jk} = \delta_{i_j i_k}[\Theta(x_j - x_k) - \Theta(x_k - x_j)] + |i_k - i_j|(-1)^{\Theta(x_j - x_k)}, \tag{4.18}$$
where $\Theta(x-y) = 1$, $x \geqslant y$, 0, otherwise.

For $i_j = i_k$ we have
$$p_{jk} = \Theta(x_j - x_k) - \Theta(x_k - x_j), \tag{4.19}$$
which equals either 0 or ± 1.

For $i_j \neq i_k$ we arrive at
$$p_{jk} = (-1)^{\Theta(x_j - x_k)}, \tag{4.20}$$
which equals ± 1. Consequently $p_{jk} = 0, \pm 1$ and in addition to the manifest antisymmetry property, $(i_j x_j) \leftrightarrow (i_k x_k) \Rightarrow \sigma \to -\sigma$, we have satisfied the property $\sigma^3 = \sigma$ as required by our previous definition. Notice that σ vanishes if and only if a pair (i, x) appears more than once in the sequence $\{(i, x)\}$.

For the mass term of H_B we get

$$H_B^m = m\int dx\,[\phi_1^*(x)\phi_2(x) + \phi_2^*(x)\phi_1(x)],$$

$$\begin{aligned} H_B^m|\alpha_1,\ldots,\alpha_n\rangle &= m\int dx_1\cdots\int dx_n\chi(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n) \\ &\times\sum_{i_1\cdots i_n}\sigma(x_1,i_1,\ldots,x_n,i_n)\exp\Big(\sum_{j=1}^n(-1)^{i_j+1}\alpha_j/2\Big) \\ &\times\Big[\sum_{k=1}^n\exp(-1)^{i_k}\alpha_k\Big]\phi_{i_1}^*(x_1)\cdots\phi_{i_n}^*(x_n)|0\rangle \\ &= m\int dx_1\cdots\int dx_n\chi(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n) \\ &\times\sum_{k=1}^n\big[\psi^*(\alpha_1,x_1)\cdots\{e^{-\alpha_k/2}\psi_1^*(\alpha_k,x_k) \\ &+ e^{\alpha_k/2}\psi_2^*(\alpha_k,x_k)\}\cdots\psi^*(\alpha_n,x_n)\big]|0\rangle \\ &= H_F^M|\alpha_1\cdots\alpha_n\rangle, \end{aligned} \tag{4.21}$$

and for the interaction term

$$\begin{aligned} H_B^{\rm int}|\alpha_1,\ldots,\alpha_n\rangle &= 2g\int dx_1\cdots\int dx_n\chi(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n) \\ &\times\sum_{i_1\cdots i_n}\Big[\exp\Big(\sum_{j=1}^n(-1)^{i_j+1}\alpha_j/2\Big)\Big] \\ &\times\sum_k\sum_{l\neq k}\delta(x_k-x_l)[\delta_{1i_k}\delta_{2i_l}+\delta_{1i_l}\delta_{2i_k}] \\ &\times\psi_{i_1}^*(x_1)\cdots\psi_{i_n}^*(x_n)|0\rangle = 4g\int dx_1\cdots \\ &\int dx_n\chi(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n) \\ &\times\sum_k\sum_{l\neq k}\delta(x_k-x_l)(-1)^{k-l+1}\epsilon(k-l) \\ &\times\sinh\tfrac{1}{2}(\alpha_k-\alpha_l)\psi_1^*(x_k)\psi_2^*(x_l) \\ &\times\psi^*(\alpha_1,x_1)\cdots\psi^*(\alpha_nx_n)|0\rangle = H_F^{\rm int}(\alpha_1,\ldots,\alpha_n), \end{aligned} \tag{4.22}$$

and by virtue of Ref. 23 we thus arrive at

$$\begin{aligned} (H_F^{\rm kin}+H_F^m+H_F^{\rm int})|\alpha_1,\ldots,\alpha_n\rangle &= H_F|\alpha_1,\ldots,\alpha_n\rangle \\ &= H_B|\alpha_1,\ldots,\alpha_n\rangle. \end{aligned} \tag{4.23}$$

Needless to say,

$$1_F|\alpha_1,\ldots,\alpha_n\rangle = |\alpha_1,\ldots,\alpha_n\rangle, \tag{4.24}$$

i.e., indeed

$$H_F = 1_FH_B1_F, \quad [H_B,1_F]_- = 0 \tag{4.25}$$

holds true for the continuum field theory, the massive Thirring model.

Let us notice that in contrast to lattice Fermi models we have not made any explicit transformation of Fermi variables into spin $\frac{1}{2}$ type variables. The formula (4.24) proves that such a transformation exists.

Remark 1: By Ref. 4 the spectrum of H_B and H_F is the same in the physical Hilbert space (which is not the Fock space), but the spectrum of H_B appears as infinitely degenerate. As a conjecture we have suggested that the Bose massive Thirring model can be rewritten as a reducible Fermi one.

From the heuristic point of view we find evidence that

$$\begin{aligned} &H_B = \sum_{k=1}^\infty 1_F^kH_B1_F^k = \sum_{k=1}^\infty H_F^k, \\ &1_F^k1_F^l = \delta_{kl}1_F^k, \quad \mathscr{H}_B = \bigoplus_{k=1}^\infty\mathscr{H}_F^k, \quad \mathscr{H}_F^k = 1_F^k\mathscr{H}_B, \end{aligned} \tag{4.26}$$

$$\operatorname{tr}\exp(-iH_Bt) = \sum_{k=1}^\infty\operatorname{tr}\exp(-iH_F^kt).$$

Remark 2: A relationship between the c-number (classical) and Fermi massive Thirring models is thus established as follows.

(1) Take $H_B\equiv H_F(\phi^*,\phi)$ and a coherent state of the Bose spinor field $|\varphi\rangle$ such that $\langle\varphi|\phi|\varphi\rangle = \langle 0|\phi+\varphi|0\rangle = \varphi$ is a classical c-number solution of the MT model field equations.

(2) Construct a separable Hilbert space IDPS($|\varphi\rangle$) by using the CCR algebra generators $\{\phi^*,\phi\}$.

(3) Check (see Refs. 3 and 25) whether $|\varphi\rangle$ allows for the existence of Fermi states in IDPS($|\varphi\rangle$), i.e., that $1_F|\psi\rangle = |\psi\rangle\in$ IDPS($|\varphi\rangle$). If so, then look for eigenvectors of H_B which are Fermi vectors, they are then the eigenvectors of H_F. So the Fermi model appears and the (irreducible) Fermi fields can be introduced.

Remark 3: One must realize that Fermi states of the Bose system are allowed to exist in a very restrictive subset of the set of all non-Fock sectors of the Bose model.[3,25] This restriction follows from the assumption that Fermi states are admitted to arise in Hilbert spaces (in fact in the incomplete direct product ones of von Neumann) which are generated about coherent states of the Bose system. The latter are necessary to satisfy the (weak) correspondence principle

$$\langle\varphi|:H_B(\phi^*,\phi):|\varphi\rangle = H(\varphi^*,\varphi) = H_{\rm classical}, \tag{4.27}$$

where $H_{\rm classical}$ is a classical Hamiltonian of the c-number spinor field satisfying the field equations of the massive Thirring model.[26]

5. MASSIVE THIRRING MODEL: *c*-NUMBER PATH INTEGRAL REFORMULATION AND ALL THAT

To strengthen the above introduced links between the classical c-number and Fermi massive Thirring models, we shall try to derive the path integral expression for $\operatorname{tr}\exp(-iH_Ft)$ in terms of genuine c-number trajectories. We are not aiming at any practical application (for which the Grassmann algebra formulation perfectly suffices); our problems are rather of the foundational nature. The observation that $H_F = 1_FH_B1_F$ suggests that Bose operators in H_B should be replaced by spin $\frac{1}{2}$ objects like in the lattice cases. The lattice spins suitable for our purposes are given by

$$\begin{aligned} &\sigma_i^+(k) = \phi_i^*(k):\exp(-\phi_i^*(k)\phi_i(k)):, \\ &\sigma_i^-(k) = :\exp(-\phi_i^*(k)\phi_i(k)):\phi_i(k), \\ &\quad k = 0,\pm 1,\ldots,\quad i = 1,2, \\ &[\sigma_i^-(k),\sigma_i^+(k)]_+ = 1_F^i(k), \\ &1_F(k) = \prod_{i=1}^2 1_F^i(k),\quad 1_F = \prod_k 1_F(k), \end{aligned} \tag{5.1}$$

where we assume

$$\phi_i^\#(k) = (1/\sqrt{\delta}) \int_{R^1} dx\, \chi_k(x) \phi_i^\#(x),$$

$$\chi_k(x) = \begin{cases} 1, & x \in \Delta_k, \\ 0, & \text{otherwise,} \end{cases} \quad [\phi_i(k), \phi_j^*(p)]_- = \delta_{kp}\delta_{ij}\,, \tag{5.2}$$

δ being the length of each interval Δ_k. Since we assume $\delta \ll 1$, one can formally write $\phi_i^\#(k) \cong \sqrt{\delta}\phi_i^\#(x_k)$, i.e., a reasonable continuum limit would arise after rescaling fields $\phi_i^\#(k)$ by $1/\sqrt{\delta}$ as then $\delta_{kp} \to \delta_{kp}/\sqrt{\delta} \to \delta(x-y)$. It is inconvenient to work explicitly on the continuum level of quantum field theory, hence to avoid inconsistencies we shall use the appropriately discretized model and the transition to continuum will be investigated after achieving the c-number level of the theory.

The mass and interaction terms of H_B we discretize as follows:

$$\begin{aligned} &H_B \to \sum_k H_B(k), \\ &H_B^m(k) = m[\phi_1^*(k)\phi_2(k) + \phi_2^*(k)\phi_1(k)], \\ &H_B^{\text{int}}(k) = (2g/\delta)\phi_1^*(k)\phi_2^*(k)\phi_2(k)\phi_1(k), \end{aligned} \tag{5.3}$$

so that for $\delta \ll 1$ we would formally have

$$\begin{aligned} &H_B^m(k) \cong \delta H_B^m(x_k), \\ &H_B^{\text{int}}(k) \cong \delta H_B^{\text{int}}(x_k). \end{aligned} \tag{5.4}$$

The main difficulty, as usual with Fermi models, comes from the kinetic term, and we shall use a trick which differs from those used in the literature, see for example Ref. 21.

We introduce

$$\begin{aligned} &\phi_i(k,\alpha) = \frac{1}{\sqrt{\delta}} \int_{R^1} \chi_k(x)\phi_i(x+\alpha), \\ &\frac{\partial}{\partial\alpha} \phi_i^*(k)\phi_i(k,\alpha)|_{\alpha=0} \\ &\quad = \frac{1}{\sqrt{\delta}} \phi_i^*(k) \int dx\, \chi_k(x)\partial_x\phi_i(x), \end{aligned} \tag{5.5}$$

which allows for the following computation:

$$\begin{aligned} &\frac{\partial}{\partial\alpha} \sigma_i^*(k)\sigma_i(k,\alpha)|_{\alpha=0}: \\ &\quad = \phi_i^*(k)\text{: } \exp(\phi_i^*(k)\phi_i(k))\text{:} \\ &\qquad \times \frac{\partial}{\partial\alpha}\{\text{: } \exp(-\phi_i^*(k,\alpha)\phi_i(k,\alpha))\text{ :}\phi_i(k,\alpha)\}|_{\alpha=0} \\ &\quad = \phi_i^*(k)\text{: } \exp(-\phi_i^*(k)\phi_i(k))\text{:}\partial_\alpha\phi_i(k,\alpha)|_{\alpha=0} \\ &\qquad - \phi_i^*(k)\text{: } \exp(-\phi_i^*(k)\phi_i(k))\text{: } \{(\partial_\alpha\phi_i^*(k,\alpha)|_{\alpha=0}) \\ &\qquad \times\text{: } \exp(-\phi_i^*(k)\phi_i(k))\text{ : }\phi_i^2(k) + \phi_i^*(k) \\ &\qquad \times\text{:}\exp(-\phi_i^*(k)\phi_i(k))\text{ : }(\partial_\alpha\phi_i(k,\alpha)|_{\alpha=0})\phi_i(k)\} \\ &\quad \equiv \phi_i^*(k)\text{: } \exp(-\phi_i^*(k)\phi_i(k))\text{ : }\partial_\alpha\phi_i(k,\alpha)|_{\alpha=0}\,. \end{aligned} \tag{5.6}$$

Consequently we introduce

$$H_B^{\text{kin}}(k) = -i[\phi_1^*(k)\partial_\alpha\phi_1(k,\alpha) - \phi_2^*(k)\partial_\alpha\phi_2(k,\alpha)]|_{\alpha=0}. \tag{5.7}$$

With such H_B, the formula $H_B(k) \to 1_F(k) H_B(k) 1_F(k)$ arises through replacing all Bose operators by appropriate spin $\frac{1}{2}$ operators:

$$\begin{aligned} &H_B(k) \to H_F(k) \\ &\quad = -i[\phi_1^*(k)\text{: } \exp(-\phi_1^*(k)\phi_1(k))\text{ : }\partial_\alpha\phi_1(k,\alpha)|_{\alpha=0} \\ &\qquad - \phi_2^*(k)\text{: } \exp(-\phi_2^*(k)\phi_2(k))\text{ : }\partial_\alpha\phi_2(k,\alpha)|_{\alpha=0}] \\ &\qquad + m\Big[\phi_1^*(k)\text{: } \exp\Big(-\sum_{j=1}^2 \phi_i^*(k)\phi_i(k)\Big)\text{ : }\phi_1(k) \\ &\qquad + \phi_2^*(k)\text{: } \exp\Big(-\sum_{j=1}^2 \phi_i^*(k)\phi_i(x)\Big)\text{ : }\phi_2(k)\Big] \\ &\qquad + \frac{2g}{\delta}\phi_1^*(k)\phi_2^*(k)\text{:} \\ &\qquad \times \exp\Big(-\sum_{j=1}^2 \phi_j^*(k)\phi_j(k)\Big)\text{ : }\phi_2(k)\phi_1(k). \end{aligned} \tag{5.8}$$

An infinitesimal propagator for $H_F(k)$ reads

$$\hat{U}_F^k(\Delta t) = \exp(iH_F(k)\Delta t) \cong 1_F(k) - i\Delta t H_F(k), \tag{5.9}$$

and its functional kernel is

$$\begin{aligned} U_F^k(\Delta t) &= \Big[\exp \sum_{j=1}^2 \beta_j^*(k)\beta_j(k)\Big] \cdot \langle\beta|U_F^k(\Delta t)|\beta\rangle \\ &= \{-i[\beta_1^*(k)\partial_\alpha\beta_1(k,\alpha)|_{\alpha=0} \exp\beta_2^*(k)\beta_2(k) \\ &\quad - \beta_2^*(k)\partial_\alpha\beta_2(k,\alpha)|_{\alpha=0} \exp\beta_1^*(k)\beta_1(k)] \\ &\quad + m[\beta_1^*(k)\beta_2(k) + \beta_2^*(k)\beta_1(k)] \\ &\quad + (2g/\delta)|\beta_1(k)|^2|\beta_2(k)|^2\}(-i\Delta t) \\ &\quad + \Big\{1 + \sum_{i=1}^2 \beta_i^*(k)\beta_i(k) + |\beta_1(k)|^2|\beta_2(k)|^2\Big\} \\ &:= (1 - i\Delta t H(k)/\Pi_{i=1,2}(1 + |\beta_i(k)|^2)) \\ &\quad \times \exp\sum_{i+1}^2 \ln(1 + |\beta_i(k)|^2) \\ &\cong \exp\Big\{\sum_{i=1}^2 \ln[1 + \beta_i^*(k)\beta_i(k)] \\ &\quad - i\Delta t H(k)/\Pi_{i=1,2}(1 + \beta_i^*(k)\beta_i(k))\,. \end{aligned} \tag{5.10}$$

We compose a product of such kernels for all sites: $U_F = \Pi_k U_F^k$, and by repeating the Fermi oscillator arguments, we arrive at the following expression on a lattice (time is continuous):

$$\begin{aligned} \text{tr}\exp(-iH_F t) = \int [d\beta][d\beta^*] \exp\sum_k i\int_0^t dt \\ \times\Big[\sum_{i=1}^2 i\frac{\beta_i^*(k)\dot{\beta}_i(k)}{1 + \beta_i^*(k)\beta_i(k)} \\ - \frac{H(k)}{\Pi_{i=1,2}(1 + \beta_i^*(k)\beta_i(k))}\Big]. \end{aligned} \tag{5.11}$$

If now to notice that the coherent state representation of $\phi_i(k)$ reads

$$\beta_i(k) = \frac{1}{\sqrt{\delta}} \int \chi_k(x)\beta_i(x)dx \cong \sqrt{\delta}\beta_i(x) \quad x \in \Delta_k\,, \tag{5.12}$$

we arrive at

$$\mathrm{tr}_\delta \exp(-iH_F t)$$

$$= \int [d\beta][d\beta^*] \exp \sum_k i \int_0^t dt$$

$$\times \Big\{ \sum_{j=1}^{2} i\delta \frac{\beta_j^*(x)\dot\beta_j(x)}{1+\delta\beta_j^*(x)\beta_j(x)} - [(-i\delta)(\beta_1^*(x)$$

$$\times \partial\beta_1(x) \exp(\delta\beta_2^*(x)\beta_2(x)) - \beta_2^*(x)\partial\beta_2(x)$$

$$\times \exp(\delta\beta_1^*(x)\beta_1(x))) + m\delta(\beta_1^*(x)\beta_2(x) + \beta_2^*(x)\beta_1(x))$$

$$+ 2g\delta|\beta_1(x)|^2|\beta_2(x)|^2] \Big/ \prod_{i=1,2}(1+\delta\beta_i^*(x)\beta_i(x))\Big\}. \tag{5.13}$$

Under an assumption that we restrict path integrations to these c-number paths only for which $\sigma_i(x) = \beta_i^*(x)\beta_i(x) < \infty$ on the whole space axis, we can consistently achieve a continuum limit. Then contributions from factors $\exp(\delta\beta_i^*(x)\beta_i(x))$, $\Pi_{i=1}^2(1+\delta\beta_i^*(x)\beta_i(x))$ become negligible; if compared with 1; hence

$$\mathrm{tr}_A \exp(-iH_F t)$$

$$= \int_{\{\sigma_i \le A < \infty\}} [d\beta][d\beta^*] \exp i \int_0^t dt \int dx$$

$$\times \Big\{ \sum_{j=1}^{2} i\beta_j^*(x)\dot\beta_j(x) - [(-i)(\beta_1^*\partial\beta_2 - \beta_2^*\partial\beta_1)$$

$$+ m(\beta_1^*\beta_2 + \beta_2^*\beta_1) + 2g|\beta_1|^2|\beta_2|^2](x)\Big\}$$

$$= \int_{\{\sigma_i \le A < \infty\}} [d\beta][d\beta^*] \exp i \int_0^t dt \int dx$$

$$\times \Big\{ \sum_{j=1}^{2} i\beta_j^*(x)\dot\beta_j(x) - H_{\text{classical}}(x)\Big\}. \tag{5.14}$$

It does not overcome problems with the continuum limit if no restrictions on trajectories are imposed, but at the same time it selects a subset of trajectories on which these problems disappear. Then the Fermi and Bose formula for the trace has an identical contribution from this subset, and the solutions of the original classical field equations do make stationary the action for both cases. Because of the importance of such stationary points in the semiclassical physics, we conclude that it makes sense to talk about the quantum meaning of classical c-number spinor fields in the Fermi quantized case.

6. REMARKS ON THE CHIRAL INVARIANT GROSS–NEVEU MODEL

The Hamiltonian density of the model reads

$$H_F(x) = -i \sum_{a=1}^{N} (\psi_{a+}^* \partial\psi_{a+} - \psi_{a-}^*\partial\psi_{a-})$$

$$+ \sum_{a,b}^{N} 4g\psi_{a+}^*\psi_{b-}^*\psi_{b+}\psi_{a-}. \tag{6.1}$$

The Hamiltonian H_F is diagonalizable in the Fock space of the Fermi fields $\psi_{\alpha a}(x)$, $a = 1,2,\ldots, N$, $\alpha = \pm$,

$$[\psi_{\alpha a}(x), \psi_{\beta b}^*(y)]_+ = \delta_{\alpha\beta}\delta_{ab}\delta(x-y),$$
$$\psi_{\alpha a}(x)|0\rangle = 0, \quad \forall\ \alpha, a, x. \tag{6.2}$$

The respective eigenvectors have the general form[27]

$$|F,\xi\rangle = \int dx_1 \cdots \int dx_n \sum_{\{\alpha,a\}} F(x_1,\ldots,x_n,\alpha_1,\ldots,\alpha_n)$$

$$\cdot\, \xi(a_1,\ldots,a_n) \prod_{i=1}^{n} \psi_{a_i\alpha_i}^*(x_i)|0\rangle, \tag{6.3}$$

where $F(x,\alpha)$ is an eigenfunction of the n-particle Hamiltonian

$$h = -i\sum_{j=1}^{n} \alpha_j\partial_j - 4g\sum_{i,j}\delta(x_i - x_j)P^{ij}\Big[\frac{1}{2}(1-\alpha_i\alpha_j)\Big]. \tag{6.4}$$

P^{ij} is an operator which exchanges chiralities α_i and α_j. Let us assume that the CAR algebra representation (6.2) is embedded in the CCR algebra representation, as in Ref. 24. For this purpose we need an antisymmetric function

$$\sigma = \sigma(x_1,\alpha_1,a_1,\ldots,x_n,\alpha_n,a_n) = \sigma^3 \tag{6.5}$$

which changes a sign if triplets (x,α,a) in $\{(x,\alpha,a)\}$ are interchanged and takes the values 0, ± 1. Then, we arrive at

$$\prod_{i=1}^{n} \psi_{a_i\alpha_i}^*(x_i)|0\rangle$$

$$= \sigma(x_1,\alpha_1,a_1,\ldots,x_n,\alpha_n,a_n)\phi_{a_1\alpha_1}^*(x_1)\cdots\phi_{a_n\alpha_n}^*(x_n)|0\rangle, \tag{6.6}$$

where ϕ, ϕ^* are the corresponding Bose operators (with the same iso-indices as the fermions).

If $F(x,\alpha)$ is an eigenfunction of h, then

$$F^a(x,\alpha) = F(x,\alpha)\cdot\sigma(x,\alpha,a) \tag{6.7}$$

is an eigenfunction again, because formally

$$\partial_i\sigma = \partial_i\sigma^{2n+1} = (2n+1)\sigma^{2n}\partial_i\sigma = (2n+1)\sigma^2\partial_i\sigma$$
$$\Rightarrow \partial_i\sigma \equiv 0, \tag{6.8}$$

which yields (up to a set of measure zero)

$$-i\sum_j \alpha_j\partial_j F^a(x,\alpha) \equiv \Big(-i\sum_j\alpha_j\partial_j F\Big)\cdot\sigma. \tag{6.9}$$

On the other hand

$$\sum_{ij}\delta(x_i-x_j)P^{ij}\Big[\frac{1}{2}(1-\alpha_i\alpha_j)\Big]F^a(x_1\alpha)$$

$$= \sum_{ij}\delta(x_i-x_j)P^{ij}\Big[\frac{1}{2}(1-\alpha_i\alpha_j)\Big]$$

$$\times F(\cdots x_i\cdots x_j,\ldots,\ldots,\alpha_i\cdots\alpha_j\cdots)$$

$$\times\sigma(\cdots x_i\cdots x_j\cdots\alpha_i\cdots\alpha_j\cdots)$$

$$= \Big\{\sum_{ij}\delta(x_i-x_j)P^{ij}\Big[\frac{1}{2}(1-\alpha_i\alpha_j)\Big]F\Big\}\cdot\sigma \tag{6.10}$$

provided σ is symmetric under an interchange of α_i and α_j at $x_i = x_j$ and $(a_1,\ldots,a_n)$ fixed. As an example of such σ we propose

$$\sigma(x_1,a_1,\alpha_1,\ldots,x_n,a_n,\alpha_n) = \prod_{1\le j<k\le n} p_{jk},$$

$$p_{jk} = \delta_{\alpha_j\alpha_k}\delta_{a_ja_k}[\Theta(x_j-x_k) - \Theta(x_k-x_j)]$$

$$+ \delta_{a_ja_k}\Theta(|\alpha_j-\alpha_k|)(-1)^{1+\Theta(x_j-x_k)}$$

$$+ \delta_{\alpha_j\alpha_k}\Theta(|a_j-a_k|)(-1)^{\Theta(x_j-x_k)}$$

$$+ (1-\delta_{\alpha_j\alpha_k})(1-\delta_{a_ja_k})(-1)^{\Theta(x_j-x_k)}. \tag{6.11}$$

Notice that at $x_i = x_j$ we have

$$p_{ik}(x_j = x_k) = \delta_{a_j a_k}\Theta(|\alpha_j - \alpha_k|) - \delta_{\alpha_j\alpha_k}\Theta(|a_j - a_k|) + (1-\delta_{\alpha_j\alpha_k})(1-\delta_{a_j a_k})(-1). \quad (6.12)$$

One should notice also that an eigenvalue problem for h arises from an eigenvalue problem for H_F after commuting the annihilation operators to $|0\rangle$ through the product of ψ^*'s and then integrating the gradients by parts. If now to use the Bose CGN Hamiltonian $H_B = H_F(\psi^* \to \phi^*, \psi \to \phi)$ the procedure is exactly the same and the eigenvalue problem for h arises again, but with a wave function $F^a(x, \alpha)$ instead of $F(x, \alpha)$.

Consequently the Bethe ansatz eigenvectors of H_F are also eigenvectors of H_B, and satisfy the property $1_F|F, \xi\rangle = |F, \xi\rangle$, where 1_F is the operator unit of the *bosonized* Fermi algebra. All the arguments applied before to the massive Thirring model apparently apply in the CGN case, and for example it is not difficult to check that the coherent state expectation value of the *bosonized* Fermi Hamiltonian reads

$$\begin{aligned}\langle\beta|1_F H_B 1_F|\beta\rangle &\cong \sum_k \Big\{\delta(-i)\sum_a [\beta^*_{a+}\exp(-\delta\beta^*_{a+}\beta_{a+})\partial\beta_{a+} \\ &\quad - \beta^*_{a-}\exp(-\delta\beta^*_{a-}\beta_{a-})\partial\beta_{a-}] \\ &\quad + 4g\delta\sum_{ab}\beta^*_{a+}\beta^*_{b-}\exp\Big(-\delta\sum_\alpha[\beta^*_{a\alpha}\beta_{a\alpha} \\ &\quad + \beta^*_{b\alpha}\beta_{b\alpha}]\Big)\beta_{b+}\beta_{b-} \underset{\delta\to 0}{\to} \int dx \\ &\quad \times\Big\{\sum_a(-i)(\beta^*_{a+}\partial\beta_{a+} - \beta^*_{a-}\partial\beta_{a-} \\ &\quad + 4g\sum_{ab}\beta^*_{a+}\beta^*_{b-}\beta_{b+}\beta_{a-}\Big\}\end{aligned} \quad (6.13)$$

for all c-number spinor functions which are regular enough, i.e., satisfy $\sigma_{\alpha a}(x) = \beta^*_{\alpha a}(x)\beta_{\alpha a}(x) \leqslant A < \infty$. The respective solutions of the classical CGN model field equations thus have a quantum meaning both in the Bose and Fermi quantization cases (see in this connection also Ref. 28).

[1] P. Garbaczewski, J. Math. Phys. **19**, 642 (1978).
[2] P. Garbaczewski, J. Math. Phys. **23**, 442 (1982).
[3] P. Garbaczewski, J. Math. Phys. **24**, 341 (1983).
[4] P. Garbaczewski, J. Math. Phys. **24**, 1806 (1983).
[5] R. Jackiw, Rev. Mod. Phys. **49**, 682 (1977).
[6] R. Dashen, B. Hasslacher, and A. Neveu, Phys. Rev. D **12**, 2443 (1975).
[7] S. Shei, Phys. Rev. D **14**, 535 (1976).
[8] S. Y. Lee and A. Gavrielides, Phys. Rev. D **12**, 3880 (1975).
[9] J. R. Klauder, Ann. Phys. (NY) **11**, 123 (1960).
[10] F. A. Berezin, *Methods of Second Quantization* (Nauka, Moscow, 1965).
[11] L. D. Faddeev, and A. A. Slavnov, *Introduction to Quantum Theory of Gauge Fields* (Nauka, Moscow, 1978).
[12] A. Jevicki and N. Papanicolau, Ann. Phys. (NY) **120**, 107 (1979).
[13] L. Schulman, Phys. Rev. **176**, 1558 (1970).
[14] J. M. Radcliffe, J. Phys. A **4**, 413 (1971).
[15] F. T. Arecchi, E. Courtens, R. Gilmore, and H. Thomas, Phys. Rev. A **6**, 2210 (1972).
[16] J. R. Klauder, J. Math. Phys. **23**, 1797 (1982).
[17] G. F. De Angelis, D. de Falco, and F. Guerra, Phys. Rev. D **23**, 1747 (1981).
[18] D. de Falco *et al.*, Phys. Lett. B **109**, 54 (1982).
[19] Ph. Combe, R. Rodriguez, M. Sirugue-Collin, and M. Sirugue, J. Math. Phys. **20**, 611 (1979).
[20] J. B. Zuber and C. Itzykson, Phys. Rev. D **15**, 2875 (1977).
[21] L. Susskind, Phys. Rev. D **16**, 3031 (1977).
[22] M. Lüscher, Nucl. Phys. B **117**, 476 (1976).
[23] H. Bergknoff and H. B. Thacker, Phys. Rev. D **19**, 3666 (1979).
[24] P. Garbaczewski, Commun. Math. Phys. **43**, 131 (1975).
[25] P. Garbaczewski, "Representations of the CAR generated by representations of the CCR," Wroclaw University preprint (unpublished).
[26] V. A. Kuznetsov and A. V. Mikhailov, Teor. Mat. Fiz. **30**, 323 (1977).
[27] N. Andrei and J. H. Lowenstein, Phys. Rev. Lett. **43**, 1698 (1979).
[28] P. Garbaczewski, Ann. Phys (NY) **150**, 22 (1983).

Evaluation of an integral involving associated Legendre polynomials and inverse powers of $(1-X^2)$

Nazakat Ullah
Tata Institute of Fundamental Research, Bombay 400 005, India

(Received 3 January 1983; accepted for publication 8 April 1983)

An integral involving the product of two associated Legendre polynomials and inverse powers of $(1-X^2)$ is carried out using the operator form of Taylor's theorem and an integral over a single associated Legendre polynomial.

PACS numbers: 02.90. + p, 14.40. − n, 23.20.Js

I. INTRODUCTION

The integrals involving associated Legendre polynomials occur in many branches of physics.[1] Since associated Legendre polynomials $P_l^m(X)$ multiplied by $(1-X^2)$ are polynomials in X, in principle such integrals can easily be evaluated involving double or triple sums. However, it is not an easy matter to resum such series into a simpler form. Because of this, various other techniques are used to evaluate these integrals which directly give simpler forms. The purpose of the present work is to describe a new technique, based on the operator form of Taylor's theorem, to evaluate integrals involving the product of two associated Legendre polynomials and inverse powers of $(1-X^2)$. Laursen and Mita[2] have recently pointed out that such integrals are needed in meson physics and have evaluated some of them using special properties of hypergeometric functions. Since associated Legendre polynomials can be coupled using vector-coupling coefficients,[1] such integrals can be used to derive additional sum rules for the vector-coupling coefficients.[3] We present this formulation in Sec. II. Concluding remarks are presented in Sec. III.

II. FORMULATION

Let us consider the integral

$$I_p = \int_{-1}^{1} dx \frac{[P_l^m(x)P_k^n(x)]}{(1-X^2)^{p+1}}, \tag{1}$$

with l, m, k, n, p being positive integers, $l \geqslant m$, $k \geqslant n$, $p \leqslant (m+n-1)/2$, and $l+k-(m+n)$ even. We shall use Taylor's theorem to write it as an integral over a single associated Legendre polynomial. The operator form of this theorem is written as

$$f(x+h) = \exp(hD)f(x), \tag{2}$$

where $D = d/dX$ is the differential operator. Using expression (2), the integral in (1) can be rewritten as

$$I_p = \int_{-1}^{1} dx(1-X^2)^{-n/2-p-1}P_l^m(x) \times \exp\left[(1-X^2)\frac{\partial}{\partial\lambda}\right]\lambda^{n/2}P_k^n(\sqrt{1-\lambda})\Big|_{\lambda=0}. \tag{3}$$

The function $\lambda^{m/2}P_l^m(\sqrt{1-\lambda})$ can be written in terms of a hypergeometric function using the relation[4]

$$F(a,b;a+b+\tfrac12;X) = 2^{a+b-1/2}\Gamma(\tfrac12+a+b) \times X^{1/2(1/2-a-b)}P_{a-b-1/2}^{1/2-a-b}(\sqrt{1-X}). \tag{4}$$

Substituting expression (4) in expression (3), expanding the exponential, and making use of the integral[4]

$$\int_0^\pi dt\,(\sin t)^{\alpha-1}P_\nu^{-\mu}(\cos t) = 2^{-\mu}\pi\Gamma(\tfrac12\alpha+\tfrac12\mu) \times \Gamma(\tfrac12\alpha-\tfrac12\mu)[\Gamma(\tfrac12+\tfrac12\alpha+\tfrac12\nu)\Gamma(\tfrac12\alpha-\tfrac12\nu) \times \Gamma(\tfrac12\mu+\tfrac12\nu+1)\Gamma(\tfrac12\mu-\tfrac12\nu+\tfrac12)]^{-1}; \tag{5}$$

$$\mathrm{Re}(\alpha\pm\mu)>0,$$

we get

$$I_p = (-2)^{m+n}\pi[\Gamma((k-n+1)/2)\Gamma(-(k+n)/2) \times \Gamma((-l-m+1)/2)\Gamma((l-m+2)/2)]^{-1} \times \sum_{t=0}^{\infty}\frac{1}{t!}\frac{\Gamma((k+n+1)/2+t)\Gamma((-k+n)/2+t)}{\Gamma(n+1+t)\Gamma((n-2p+l+1)/2+t)} \times \frac{\Gamma((n-2p-m)/2+t)\Gamma((n-2p+m/2)+t)}{\Gamma((n-2p-l)/2)}. \tag{6}$$

By using specific values of P_l^m in expression (4) we find that a phase factor has to be added to it. Similarly, there is a phase factor in expression (5). These phases are added to expression (6). Further in the derivation, we have assumed $l+m$ to be even; this restriction will be dropped in the final result. To arrive at the final expression, we express the sum over t as a generalized hypergeometric series. The final expression for I_p is given by

 0022-2488/84/040872-02$02.50

$$I_p = (-2)^{m+n}\pi\Big[\Gamma\Big(\frac{k-n+1}{2}\Big)\Gamma\Big(-\frac{k+n}{2}\Big)$$

$$\times\Gamma\Big(\frac{-l-m+1}{2}\Big)\Gamma\Big(\frac{l-m+2}{2}\Big)\Gamma(n+1)$$

$$\times\Gamma\Big(\frac{n-2p+l+1}{2}\Big)\Gamma\Big(\frac{n-2p-l}{2}\Big)\Big]^{-1}$$

$$\times\Big[\Gamma\Big(\frac{k+n+1}{2}\Big)\Big(\frac{-k+n}{2}\Big)$$

$$\times\Gamma\Big(\frac{-2p+n-m}{2}\Big)\Gamma\Big(\frac{-2p+n+m}{2}\Big)\Big]$$

$$\times{}_4F_3\Big(\frac{k+n+1}{2},\frac{-k+n}{2},$$

$$\frac{-2p+n-m}{2},\frac{-2p+n+m}{2};$$

$$n+1,\frac{n-2p+l+1}{2},\frac{n-2p-l}{2};1\Big), \tag{7}$$

where ${}_4F_3$ is the generalized hypergeometric series.

For the special case in which one has squares of associated Legendre polynomials, we get by putting $k=l$ and $n=m$, the following expression:

$$I_p = \frac{(l+m)!(m-p-1)!\Gamma\Big(\frac{l+m+1}{2}\Big)\Gamma\Big(\frac{l-m+2}{2}+p\Big)}{(l-m)!m!\,p!\Gamma\Big(\frac{l+m+1}{2}-p\Big)\Gamma\Big(\frac{l-m+2}{2}\Big)}$$

$$\times{}_4F_3\Big(\frac{l+m+1}{2},-\frac{l-m}{2},m-p,-p;m+1,$$

$$\frac{l+m+1-2p}{2},\frac{-l+m-2p}{2};1\Big). \tag{8}$$

As a simple check on expression (8) we find by setting $p=0$, that

$$I_0 = \frac{(l+m)!}{(l-m)!}\frac{l}{m}, \tag{9}$$

the same value as given in Ref. 4.

III. CONCLUDING REMARKS

We have shown that a new technique based on the operator form of Taylor's theorem can be used to evaluate integrals involving associated Legendre polynomials and inverse powers of $(1-X^2)$. The technique enables us to express this integral in terms of the generalized hypergeometric series ${}_4F_3$.

[1]D. M. Brink and G. R. Satchler, *Angular Momentum* (Oxford U.P., Princeton, 1962); A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).

[2]M. L. Laursen and K. Mita, J. Phys. A: Gen. Phys. **14**, 1065 (1981).

[3]N. Ullah, Indian J. Phys. **56A**, 296 (1982).

[4]M. Abramowitz and A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), Chaps. 8, 15.

A projective constrained variational principle for a classical particle with spin

R. Amorim[a)]
Universidade Federal do Rio de Janeiro, Instituto de Física, Cidade Universitária-Ilha do Fundão, Rio de Janeiro-CEP.: 21.944-RJ, Brazil

(Received 24 June 1983; accepted for publication 28 October 1983)

A geometric approach for variational principles with constraints is applied to obtain the equations of motion of a classical charged point particle with magnetic moment interacting with an external eletromagnetic field.

PACS numbers: 03.20. + i, 03.30. + p

I. INTRODUCTION

In this paper we obtain a set of equations of motion for a classical charged point particle with spin and magnetic dipole moment. These equations are completely consistent and are generated according to a projective Lagrangian formalism introduced by C. M. Amaral and P. Pitanga.[1] This formalism is based on a geometric view of the constraints, which represent hypersurfaces in the space of coordinates and velocities. The basic assumption is that the physical movement of the representative point belongs to the intersection of those hypersurfaces.

The nonintegrable condition of orthogonality between the spin tensor and the four-velocity is conserved by the theory. The equations of motion do not have, however, the spurius helicoidal solutions usually present in these theories[2–5] and that appear even in the free field case. These helicoidal motions are usually eliminated when one assumes the orthogonality between the spin and the mechanical movement.[6–8] In the present theory both conditions are equivalent since the momentum and the velocity are parallel.

We could obtain the same set of equations[9] by using in a correct form the Lagrangian formalism with nonintegrable constraints via Lagrangian multipliers.[10,11]

In Sec. II we define the space of coordinates and velocities for a particle with spin as well as the form of the constrained variational principle. In Sec. III we propose a specific Lagrangian so as to obtain the equations of motion. In Sec. IV some conclusions are pointed out.

II. THE SPACE OF VELOCITIES AND COORDINATES

Consider a classical point particle with spin described by a set of four coordinates

$$x^\mu = x^\mu(\tau) \tag{2.1}$$

related to an arbitrary Lorentz frame. The parameter τ is the proper time, in the sense that

$$u^\mu = \frac{dx^\mu}{d\tau} \tag{2.2}$$

satisfies the canonical condition[12]

$$u^\mu u_\mu = 1. \tag{2.3}$$

We can introduce additional degrees of freedom defining two spacelike four-vectors

$$b^{(i)\mu} = b^{(i)\mu}(\tau), \quad i = 1,2$$

on the trajectory of the particle. Later they will be associated with the spin. We assume that these vectors are normalized to -1 and are orthogonal to each other and to the velocity u^μ.

We could introduce a third spacelike four-vector $b^{(3)\mu}$ to complete a vierbein field, where $b^{(0)\mu}$ is identified with the four-velocity u^μ. This is not necessary, however, because the spin tensor is antisymmetric [see expression (3.3)] and its six non-null components can be described only by $b^{(1)\mu}$ and $b^{(2)\mu}$ put in a convenient form.

The conditions of orthonormality satisfied by $b^{(i)\mu}$ and u^μ are expressed by the constraints

$$\phi^{(1)} = b^{(1)\mu}u_\mu = 0, \tag{2.4a}$$

$$\phi^{(2)} = b^{(2)\mu}u_\mu = 0, \tag{2.4b}$$

$$\tilde\phi^{(3)} = b^{(1)\mu}b^{(1)}{}_\mu + 1 = 0, \tag{2.4c$'$}$$

$$\tilde\phi^{(4)} = b^{(2)\mu}b^{(2)}{}_\mu + 1 = 0, \tag{2.4d$'$}$$

$$\tilde\phi^{(5)} = b^{(1)\mu}b^{(2)}{}_\mu = 0. \tag{2.4e$'$}$$

We can use an equivalent set of constraints, replacing (2.4c′)–(2.4e′) by

$$\phi^{(3)} = b^{(1)\mu}\frac{db^{(1)}{}_\mu}{d\tau} = 0, \tag{2.4c}$$

$$\phi^{(4)} = b^{(2)\mu}\frac{db^{(2)}{}_\mu}{d\tau} = 0, \tag{2.4d}$$

$$\phi^{(5)} = b^{(1)\mu}\frac{db^{(2)}\mu}{d\tau} + \frac{db^{(1)\mu}}{d\tau}b^{(2)}{}_\mu = 0. \tag{2.4e}$$

The constraints (2.4) represents five hypersurfaces in the space of coordinates and velocities (CV-space). CV-space has dimension $D = 24$ and is spanned by all the representative 24-dimension "position vectors" of the particle

$$|R\rangle = \left(x^\mu, b^{(1)\mu}, b^{(2)\mu}; \frac{dx^\mu}{d\tau}, \frac{db^{(1)\mu}}{d\tau}, \frac{db^{(2)\mu}}{d\tau}\right) \equiv \left(\chi^{(i)\mu}; \frac{d\chi^{(i)\mu}}{d\tau}\right), \tag{2.5}$$

where

$$\chi^{(0)\mu} = x^\mu,\ \chi^{(1)\mu} = b^{(1)\mu},\ \text{and}\ \chi^{(2)\mu} = b^{(2)\mu}.$$

The metric of CV-space is

$$g^{mn} = g_{mn} = \operatorname{diag}(+ - - -, \dots, + - - -), \tag{2.6}$$

since $\chi^{(i)\mu}$ and $d\chi^{(i)\mu}/d\tau$ are Lorentz four-vectors.

The physical trajectories of the representative point in

[a)] Partially supported by FINEP.

CV-space must belong to the intersection of the hypersurfaces $\phi^{(I)}=0$. Consequently the tangent vectors of these trajectories have to be orthogonal to each one of the vectors

$$|e^I\rangle = \overset{\mathrm{CV}}{\nabla}\phi^I = \left(\frac{\partial\phi^I}{\partial\chi^{(i)\mu}};\frac{\partial\phi^I}{\partial(d\chi^{(i)\mu}/d\tau)}\right). \tag{2.7}$$

Because $\phi^I=0$ are five independent conditions, we can define a local five-dimensional metric

$$g^{IJ}=\langle e^I|e^J\rangle \tag{2.8}$$

and its inverse g_{IJ}.

The projector[1]

$$\Lambda = 1 - g_{IJ}\langle e^I|e^J\rangle \tag{2.9}$$

will subtract from an arbitrary vector belonging to CV-space its linearly dependent part with the vectors $|e^I\rangle$.

In this way, if we define

$$|\bar{A}\rangle = \Lambda|A\rangle, \tag{2.10}$$

we see that

$$\langle e^J|\bar{A}\rangle = 0,\quad J=1,2,...,5. \tag{2.11}$$

We will now see how to apply these ideas to the variational principle. In its free form, we have

$$\delta A = \int_{\lambda_1}^{\lambda_2}\delta L\,d\lambda$$

$$=\int_{\lambda_1}^{\lambda_2}\left\{\frac{\partial L}{\partial\chi^{(i)\mu}}\delta\chi^{(i)\mu}+\frac{\partial L}{\dot{\chi}^{(i)\mu}}\delta\dot{\chi}^{(i)\mu}\right\}d\lambda=0, \tag{2.12}$$

where $\lambda=\lambda(\tau)$ is an arbitary function of proper time and the dot represents $d/d\lambda$.

By assuming a null variation at the end points, and by making an integration by parts, we see that

$$\delta A = -\int_{\lambda_1}^{\lambda_2} E^{(i)}_{\mu}\delta\chi^{(i)\mu}=0, \tag{2.13a}$$

where

$$E^{(i)}_{\mu}=\frac{d}{d\lambda}\frac{\partial L}{\partial\dot{\chi}^{(i)\mu}}-\frac{\partial L}{\partial\chi^{(i)\mu}} \tag{2.13b}$$

is the Euler vector. We assume the sum convention for the μ and (i) indices.

Equations (2.13) are equivalent to

$$E^{(i)}_{\mu}\frac{\delta\chi^{(i)\mu}}{\delta\lambda}=0 \tag{2.14a}$$

for an arbitrary $\delta\chi^{(i)\mu}$.

In CV-space this corresponds to the equation

$$\left\langle\frac{\delta\chi}{\delta\lambda}\middle|E\right\rangle=0, \tag{2.14b}$$

if we define the 24-dimensional Euler vector

$$|E\rangle = (0,0,0;E^{(0)\mu},E^{(1)\mu},E^{(2)\mu}). \tag{2.14c}$$

We note that this vector has non-null components only in V-space, i.e., the subspace of CV-space spanned by the arbitrary

$$|R\rangle = \left(0;\frac{d\chi^{(i)\mu}}{d\tau}\right). \tag{2.15}$$

To introduce the constraints, we replace the free variation $\delta\chi$ by

$$\bar{\delta}\chi = \Lambda\delta\chi, \tag{2.16}$$

and the corresponding one to Eq. (2.14b) is now

$$\left\langle\frac{\bar{\delta}\chi}{\delta\lambda}\middle|E\right\rangle=0, \tag{2.17a}$$

which is equivalent to

$$\left\langle\frac{\delta\chi}{\delta\lambda}\middle|\Lambda^TE\right\rangle=0, \tag{2.17b}$$

which implies that

$$\Lambda^T|E\rangle = 0 \tag{2.18}$$

since the $\delta\chi$ are all independent.

$|E\rangle$ belongs to the V-space and it is, however, enough to calculate the projector Λ^T in this space. In this way, we will list only the V-projection of $|e^I\rangle$, $I=1$ to 5.

From (2.7) and (2.4) they are

$$\begin{aligned}|e^1\rangle &= (b^{(1)\mu},0,0),\\ |e^2\rangle &= (b^{(2)\mu},0,0),\\ |e^3\rangle &= (0,b^{(1)\mu},0),\\ |e^4\rangle &= (0,0,b^{(2)\mu}),\\ |e^5\rangle &= \frac{1}{\sqrt{2}}(0,b^{(2)\mu},b^{(1)\mu}),\end{aligned} \tag{2.19}$$

where the $|e^I\rangle$ are presented in a normalized form.

It is trivial to see that

$$g^{IJ}=\langle e^I|e^J\rangle = -\delta^{IJ}, \tag{2.20}$$

and so $g_{IJ}=g^{IJ}$.

The projector

$$Q = 1-\Lambda \tag{2.21}$$

is

$$(Q^{(i)(j)}_{\mu\nu}) = \begin{bmatrix} b^1b^1+b^2b^2 & 0 & 0\\ 0 & b^1b^1+\dfrac{b^2b^2}{2} & \dfrac{b^2b^1}{2}\\ 0 & \dfrac{b^1b^2}{2} & b^2b^2+\dfrac{b^1b^1}{2}\end{bmatrix}_{\mu\nu}. \tag{2.22}$$

We note that

$$(Q^{(i)(j)}_{\mu\nu})^T = Q^{(j)(i)}_{\nu\mu} = Q^{(i)(j)}_{\mu\nu},\ \text{i.e.,}$$

$$Q=Q^T. \tag{2.23}$$

III. EQUATIONS OF MOTION

To generate equations of motion for a classical point particle with spin in an external electromagnetic field, we choose[13]

$$L=(m_0+k/2S^{\mu\nu}F_{\mu\nu})(v^2)^{1/2}$$
$$+eA_\mu v^\mu + (h/2)(\dot{b}^{(i)\mu}b^{(2)}_\mu - b^{(1)\mu}\dot{b}^{(2)}_\mu), \tag{3.1}$$

which is the simplest Lagrangian that is homogeneous in $v^\mu = dx^\mu/d\lambda$ describing the interaction of a charge e with a magnetic dipole moment

$$M^{\mu\nu}=kS^{\mu\nu} \tag{3.2}$$

with the electromagnetic field. Although not present in Ref. 13, the included $\dot{b}^{(i)\mu}$ terms are necessary to generate the precession equation for the spin, defined as

$$S^{\mu\nu} = h\,(b^{(1)\mu} b^{(2)\nu} - b^{(1)\nu} b^{(2)\mu}), \tag{3.3}$$

where h is a constant related to the spin modulus.

Those terms represent the energy of proper rotation of the particle.[4]

Let us obtain the precession equations. The Euler vectors for $b^{(i)\mu}$ are

$$E^{(1)\mu} = h\,(\dot{b}^{(2)\mu} - kF^{\mu\nu} b^{(2)}_{\nu} (v^2)^{1/2}), \tag{3.4a}$$

$$E^{(2)\mu} = -h\,(\dot{b}^{(i)\mu} - kF^{\mu\nu} b^{(1)}_{\nu} (v^2)^{1/2}), \tag{3.4b}$$

and Eqs. (2.18), (2.22), and (2.23) imply that

$$E^{(1)\mu} + (b^1 b^1 + b^2 b^2/2)^{\mu\nu} E^{(1)}_{\nu} + (b^2 b^1/2)^{\mu\nu} E^{(2)}_{\nu} = 0. \tag{3.5}$$

From (3.5) and the constraints, we get

$$\frac{db^{(2)\mu}}{d\tau} = kF^{\mu\nu} b^{(2)}_{\nu}. \tag{3.6a}$$

In an analogous way, we see that

$$\frac{db^{(1)\mu}}{d\tau} = kF^{\mu\nu} b^{(1)}_{\nu} \tag{3.6b}$$

and from (3.3) we get

$$\frac{dS^{\mu\nu}}{d\tau} = 2kF^{[\mu}_{\alpha} S^{\nu]\alpha}, \tag{3.7}$$

where μ and ν appear in an antisymmetric form.

From (3.7) we see that the spin modolus is conserved, i.e.,

$$\frac{d}{d\tau}(S^{\mu\nu} S_{\mu\nu}) = 0. \tag{3.8}$$

Let us now obtain the last equation of motion. From (3.1)

$$\frac{d\lambda}{d\tau} E^{(0)}{}_{\mu} = \frac{d}{d\tau}(mu_{\mu}) - eF_{\mu\nu} u^{\nu} - \frac{k}{2} S^{\alpha\beta} F_{\alpha\beta,\mu}, \tag{3.9}$$

where

$$m = m_0 + (k/2) S^{\alpha\beta} F_{\alpha\beta} \tag{3.10}$$

is the effective mass of the particle.

If we use (2.18), (2.22), (2.23), and the identity

$$\frac{S^{\mu\alpha} S^{\nu}_{\alpha}}{h^2} = (b^1 b^2 - b^2 b^1)^{\mu\nu}, \tag{3.11}$$

which is a consequence of (3.3) and the constraints (2.4), we obtain

$$\frac{d}{d\tau}(mu_{\mu}) = eF_{\mu\alpha} u^{\alpha} + \frac{k}{2} S^{\alpha\beta} F_{\alpha\beta,\mu} - \frac{S_{\mu\alpha} S^{\alpha\nu}}{h^2} \left\{ (e - mk) F_{\nu\rho} u^{\rho} + \frac{k}{2} S^{\beta\gamma} F_{\beta\gamma,\nu} \right\}. \tag{3.12}$$

From (3.7), (3.10), and (2.4) we see that

$$u^{\mu} \frac{du_{\mu}}{d\tau} = \frac{k}{2m} \frac{dS^{\alpha\beta}}{d\tau} F_{\alpha\beta} = 0, \tag{3.13}$$

which expresses the conversation of (2.3).

Similarly, we prove that

$$u_{\mu} \frac{dS^{\mu\nu}}{d\tau} + \frac{du_{\mu}}{d\tau} S^{\mu\nu} = 0 \tag{3.14}$$

and as a consequence, the conservation of the condition

$$S^{\mu\nu} u_{\nu} = 0 \tag{3.15}$$

which is in agreement with (2.4) and (3.3).

Equations (3.8), (3.13), and (3.14) are the consistence conditions of the theory.

IV. CONCLUSION

The present set of equations obtained by a geometrized constrained variational principle could also be obtained in a more traditional way by using Lagrangian multipliers.[9]

The geometrical treatment is, however, more intuitive and it can avoid the null momenta conjugate to the Lagrangian multipliers, which may be inconvenient for the quantization of constrained systems.

It is interesting to point out that the coupled equations (3.7) and (3.12) are more satisfactory than those similar equations found in the literature. For example, in the elucidative text of A. Barut[13] we have

$$\frac{dP_{\mu}}{d\tau} = eF_{\mu\nu} u^{\nu} + \frac{k}{2} S^{\lambda\nu} F_{\lambda\nu,\mu} \tag{4.1}$$

and

$$\frac{dS_{\mu\nu}}{d\tau} + 2P_{[\mu} u_{\nu]} = 2kF_{\rho[\mu} S_{\nu]}{}^{\rho} \tag{4.2}$$

in lieu of (3.12) and (3.7), when we consider (2.2). The mechanical moment P_{μ} satisfies the relation

$$(\delta^{\alpha}_{\mu} - u^{\alpha} u_{\mu}) P_{\alpha} = S_{\mu\nu} \frac{du^{\nu}}{d\tau} + 2kF_{\rho[\mu} S_{\nu]}{}^{\rho} u^{\nu}, \tag{4.3}$$

which follows from (3.2) and (4.2).

Equation (4.3) has two negative features: (i) it cannot be truly inverted because it is a projective relation, and (ii) it depends on the acceleration, which implies that Eq. (4.1) is of third order in proper time and so it admits free helicoidal solutions. These spurious solutions are characteristic of all theories that treat condition (3.15) as a holonomic one.[2–4]

It is possible to avoid the third-order character of the equations of motion if we assume the orthogonality between spin and mechanical moment,[6–8] which in general is not parallel to four-velocity. Thus it is difficult, however, to find a closed simple form for the mechanical moment and in these theories the canonical condition (2.3) is not preserved.[14]

None of these negative features are present in Eqs. (3.7) and (3.12), and the conditions (2.3) and (3.15) are conserved in a simple way.

We could also point out that if we define the spin four-vector as

$$S_{\rho} = \tfrac{1}{2} \epsilon_{\tau\mu\nu\rho} u^{\tau} S^{\mu\nu}, \tag{4.4}$$

where $\epsilon_{\tau\mu\nu\rho}$ is the usual Levi-Cevità tensor, we obtain from (3.7), (3.10), and (3.12) the precession equation

$$\frac{dS}{d\tau} = k\,\{F_{\rho\mu} S^{\mu} - (u_{\tau} F^{\tau\mu} S_{\mu}) u_{\rho}\}$$

$$- u_\rho \left\{ \frac{e}{m} F_{\mu\nu} u^\nu + \frac{k}{2m} S^{\nu\tau} F_{\nu\tau,\mu} \right\} S^\mu, \tag{4.5}$$

which was first established by Bargmann, Michel, and Teledgi[15] for the special case of the homogeneous electromagnetic field.

Our equations can be generalized so as to include curvature and torsion effects if considering space-time as a Riemann–Cartan variety, by using a minimal coupling procedure. These considerations will be presented in a later paper.[16]

[1] C. M. Amaral and P. Pitanga, Rev. Bras. Física **3**, 473 (1982).

[2] J. Frenkel, Z. Phys. **37**, 243 (1926).

[3] C. Lattes, M. Schönberg, and W. Schützer, An. Acad. Bras. Ciênc. **19**, 193 (1947).

[4] F. Halbwachs, Prog. Theor. Phys. **29**, 291 (1960).

[5] A. Shild, *Lectures in Applied Mathematics* (American Mathematical Society, Providence, R.I., 1967) Vol. 8.

[6] W. G. Dixon, Proc. R. Soc. London Ser. A **314**, 499 (1968).

[7] L. Suttorp and S. de Groot, Nuovo Cimento A **65**, 245 (1970).

[8] G. Cognola, R Soldati, L. Vanzo, and S. Zerbini, Phys. Rev. D **25**, 3109 (1982).

[9] R. Amorim, "The Motion of a Classical Particle with Spin," If/UFRJ/83/23, Instituto de Física, Universidade Federal do Rio de Janeiro, 1983. All the relevant results quoted here can also be found in Ref. 16.

[10] C. Lanczos, *The Variational Principle of Mechanics* (Univ. of Toronto, Toronto, 1974).

[11] J. L. Leech, *Classical Mechanics* (Methuen, New York, 1965), 2nd ed.

[12] $\eta_{\mu\nu} = \eta^{\mu\nu} = \text{diag.}(+1, -1, -1, -1)$.

[13] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).

[14] See, for example, Ref. 8.

[15] V. Bargmann, L. Michel, and V. Teledgi, Phys. Rev. Lett. **2**, 435 (1959).

[16] R. Amorim, "The Motion of a Classical Spinning Particle in a Riemann–Cartan Space-Time," Universidade Federal Do Rio de Janeiro, Instituto De Fisica 83126 (1983).

Integrability and the Painlevé property for low-dimensional systems

A. Ramani
Centre de Physique Théorique, Ecole Polytechnique, F-91128 Palaiseau, France

B. Dorizzi and B. Grammaticos
Département de Mathématiques (M.T.I.), Centre National d'Etudes de Télécommunications, F-92131 Issy les Moulineaux, France

T. Bountis
Department of Mathematics and Computer Science, Clarkson College, Potsdam, New York 13676

(Received 16 June 1983; accepted for publication 7 October 1983)

We examine a system described by two first-order nonlinear differential equations from the point of view of integrability. The singularity analysis in the complex-time plane is used to investigate the Painlevé property, which according to the Ablowitz–Ramani–Segur conjecture is a prerequisite for integrability for infinite-dimensional systems. We show that for such low-dimensional systems, the Painlevé analysis is still a most useful guide, but integrable cases also exist which do not possess the Painlevé property.

PACS numbers: 03.20. + i, 02.30. + g

I. INTRODUCTION

In the last few years, a new method has been developed for investigating the integrability of dynamical systems. The principle of the method is old. It is based on the relation between integrability of nonlinear differential equations and the analytic behavior of their solutions. The names of Fuchs,[1] Kowalevskaya,[2] Painlevé,[3] and Gambier [4] have been associated with the early developments of this approach. This method has been resurrected recently by Ablowitz, Ramani, and Segur (ARS)[5] in relation to their conjecture on the integrability of nonlinear partial differential equations (PDE's). They have conjectured that a PDE is integrable whenever it possesses the Painlevé property, i.e., the only movable singularities are poles. This conjecture was implemented through a powerful algorithm for the study of the singularity structure of all the reductions of the PDE's to ordinary differential equations (ODE's).

The fact that the initial studies of the past century mathematicians concerned ODE's, combined with the existence of numerous interesting physical systems described by ODE's led to a natural "tacit" extension of the ARS conjecture[5] to systems of ODE's. The works of Bountis, Segur, and Vivaldi[6] as well as of Tabor *et al.*[7] and Segur[8] have demonstrated the usefulness of the conjecture for a dynamical system. They have led to the discovery of new cases of integrability for dissipative as well as for Hamiltonian systems.[9] One important question which can be formulated in relation to the conjecture concerning ODE's is whether the Painlevé property, i.e., absence of movable singularities in the complex time plane "worse" than poles, is a prerequisite for integrability. Now, there certainly exist examples of integrable systems which do not possess the Painlevé property.[10,11] Such examples are mostly encountered in low-dimensional systems. This remark has motivated the present study.

In this work, we explore the connection between the Painlevé property and integrability in systems of two first-order ODE's of the form

$$\dot{x} = f(x,y), \quad \dot{y} = g(x,y),$$

where f,g are polynomials of second degree in x, y (see Sec. II). Such systems are known to arise in several conservative or dissipative physical models.[12] Complete integrability in this case is associated to the existence of two time-dependent (analytical, single valued) integrals of motion. Alternatively, a single time-independent integral suffices to determine the trajectories in the x-y plane, and then the time dependence of the motion can be obtained through quadratures. Systems of this kind are encountered in the modelization of dissipative physical phenomena. However, we will also call completely integrable, systems which reduce to linear, nonautonomous differential equations.

For the system at hand we will show that the singularity analysis leads to cases where the Painlevé property is associated to integrability. However, we have been able to show that there exist whole classes of cases which do not possess the Painlevé property, and yet do possess one time-independent integral. Sometimes, they even reduce to $\dot{x} = h(x)$, say, in which case they are, of course, immediately integrated by quadratures.

In Sec. III, the singularity analysis is presented, and the integrals of motion are obtained in Sec. IV for cases with and without the Painlevé property. The paper ends with a more extensive discussion in Sec. V, on the relation between the Painlevé property and the question of integrability of dynamical systems, in general.

II. A SIMPLE NONLINEAR DIFFERENTIAL SYSTEM

Let us consider the following system:

$$\dot{x} = -x^2 + axy + \alpha x + \beta y + \lambda,$$
$$\dot{y} = -y^2 + bxy + \gamma x + \delta y + \mu, \tag{1}$$

where the dot denotes differentiation with respect to the time t.

If we limit ourselves to the higher-order part

$$\dot{x} = -x^2 + axy = -x(x - ay),$$
$$\dot{y} = -y^2 + bxy = -y(y - bx), \tag{2}$$

 0022-2488/84/040878-06$02.50

we see that each variable can be factorized in its respective time derivative. It is interesting to introduce here the unique (up to a multiplicative constant) extra linear combination of x and y that has the same property, namely

$$z = (1+b)x - (a+1)y .$$

Indeed

$$\dot z = -z(x+y) .$$

Note that if either $a = -1$ or $b = -1$, z is not distinct of x or y.

It should be made clear that our aim is to study the relation between integrability and the singularity structure and illustrate it through the above system, rather than study this system for its own sake. Therefore, we will not attempt to study it exhaustively. In particular, we will limit ourselves to the generic case where x, y and z are all distinct, i.e.,

$$(a+1)(b+1) \neq 0 .$$

Moreover, *generically*, one can define new sets of variables (x',z') and (y'',z'') such that the reduced system (2) can be written

$$\begin{aligned} \dot x' &= -x'^2 + ax'z' \\ \dot z' &= -z'^2 + cx'z' \end{aligned} \tag{3}$$

or, alternatively

$$\begin{aligned} \dot y'' &= -y''^2 + by''z'' \\ \dot z'' &= -z''^2 + cy''z'' , \end{aligned} \tag{4}$$

where

$$\frac{1}{a+1} + \frac{1}{b+1} + \frac{1}{c+1} = 1 , \tag{5}$$

with

$$z' = -\frac{z}{a+1}, \quad x' = -\frac{(b+1)x}{c+1}$$

and

$$z'' = +\frac{z}{b+1}, \quad y'' = -\frac{(a+1)y}{c+1} .$$

Since we start with $(a+1)(b+1) \neq 0$, we will never get $c+1=0$ from (5). We will also discard the case where (5) does not define a finite c, namely $ab=1$. Note that we will never get either $bc=1$ or $ac=1$, as that would correspond to either a or b infinite. With the above restrictions, the systems (2), (3), and (4) are completely equivalent. In an analogous way, the full system (1) possesses two other equivalent expressions in terms of (x',z') and (y'',z'').

System (1) can be further simplified through translations. Indeed, if $a \neq 0$ and $b \neq 0$ the terms βy and γx can be eliminated. The case $ab = 0$ will be studied separately.

Other convenient translations (which exist even if $ab = 0$) are those for which λ and μ are made to vanish. There are in general four such translations. In these frames, the origin is a fixed point. The character of this fixed point can be very different, depending on the values of $\alpha,\beta,\gamma,\delta$. If $\alpha\delta - \beta\gamma < 0$, the origin is a saddle point.[12] If $\alpha\delta - \beta\gamma > 0$, it is unstable when $\alpha + \beta > 0$ ["unstable spiral" if moreover $(\alpha-\delta)^2 + 4\beta\gamma < 0$], while it is stable if $\alpha + \delta < 0$ ["stable spiral" if $(\alpha-\delta)^2 + 4\beta\gamma < 0$]. The most interesting case is the one where $\alpha + \delta = 0$ and $\alpha\delta - \beta\gamma > 0$ [or equivalently $(\alpha-\delta)^2 + 4\beta\gamma < 0$]. Then the two linear eigenvalues are pure imaginary. This would suggest that the origin is a center, i.e., that all trajectories in a finite neighborhood of this fixed point are closed orbits. This, however is not necessarily true due to the nonlinear terms. In fact, in general, the origin will be a spiral point either stable or unstable and only exceptionally a center.

We have numerically observed all these behaviors in one particular case, namely $a = 1$, $b = 2$, and $\beta = -\alpha$. Along the $\alpha + \delta = 0$ line suggested by the linear stability analysis, we found a center only for $\gamma = 2\alpha$ which translates to a frame where β and γ vanish giving $\delta = (-3/2)\alpha$, $\lambda = \mu = 0$. Quite remarkably, this very same set of values will be encountered again in the Painlevé analysis of the singularity behavior.

Moving away from the $\alpha + \delta = 0$ line, we have also observed limit cycles enclosing the origin, where the linear instability is balanced by a nonlinear stability.

Thus, this quite simple nonlinear system can exhibit the full variety of orbit behavior of two-dimensional systems.

III. PAINLEVÉ ANALYSIS

System (1) can be written as a simple second-order differential equation in terms of either x or y (or z!), and one could wonder whether any of these equations can be reduced to one of the fifty canonical forms enumerated by Painlevé and Gambier.[3,4] There does not exist, however, a systematic way to ascertain that a given equation does or does not reduce to one of these canonical forms. On the other hand, there is a straightforward, if somewhat tedious, algorithm to check directly for the Painlevé property.[5]

The first step of this Painlevé analysis is the research of the leading behaviors. For this system, there are three possible leading behaviors near the (arbitrary) singularity at $t = t_0$:

$$x \text{ and } y \sim \frac{1}{(t-t_0)}, \quad z \sim (t-t_0)^c ,$$

$$x \text{ and } z \sim \frac{1}{(t-t_0)}, \quad y \sim (t-t_0)^b ,$$

$$y \text{ and } z \sim \frac{1}{(t-t_0)}, \quad x \sim (t-t_0)^a .$$

Any of these behaviors is absent if the corresponding exponent a,b,c, is less than -1. Because of (5), however, at least one of these numbers must be larger than -1. Therefore, there exists at least one possible leading behavior. For the Painlevé property to be satisfied, one immediately sees that a necessary condition is that those of the exponents a,b,c that are larger than -1, be integers. (Incidentally, if a and b are equal to -1, z and c are not defined, but the Painlevé property could still hold, depending on the linear and constant terms in (1). As we explained in the previous subsection, this degenerate case will not be dealt with.) It is easy to find all solutions of (5) where the exponents a,b,c are either less than -1 or nonnegative integers. Up to permutations, these solutions are:

(i) $a=0 \quad b=n \quad c=-2-n$, n nonnegative integer,
(ii) $a=1 \quad b=2 \quad c=5$,
(iii) $a=1 \quad b=3 \quad c=3$,
(iv) $a=2 \quad b=2 \quad c=2$.

Note that $a=b=1$ allows us to define z, but not c, because the equation for x in terms of x and z is

$$\dot{x}=-xz/2$$

and cannot be rewritten in the form (3) by an appropriate scaling. We will again discard this degenerate case.

Consider first the case $a=0$. If β is zero in (1), the equation for x separates and is of Riccati type, and therefore has the Painlevé property.[3,13] Once the equation for x is integrated, and a solution for x is chosen, the equation for y

$$\dot{y}=-y^2+(bx+\delta)y+(\gamma x+\mu) \tag{6}$$

is again of Riccati type, and the movable singularities for y (at given x!) are poles. However, whenever x has a pole, the behavior of y must be studied in more detail. Indeed the Riccati equation (6) could *a priori* exhibit a "fixed" singularity worse than a pole. However, this "fixed" singularity for y is really a movable singularity of the original system (1), because the pole of x is movable. If it were worse than a pole, the original system (1) would not possess the Painlevé property. A first necessary condition for this property to hold is just that b be an integer other than -1, which is precisely condition (i), up to the permutation of b and c. A second necessary condition, which will then be sufficient, can be obtained as the vanishing of some polynomial in α, γ, δ, λ, and μ. This polynomial is homogeneous of degree $b+1$, provided each λ and μ is counted as a factor of degree two. (Here b is the nonnegative integer, while $c=-b-2$ is negative.)

For instance, this necessary condition for $b=0$ is obviously

$$\gamma=0.$$

For b strictly positive, γ can be absorbed by a translation of y, and in this frame the condition is

$$\begin{aligned}&\text{for } b=1 \quad \mu=0,\\ &\text{for } b=2 \quad \mu(\alpha+\delta)=0,\\ &\text{for } b=3 \quad \mu(2\alpha^2+3\alpha\delta+\delta^2-\lambda+\mu/2)=0.\end{aligned} \tag{7}$$

For b strictly positive, the vanishing of μ in this frame is always sufficient, but clearly not necessary (for $b \geqslant 2$).

If β is not zero, then the system (1) does not have the Painlevé property. In fact, whenever y diverges faster than x, then the leading behavior of y is $1/(t-t_0)$ and x behaves as $\ln(t-t_0)$. Moreover, because of the $-x^2$ term in the time derivative of x, even $\exp(x)$ is not free from logarithms.

In the three remaining cases, it is immediate to show that the reduced system, where all linear and constant terms are taken equal to zero is indeed Painlevé. We are now interested in the different choices of α, β, γ, λ, μ that preserve this property. Since neither a nor b vanish, one can translate x and y so as to absorb β and γ. System (1) thus becomes

$$\begin{aligned}\dot{x}&=-x^2+axy+\alpha x+\lambda,\\ \dot{y}&=-y^2+bxy+\delta y+\mu\end{aligned} \tag{8}$$

with properly redefined α, δ, λ and μ.

In case (ii) it is easy to check that

$$\lambda=0$$

is a first necessary condition for the Painlevé property. Finally, there are eight distinct cases unique up to a scaling, that preserve the Painlevé property, namely

$$\delta=-3\alpha/2 \quad \mu=0, \tag{9}$$
$$\delta=\alpha=0 \quad \mu \text{ free}, \tag{10}$$
$$\delta=\alpha \quad \mu=0, \tag{11}$$
$$\delta=-4\alpha \quad \mu=0, \tag{12}$$
$$\delta=-\alpha \quad \mu=6\alpha^2/25, \tag{13}$$
$$\delta=-\alpha \quad \mu=-6\alpha^2/25, \tag{14}$$
$$\delta=-8\alpha/7 \quad \mu=0, \tag{15}$$
$$\delta=-13\alpha/7 \quad \mu=0. \tag{16}$$

Once these cases are identified, one can look for the reduction to a canonical form[3,4,13] of one of the second-order ordinary differential equations equivalent to this system. It is easier to do so for the equation in terms of x. We find that all these cases are reducible to equation (X).[13]

$$\frac{d^2W}{dZ^2}=-W\frac{dW}{dZ}+W^3-12q(Z)W+12q'(Z).$$

For cases (9) and (10) q is zero while it is a nonzero constant proportional to α^2 in case (11)–(13). Finally for cases (14)–(16), $q(Z)$ is a Weierstrass elliptic function. The relation between (W,Z) on the one hand, and (x,t) on the other is complicated and need not be explicited here.

In case (iii), one can also easily check that

$$\lambda=0$$

is again a necessary condition. We finally obtain three discrete cases (up to a scaling)

$$\delta=-5\alpha \quad \mu=0, \tag{17}$$
$$\delta=-\alpha \quad \mu=4\alpha^2/9, \tag{18}$$
$$\delta=\alpha \quad \mu=0, \tag{19}$$

and, in addition, a continuous family governed by

$$\delta=-2\alpha \quad \mu \text{ free}. \tag{20}$$

Since $b=c$, there exists a symmetry with respect to the exchange of y and z as far as the dominant terms are concerned, i.e., (2) and (3) have the same form. When the system is complete with linear and constant terms, this symmetry exchanges the cases (17) and (18) while cases (19) and, of course, (20) remain invariant.

Again, one can look for reductions of the second-order differential equation for x. Cases (17–19) reduce to the canonical form (VII)[13]

$$\frac{d^2W}{dZ^2}=2W^3,$$

while case (20), in general, reduces to the type (VIII)

$$\frac{d^2W}{dZ^2}=2W^3+\beta W+\gamma.$$

Case (iv) has the Painlevé property in one discrete case

$$\delta = \alpha \quad \lambda = \mu = 0\,, \tag{21}$$

and one doubly continuous family

$$\delta = -\alpha \quad \lambda \text{ free}, \mu \text{ free}. \tag{22}$$

Both cases (21) and (22) can be reduced to the canonical form (XXX)

$$\frac{d^2W}{dZ^2} = \frac{1}{2W}\left(\frac{dW}{dZ}\right)^2 + \frac{3W^3}{2} + 4\alpha W^2 + 2\beta W - \frac{\gamma^2}{2W}.$$

Here, $a = b = c$, and (2), (3), and (4) all have the same form. Clearly cases (21) and (22) are each invariant with respect to permutations of x, y, and z.

To conclude, we can remark that all the Painlevé subcases of (ii), (iii), and (iv) can be integrated in terms of elliptic functions, but the explicit form of the solutions are usually very complicated.

IV. INTEGRATION OF THE EQUATIONS OF MOTION

While the explicit expression of the solutions of (ii), (iii), and (iv) are quite involved, one can obtain, rather easily, integrals of the motion. Since several changes of variable must be made "en route," these integrals are not always time independent in the original coordinates. In fact this is true for cases (11)–(19) and (21) while in cases (9), (10), (20), and (22) we find genuine constants of the motion. The latter are

case (9) $x^3y^2(z-3\alpha)$

$$\equiv[\dot x + x^2 - \alpha x]^2[-2\dot x + x^2 - \alpha x] = \text{const}\,, \tag{23}$$

case (10) $(xy+\mu)^2(xz+\mu)$

$$\equiv[\dot x + x^2 + \mu]^2[-2\dot x + x^2 + \mu] = \text{const}\,, \tag{24}$$

case (20) $\dot x^2 - x^4 + 2\alpha x^3 - (\alpha^2+\mu)x^2 = \text{const}\,, \quad (25)$

case (22) $[\dot x^2 - x^4 + 2\alpha x^3 - (\alpha^2 + 4\mu - 2\lambda)x^2 - \lambda^2]/x = \text{const}\,. \quad (26)$

We can immediately note that all these constants (and also the time-dependent integrals of the other cases) are "perturbations" of a fundamental constant of the motion for system (2), namely

$$x^{1/(a+1)}y^{1/(b+1)}z^{1/(c+1)} = \text{const}\,,$$

which can be rewritten as

$$x^{(1-a)/(1+a)}(\dot x + x^2)^{1/(b+1)}\left(\dot x - \frac{b+1}{c+1}x^2\right)^{1/(c+1)} = \text{const}. \tag{27}$$

This quantity is a genuine constant for any value of a and b [with c given by (5)], whether integer or not. Therefore, the homogeneous system (2) is always integrable, whether it has the Painlevé property or not. In some cases [other than (i)–(iv)] one can even solve Eq. (27) for $\dot x = h(x)$, whence the corresponding system would be explicitly integrable by quadratures. Although this contradicts the naive form of the Painlevé conjecture, this is not too surprising due to the simplicity of the homogeneous system (2).

At this point the question arises, as to whether the only ways to perturb system (2) through linear and constant terms, while preserving integrability, lead to Painlevé cases.

This turns out not to be true. There exist at least three families of integrable perturbations of (2), generically not Painlevé.

These families are, in the frame where $\beta = \gamma = 0$:

(i) $\delta = -\alpha(b+1)/(a+1)$, $\lambda = \mu = 0$; a,b,α free, (28)

$$x^{1/(a+1)}y^{1/(b+1)}[z-(b+1)\alpha]^{1/(c+1)} = \text{const},$$

or, equivalently

$$x^{(1-a)/(1+a)}(\dot x + x^2 - \alpha x)^{1/(b+1)} \times[-(a+1)\dot x + (ab-1)(x^2-\alpha x)]^{1/(c+1)} = \text{const}\,, \tag{29}$$

(ii) $a(b-1) = 2$ (i.e., $b = c$), $\lambda = 0$, $a\delta + 2\alpha = 0$; a, α, μ free, (30)

$$x^{(1-a)/2a}(\dot x - x^4 + 2\alpha x^3 - (\alpha^2 + \mu a^2)x^2) = \text{const}\,, \tag{31}$$

(iii) $a = 1$, $\alpha = \delta = \lambda = 0$; b, μ free, (32)

$$[xy + \mu/(b-1)]^{1/(b+1)}(xz+\mu)^{1/(c+1)} = \text{const}, \tag{33}$$

or, equivalently

$$[\dot x + x^2 + \mu/(b-1)]^2[-2\dot x + (b-1)x^2 + \mu]^{b-1} = \text{const}\,. \tag{34}$$

Case (9) is a subcase of (28), case (10) is a subcase of (32) and case (20) is a subcase of (30). There does not seem to exist any integrable family, generically non-Painlevé, that encompasses case (22). Incidentally, the Painlevé cases associated with integrals of motion depending on time in the original coordinates, can sometimes be embedded in a continuous family, generically non-Painlevé, for which there exists one such integral. Since this integral is unique and time dependent, this does not suffice for integrability in general. To our knowledge, the Painlevé subcases are the only integrable ones.

One could wonder at this point whether system (1) is not always integrable. Numerical studies cannot easily settle the question because, in two dimensions, chaos cannot be present. Although a limit cycle is usually an indication of nonintegrability, its existence does not preclude integrability, as it could be associated to a time-dependent integral. The best indication of nonintegrability is perhaps the presence of a nonlinear spiral point where the linear analysis suggested a center. And indeed, as we noted in Sec. II, for the one-dimensional family (up to a scaling)

$$a = 1, b = 2, \lambda = \mu = 0, \beta = -\alpha, \alpha + \delta = 0, \gamma \text{ free}\,,$$

we obtained a center only for $\gamma = 2\alpha$, which was the only case for which we found an integral [it translates to case (9)]. We can thus surmise that system (1) is not generically integrable. Still, the class of integrable cases is larger than the class identified by the Painlevé property.

Finally, consider case (i), with $\beta = 0$. The first equation, in terms of x only, is easy to solve. Substituting the solution $x(t)$ into Eq. (6), we obtain a Riccati equation for y. This equation can be reduced to a linear second-order equation, for all values of the other parameters, whether it possesses the Painlevé property or not. Conversely, the cases that do have this property are not any "more integrable" than the others, i.e., their solutions will generally not be expressible in terms of quadratures.

V. RELATION BETWEEN PAINLEVÉ PROPERTY AND INTEGRABILITY

As we have seen in the previous section, the Painlevé analysis has revealed several integrable cases. For some of them (14)–(16), the integrals have so intricate an expression that they could not have been obtained by pure guess work as may be done with (29), (31), and (33). Thus, the Painlevé criterion is a most useful integrability detector. However, as was explained in the previous section, there exist whole classes of integrable cases which, generically, do not possess the Painlevé property. This is not astonishing. One degree of freedom Hamiltonian systems are integrable by a simple quadrature. First-order autonomous differential equations

$$\dot{x} = f(x)$$

are also trivially integrable whatever their singularities. When the equation in question is nonautonomous, integrability is not automatic. However, one can always find integrable systems which do not satisfy the Painlevé criterion (see Appendix). So it is clear that the Painlevé approach would be an over-restrictive condition for integrability for such systems.

The case of two degree of freedom Hamiltonian systems has been dealt with in detail, in previous publications.[10,11] Our conclusion was that whenever such a Hamiltonian system is integrable without satisfying the Painlevé criterion, then either the undesirable singularities are obvious in the structure of the Hamiltonian itself, or, when this is not the case, the weak Painlevé property[9] is satisfied. In this paper, we have treated two-dimensional systems described by two first-order autonomous ODE's. Our analysis resulted in the discovery of integrable systems, quite similar in structure, some of which were Painlevé and some not. Conversely, for very similar non-Painlevé systems, some were integrable and some not. So the Painlevé approach, though undeniably fruitful, does not offer for these simple systems a clear-cut distinction between integrable and nonintegrable systems.

On the other hand, the concept of integrability needs to be more clearly defined. Consider, for instance second-order nonautonomous ODE's. Painlevé[3] and Gambier[4] have analyzed these equations from the point of view of the singularity structure. Their analysis resulted in fifty equations having no movable singularities but poles. Most, but not all, of them are integrable in the classical sense: the solution can be given through quadratures. Some (type V, Ref. 13, or Riccati) can be reduced to linear nonautonomous equations. The solution of the latter generally cannot be written in terms of quadratures. Finally, six are integrable only inasmuch as their solutions are the new transcendents introduced by Painlevé. At this point, the definition of integrability may appear somewhat circular. However, the introduction of the Painlevé transcendents as new "special functions" is precisely justified by the regularity of their structure in the whole complex plane. This point was brought to our attention by M. D. Kruskal. *A posteriori*, the particular usefulness of these transcendents, due to their frequent appearance in equations of nonlinear physics, illustrates the power of Painlevé's intuition.

For higher-order non-Hamiltonian systems, very few results exist. Our analysis of several examples of third-order, coupled, nonlinear autonomous ODE's[14] has identified several new integrable systems possessing the Painlevé property. We have also found several cases, which are integrable, without satisfying the Painlevé conditions.[14] However, these cases turn out to have a simple structure and their integrability can be assessed at a glance. When one goes to still higher orders the relation between integrability and the Painlevé property appears to become more intimate.[14,15] However, one must bear in mind that it may turn out that new transcendents are needed for the integrability of higher-order systems, just like the six ones introduced by Painlevé at order two.[16] So far, no confirmed example of such a new transcendent is known. However, some good candidates can be presented. Consider the modified Korteweg-de Vries (mKdV) hierarchy related to the Korteweg-de Vries hierarchy by the Miura transformation.[17] The nth equation of the mKdV hierarchy is a PDE of order $2n+1$, which reduces to a nonautonomous ODE of the same order through the ansatz

$$u(x,t) = t^{-1/(2n+1)}\, U(\xi) \quad \text{with } \xi = xt^{-1/(2n+1)}.$$

This equation can be integrated once, leading to a nonautonomous ODE of order $2n$. It has been shown[18] that the singularities with algebraic leading order of the three first members of this hierarchy are indeed pure poles. Although this is not a full proof of the Painlevé property, the ARS conjecture[5] asserts that all of those ODE's have only movable poles. The lowest member, in fact, is nothing else but the second Painlevé transcendental equation. It is tempting to surmise that these equations could lead to new transcendents.

In Hamiltonian systems of N degrees of freedom ($2N$-dimensional phase space) the concept of complete integrability is defined by the existence of N independent, analytic, single valued, global integrals in involution.[19] As mentioned earlier, several such integrable systems have already been identified by the Painlevé property.[6,10] Now, it does happen sometimes that upon integrating such a system by quadratures, one arrives at one degree of freedom motion, which is plagued by the "uglier" singularities imaginable.[20] It has been our experience, however, that the "uglier" the singularities, the easier it is to see what variable transformations are needed to uncouple the degrees of freedom and integrate the equations completely.[11]

Concerning infinite-dimensional systems described by PDE's, it appears that the Painlevé property is indeed a prerequisite for integrability. For some very simple examples, integrability results from a reduction to a linear PDE by a local change in the dependent variable. For instance, the Burger's equation

$$U_t + U_{xx} + 2UU_x = 0$$

becomes

$$V_t + V_{xx} = 0 \quad \text{with } U = \frac{V_x}{V},$$

or the Liouville equation

$$U_{xt} = \exp(U)$$

reduces to $\varphi_{xt} = 0$ with $U = \ln(2\varphi_x \varphi_t / \varphi^2)$. However, for the vast majority of the so-called "integrable" nonlinear PDE's, integrability is associated with the Inverse Scattering Transform (IST).[21] In this case the solutions are obtained through linear integral equations, the kernel of which obeys a linear partial differential equation. So integrability for PDE's is related to the existence of some linear problem which is sometimes simple and sometimes of considerable complexity.[18]

Incidentally, these PDE's are related by similarity transformations to ODE's of the Painlevé type.[5] Thus the integrable character of Painlevé equations as well as of the equations of the hierarchy mentioned above is once more exemplified by the fact that they can now be linearized in terms of linear integral equations.[22]

VI. CONCLUSION

We have examined in this paper, on a second-order autonomous differential system, the relation between integrability and (movable) singularities of the solutions in the complex t-plane. We have found, for this system, that integrability can exist independent of the Painlevé property. However, we have shown that the singularity analysis offers a very useful and direct method for identifying integrable systems, which would otherwise be very difficult to discover. This appears to be all the more true for systems of higher dimensionality, where there is evidence that there are deeper connections between integrability and the Painlevé property. This is in perfect agreement with the ARS conjecture[5] for infinite-dimensional systems (i.e., PDE's).

ACKNOWLEDGMENTS

The authors want to thank M. D. Kruskal for numerous stimulating discussions.

APPENDIX

Consider the obviously integrable equation

$$\frac{dx}{dt} = x^3 + ax + b, \tag{A1}$$

where a and b are constants. This equation never has the Painlevé property in terms of x. In terms of x^2, it does not have this property either, if $b \neq 0$. If we make the following transformation

$$x = y\psi(T) + z(T),$$
$$t = \varphi(T) \quad \text{with} \quad \frac{d\varphi}{dT} = \frac{1}{\psi^2}, \tag{A2}$$

Eq. (A1) becomes

$$\frac{dy}{dT} = y^3 + A(T)y^2 + B(T)y + C(T). \tag{A3}$$

A, B, and C are given by

$$\begin{aligned} A &= 3u, \\ B &= 3u^2 + av^2 + v'/v, \\ C &= u^3 + auv^2 + bv^3 + v'u/v - u', \end{aligned} \tag{A4}$$

where $u = z/\psi$, $v = 1/\psi$.

For general A, B, and C, Eq. (A3) is not integrable. Whether A, B, C are given by (A4) or not, Eq. (A3) never has the Painlevé property. Still, it is integrable when A, B, and C are given by (A4). Of course, on the view of (A4) as it stands, one could guess that u and v can be used to change back to the "right" variables. However, when definite expressions for A, B, C, are given, for example

$$\begin{aligned} A &= T + 3, \\ B &= 2T + 3 + 1/T, \\ C &= T + 1 + 1/T, \\ &(u = 1 + T/3,\ v = T,\ a = -1/3,\ b = 2/27), \end{aligned}$$

it is not obvious anymore that a change of variables of the form (A2) is adequate. Thus, this integrable equation cannot be distinguished at a glance from similar nonintegrable equations, and would not be singled out by a Painlevé analysis either. For such systems, integrability is not related to the analytic properties of the solutions.

[1] L. Fuchs, Sitzungsber. Akad. Wiss. Wien Berlin **32**, 699 (1884).
[2] S. Kowalevskaya, Acta Math. **14**, 81 (1890).
[3] P. Painlevé, Acta Math. **25**, 1 (1902).
[4] B. Gambier, Acta Math. **33**, (1910).
[5] M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **23**, 333 (1978); J. Math. Phys. **21**, 715 and 1006 (1980).
[6] T. Bountis, H. Segur, and F. Vivaldi, Phys. Rev. A **25**, 1257 (1982).
[7] M. Tabor and J. Weiss, Phys. Rev. A **24**, 2157 (1981).
[8] H. Segur, "Solitons and the Inverse Scattering Transform," International School of Physics, "Enrico Fermi," Varenna, Italy 1980.
[9] L. Hall, Physica D. **8**, 90 (1983).
[10] A. Ramani, B. Dorizzi, and B. Grammaticos, Phys. Rev. Lett. **49**, 1539 (1982); B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **24**, 2282 (1983).
[11] B. Dorizzi, B. Grammaticos, and A. Ramani, J. Math. Phys. **25**, 481 (1984).
[12] C. Bender and A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978).
[13] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956). Note: Equation VI p. 334 should read $d^2w/dz^2 = zw(dw/dz) - w^3 + q(z)$. Ince quotes Gambier[4] who himself incorrectly quotes Painlevé[3] where the proper equation can be found.
[14] T. Bountis, A. Ramani, B. Grammaticos, and B. Dorizzi (in preparation).
[15] T. Bountis, M. Bier, and J. Mijmans, Phys. Lett. **97**, 11 (1983).
[16] M. D. Kruskal (private communications).
[17] F. Calogero and A. Degasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982).
[18] A. Ramani, in *Fourth International Conference on Collective Phenomena*, edited by J. L. Lebowitz (New York Academy of Sciences, New York, 1981) p. 54.
[19] V. I. Arnold, *Mathematical Method of Classical Mechanics* (Springer, New York, 1978).
[20] T. Bountis, "Chaos in Hamiltonian Systems and Singularity Analysis," in *Astrodynamics Conference Proceedings* (AIAA, New York, 1982).
[21] M. J. Ablowitz and H. Segur, *Solitons and Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[22] M. J. Ablowitz and H. Segur, Phys. Rev. Lett. **38**, 1103 (1977).

Higher-order symmetries of the compressible one-dimensional isentropic fluid equations

John Verosky
Department of Mathematics, Tulane University, New Orleans, Louisiana 70118

(Received 25 January 1983; accepted for publication 13 May 1983)

The equations of compressible one-dimensional isentropic flow are shown to have infinitely many higher-order symmetries by means of recursion operators. Furthermore, a new first-order conserved functional is found to be associated with a third-order symmetry via Noether's theorem.

PACS numbers: 03.40.Gc, 02.30. + g

I. INTRODUCTION

A recent theme in the study of nonlinear pde's has been to find evolution equations possessing infinitely many higher-order symmetries and recursion operators for these symmetries.[1–4] The KdV equation is the prototype of this situation.[5] An important class of nonlinear equations is that of fluid dynamics of which the simplest one-dimensional example is isentropic compressible flow. In this paper, we establish the existence of higher-order symmetries of this system, that is, evolution systems of order greater than one whose flows commute with the flow of the fluid system. Also, a new conserved functional, associated with one of the third-order symmetries via Noether's theorem, is determined.

Starting with the Riemann invariants as dependent variables, in Sec. II we find three recursion operators by using the hodograph transformation. Each of these operators, when applied to a known symmetry generator, will produce a generator of higher order. These operators immediately settle the question of the existence of infinitely many higher-order symmetries which, for systems, does not automatically follow from the existence of one higher-order symmetry as seems to be the case when there is only one dependent variable.[6] In Sec. III, we exhibit the huge variety of new higher-order symmetries and begin classifying them. In Sec. IV, the fluid equations themselves are discussed, with particular reference to the consequences of their Hamiltonian structure. Because of the correspondence between conserved functionals and Hamiltonian symmetries (asserted by Noether's theorem), it is natural to ask which of the higher-order symmetries is associated with a higher-order conserved functional. Owing to computational difficulties, so far only the third-order symmetry mentioned earlier has been shown to be Hamiltonian. However, the existence of infinitely many higher-order conserved functionals seems likely.

An infinite number of conserved densities could also be generated by a simple density by applying the infinite number of symmetries to it. It is still computationally difficult to check which of these symmetries are nontrivial.

II. DERIVATION OF THE RECURSION OPERATORS

We begin with some basic definitions. Let $u = (u^1, u^2, \ldots, u^k)^T$ be a k-tuple of dependent variables depending on an l-tuple of independent variables $x = (x^1, x^2, \ldots, x^l)^T$. The T denotes transpose, so that these vectors are column vectors. A system of pde's for u is $K = (K^1, K^2, \ldots, K^k)^T = 0$, where the K^i are pdf's (partial differential functions–smooth functions of x, u, and derivatives of the u^i with respect to the x^j). A *symmetry* of the system $K = 0$ is a (possibly local) one-parameter family of solutions of $K = 0$ with parameter s satisfying the evolution system $\partial u/\partial s \equiv u_s = P = (P^1, P^2, \ldots, p^k)^T$, where the p^i are also pdf's. P is called a symmetry generator of $K = 0$. These symmetry generators form a vector space. The condition that P generate a symmetry of $K = 0$ is $D_s K \doteq 0$. This means that if we take the total derivative of each K^i with respect to s, each resulting pdf vanishes when we make substitutions using the equations $K = 0$. More explicitly,

$$D_s = \sum D^J P^j \frac{\partial}{\partial u_J^j},$$

where $J = (j_1, j_2, \ldots, j_k)$ is a multi-index, $u_J^j = \partial^{j_1 + j_2 + \cdots + j_k} u^j / \partial x_1^{j_1} \partial x_2^{j_2} \cdots \partial x_k^{j_k}$, $D^J = D_{x_1}^{j_1} D_{x_2}^{j_2} \cdots D_{x_k}^{j_k}$ is a mixed-total derivative, and the sum is for $0 \leqslant j \leqslant k$, and $j_1, j_2, \ldots, j_k \geqslant 0$. The above operator acts on pdf's and is denoted $V(P)$. Hence the condition that P be a symmetry generator for $K = 0$ may also be written $V(P)K \doteq 0$. Often, one of the independent variables is singled out as time. An *evolution system* is of the form $\partial u/\partial t \equiv u_t = K$, where each K^i is independent of t and of t-derivatives of u. The condition that P generate a symmetry for $u_t = K$ is $D_t P - D_s K \doteq 0$, where D_t and D_s are total derivatives and $\doteq$ indicates that substitutions using $u_t = K$ are made. This may be rewritten as $\partial P/\partial t + V(K)P = V(P)K$, where $\partial P/\partial t$ means the derivative of P with respect to explicit t-dependence and $=$ is used instead of $\doteq$ since all substitutions for u_t are made automatically when we write $V(K)$. If we set $V(P)K - V(K)P = \langle P, K \rangle$, the condition is $\partial P/\partial t = \langle P, K \rangle$. The bracket $\langle\ ,\ \rangle$ is a sort of commutator for generators of evolution equations, and, in fact, the vector space of symmetry generators of a given system of pde's is a Lie algebra with respect to this bracket.[7–9]

First, we find symmetries of the evolution system

$$\begin{pmatrix} u_t \\ v_t \end{pmatrix} = \begin{pmatrix} (u + \epsilon v) u_x \\ (v + \epsilon u) v_x \end{pmatrix}. \tag{1}$$

Here, ϵ is a positive parameter not equal to one. When $\epsilon = 1$, the solutions of (1) are always simple waves, which means that they are of the form

 0022-2488/84/040884-05$02.50

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} F(\theta(x,t)) \\ G(\theta(x,t)) \end{pmatrix},$$

with F and G real-valued functions. The results of this paper depend on the fact that (1) has two dependent variables, whereas in the case of a simple wave there is essentially only one.

The *order* of a symmetry generator is the order of the highest derivative in it. Higher order means order greater than one. The classical Lie symmetries are first order and are of the form

$$\begin{pmatrix} au_t + bu_x + c \\ av_t + bv_x + d \end{pmatrix},$$

where a,b,c,d are functions of t, x,u,v. They can be solved by the method of characteristics and are equivalent to ode's in t, x,u,v. To find the first-order generators for (1) we assume the most general possible form for this generator and substitute it into the formula which a symmetry generator must satisfy:

$$\frac{\partial}{\partial t}\begin{pmatrix} P(x,t,u,v,u_x,v_x) \\ Q(x,t,u,v,u_x,v_x) \end{pmatrix} = \left\langle \begin{pmatrix} P(x,t,u,v,u_x,v_x) \\ Q(x,t,u,v,u_x,v_x) \end{pmatrix}, \begin{pmatrix} (u+\epsilon v)u_x \\ (v+\epsilon u)v_x \end{pmatrix} \right\rangle.$$

Expanding this using the definition of the bracket, we find, after a calculation, that all first-order generators of (1) are of the form

$$\begin{pmatrix} au_x + b \\ cv_x + d \end{pmatrix},$$

where a,b,c,d are functions of x,t,u,v satisfying the eight equations

$$\frac{\partial a}{\partial v} = \frac{\epsilon}{g-f}(c-a), \quad \frac{\partial a}{\partial t} - f\frac{\partial a}{\partial x} = b + \epsilon d,$$

$$\frac{\partial c}{\partial u} = \frac{\epsilon}{f-g}(a-c), \quad \frac{\partial c}{\partial t} - g\frac{\partial c}{\partial x} = d + \epsilon b,$$

$$\frac{\partial b}{\partial v} = 0, \quad \frac{\partial b}{\partial t} - f\frac{\partial b}{\partial x} = 0,$$

$$\frac{\partial d}{\partial u} = 0, \quad \frac{\partial d}{\partial t} - g\frac{\partial d}{\partial x} = 0.$$

Here $f = u + \epsilon v$ and $g = v + \epsilon u$. These equations have many solutions. Those where b and d are independent of x and t, and where a and c are linear in x and t are of special interest. There are exactly three linearly independent solutions of this type having nonzero b and d. They are

$$a = c = (1+\epsilon)t, \quad b = d = 1, \tag{2}$$

$$a = c = -x, \quad b = u, \quad d = v, \tag{3}$$

$$\begin{cases} a = \dfrac{f}{1-\epsilon}x + \left(\dfrac{f^2}{1-\epsilon} + u^2 + \epsilon v^2\right)t, \\ c = \dfrac{g}{1-\epsilon}x + \left(\dfrac{g^2}{1-\epsilon} + v^2 + \epsilon u^2\right)t, \\ b = u^2, \quad d = v^2. \end{cases} \tag{4}$$

Solution (2) corresponds to Galilean invariance. The corresponding one-parameter family is

$$\begin{pmatrix} u(x + s(1+\epsilon)t, t) + s \\ v(x + s(1+\epsilon)t, t) + s \end{pmatrix}.$$

Solution (3) is a scaling symmetry[10] with one-parameter family

$$\begin{pmatrix} e^s u(e^{-s}x, t) \\ e^s v(e^{-s}x, t) \end{pmatrix}.$$

Solution (4) has no simple interpretation.

Viewing x,t as the dependent variables and u,v as the independent variables is called the *hodograph transformation*.[11] First derivatives transform according to the formulas

$$v_t = Jx_u, \quad u_t = -Jx_v,$$
$$v_x = -Jt_u, \quad u_x = Jt_v, \tag{5}$$

where $J = u_x v_t - u_t v_x = (g-f)u_x v_x = (1-\epsilon)(v-u)u_x v_x$ is the Jacobian of the transformation. When $\epsilon = 1$, the transformation is singular. Using (5), we can rewrite (1) as a linear system

$$\begin{pmatrix} x_v + ft_v \\ x_u + gt_u \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{6}$$

If $(P,Q)^T$ generates a symmetry for (1) then $(x_u P + x_v Q, t_u P + t_v Q)^T$ will be the corresponding symmetry for (6). Derivatives with respect to x and t can be turned into derivatives with respect to u and v using the chain rule for total derivatives:

$$D_x = u_x D_u + v_x D_v,$$
$$D_t = u_t D_u + v_t D_v.$$

Conversely,

$$D_u = x_u D_x + t_u D_t,$$
$$D_v = x_v D_x + t_v D_t.$$

Hence it is possible to rewrite the derivatives of u and v in P and Q as derivatives of x and t to obtain pdf's in the new dependent variables.

Consider the first-order symmetries of (1). Their generators become

$$\begin{pmatrix} bx_u + dx_v + \dfrac{ag - cf}{g-f} \\ bt_u + dt_v + \dfrac{c-a}{g-f} \end{pmatrix}. \tag{7}$$

For these to be linear in x and t, b and d must be independent of x and t, and a and c must be linear in x and t. The three linearly independent cases where (7) is linear and truly first order are already listed as solutions (2), (3), and (4). These three generators are obtained by applying three linear operators to the vector $(x,t)^T$. By a well-known theorem in the theory of symmetries,[12] these three linear operators are recursion operators for (6). That means that by applying any of these operators to a known symmetry generator of (6), we obtain a new symmetry generator of (6). Since the operators all have order one, the new generators are of a higher order by one. Thus (6) and, in turn, (1) have infinitely many higher-order symmetries.

The three recursion operators are

$$R_1 = D_u + D_v + \begin{pmatrix} 0 & 1+\epsilon \\ 0 & 0 \end{pmatrix}, \tag{8}$$

$$R_2 = uD_u + vD_v + \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \tag{9}$$

$$R_3 = u^2 D_u + v^2 D_v + \begin{pmatrix} 0 & \dfrac{-\epsilon u^2 - 2uv - \epsilon v^2}{1-\epsilon} \\ \dfrac{1}{1-\epsilon} & \dfrac{2(u+v)}{1-\epsilon} \end{pmatrix}. \tag{10}$$

Any sum or composition of recursion operators is again a recursion operator, so these generate an infinite algebra of recursion operators. Their commutators must also be recusion operators. These are

$$[R_1, R_2] = R_1,$$

$$[R_1, R_3] = 2R_2 + \frac{3-\epsilon}{1-\epsilon} I,$$

$$[R_2, R_3] = R_3.$$

(I is the identity.)

III. THE HIGHER-ORDER SYMMETRIES

We can apply the three recursion operators repeatedly to the wide variety of first-order symmetries to obtain a multitude of higher-order symmetries. Now the higher-order symmetries of the known examples of equations possessing recursion operators are all obtained by repeated application of the recursion operators, so it is a reasonable conjecture that all of the symmetries of (6) and (1) are obtained in this manner.[13] An important subclass of symmetry generators for (1) consists of those independent of x and t. (The various well-known infinite sequences of symmetries are of this type.[14]) Since R_1, R_2, and R_3 do not depend on x and t explicitly, they preserve this property of a generator. As it is impractical to write out the symmetry generators of higher order in full owing to their length, we will classify such symmetries according to their highest-order terms.

A generator for (6) transforms to a generator of (1) according to

$$\begin{pmatrix} P \\ Q \end{pmatrix} \to \begin{pmatrix} u_x & u_t \\ v_x & v_t \end{pmatrix} \begin{pmatrix} P \\ Q \end{pmatrix} = \begin{pmatrix} u_x & fu_x \\ v_x & gv_x \end{pmatrix} \begin{pmatrix} P \\ Q \end{pmatrix} = \begin{pmatrix} (P+fQ)u_x \\ (P+gQ)v_x \end{pmatrix}.$$

Let $\equiv$ denote equality of highest-order terms. The highest-order part of the recursion operators looks like $\mathscr{D} = bD_u + dD_v$. So if R is one of the recursion operators, we have

$$R \begin{pmatrix} P \\ Q \end{pmatrix} \equiv \mathscr{D} \begin{pmatrix} P \\ Q \end{pmatrix} = \begin{pmatrix} \mathscr{D}P \\ \mathscr{D}Q \end{pmatrix}.$$

Hence R acts on the transformed symmetry by

$$R \begin{pmatrix} (P+fQ)u_x \\ (P+gQ)v_x \end{pmatrix} \equiv \begin{pmatrix} (\mathscr{D}P + f\mathscr{D}Q)u_x \\ (\mathscr{D}P + g\mathscr{D}Q)v_x \end{pmatrix}.$$

Assuming that P and Q have order two or more, it follows that $R \equiv \mathscr{D}$. Using the chain rule for total derivatives, we get $\mathscr{D} = b(x_u D_x + t_u D_t) + d(x_v D_x + t_v D_t)$. But from (1) we have

$$D_t \equiv \begin{cases} fD_x & \text{on} \quad u, u_x, \dots, \\ gD_x & \text{on} \quad v, v_x, \dots . \end{cases}$$

Thus on u and derivatives of u

$$\begin{aligned} \mathscr{D} &\equiv b(x_u D_x + t_u fD_x) + d(x_v D_x + t_v fD_x) \\ &= b(x_u + t_u f)D_x = b(-gt_u + t_u f)D_x \\ &\quad + b(f-g)t_u D_x = (b/u_x)D_x. \end{aligned}$$

All but the first are true equalities. A similar result holds for v. To sum up,

$$R \equiv \begin{cases} (b/u_x)D_x & \text{on} \quad u, u_x, \dots, \\ (d/v_x)D_x & \text{on} \quad v, v_x, \dots . \end{cases} \tag{11}$$

This gives the action of the recursion operator with highest-order part $bD_u + dD_v$ on symmetries of (1) whose generators have order two or more.

An explicit but somewhat lengthy calculation shows that there are exactly three linearly independent symmetry generators of (1) of order two that are independent of x and t. Their highest-order terms are

$$\begin{pmatrix} \dfrac{u_{xx}}{u_x^2} \\ \dfrac{v_{xx}}{v_x^2} \end{pmatrix}, \quad \begin{pmatrix} \dfrac{uu_{xx}}{u_x^2} \\ \dfrac{vv_{xx}}{v_x^2} \end{pmatrix}, \quad \begin{pmatrix} \dfrac{u^2 u_{xx}}{u_x^2} \\ \dfrac{v^2 v_{xx}}{v_x^2} \end{pmatrix}. \tag{12}$$

These could also be obtained by applying R_1^2, $R_1 R_2$, and $R_2^2 + R_2$, respectively, to the trivial symmetry generator $(x,t)^T$ of (6) and transforming back to the uv coordinates. (R_3 cannot be used in such a manner because terms containing x and t would result.) Applying (11) to (12) we see that the higher-order symmetry generators independent of x and t obtained by the recursion operators have highest-order terms

$$\begin{pmatrix} \dfrac{p(u)u_n}{u_x^n} \\ \dfrac{p(v)v_n}{v_x^n} \end{pmatrix},$$

where u_n and v_n are the nth x derivatives and p is a polynomial of degree $2(n-1)$.

IV. ONE-DIMENSIONAL ISENTROPIC COMPRESSIBLE FLOW

The equations are

$$\begin{aligned} &w_t + ww_x + (1/\rho)p_x = 0, \\ &\rho_t + \rho w_x + w\rho_x = 0, \end{aligned} \tag{13}$$

where w is fluid velocity, ρ is fluid density, and p is pressure. We will assume $p = A\rho^\gamma$, but we can get rid of A by scaling. This results in the system.

$$\begin{aligned} &w_t + ww_x + \gamma\rho^{\gamma-2}\rho_x = 0, \\ &\rho_t + \rho w_x + w\rho_x = 0. \end{aligned} \tag{14}$$

With definitions

$$\begin{matrix} u(x,t) \\ v(x,t) \end{matrix} = \frac{1+\gamma}{4}\Big[w(x,t) \pm \frac{2\gamma^{1/2}}{\gamma^{-1}}\rho^{(\gamma-1)/2}(x,t)\Big], \quad (15)$$

the fluid equations (14) take on the form (1) with $\epsilon = (3-\gamma)/(1+\gamma)$. (The forbidden value $\epsilon = 1$ corresponds to the forbidden value $\gamma = 1$.)

We note that when $\gamma = 2$, (14) is equivalent to the sloping beach equations[15]: Letting $\sigma = 2^{1/2}\rho^{1/2}$, we get

$$w_t + ww_x + 2\sigma\sigma_x = 0,$$

$$\sigma_t + \tfrac{1}{2}\sigma w_x + w\sigma_x = 0.$$

These are equivalent to (1) with $\epsilon = \frac{1}{3}$, so our results also hold for these equations.

System (1) is essentially the system of equations for the Riemann invariants,[16] but minus signs in those equations make (1) more convenient when performing long calculations.

The main point of this paper can now be made: The compressible flow equations (14) have infinitely many higher-order symmetries. [These are in one-to-one correspondence with the symmetries of (1).] Using the chain rule, we note that

$$\begin{pmatrix} w_s \\ \rho_s \end{pmatrix} = \begin{pmatrix} w_u & w_v \\ \rho_u & \rho_v \end{pmatrix}\begin{pmatrix} u_s \\ v_s \end{pmatrix};$$

hence if $(P,Q)^T$ is a symmetry generator of (1), then $J^{-1}(\widetilde{P},\widetilde{Q})^T$ is a symmetry generator of (14), where J is the Jacobian matrix of the transformation (15) and $\widetilde{P},\widetilde{Q}$ means that we have used (15) to substitute for u and v wherever they occur in P,Q. Direct manual calculation of even the second-order symmetries, independent of x and t, is almost hopeless in the w,ρ coordinates, but the results of the previous paragraphs assure us of symmetries of all higher orders. It is unknown whether the more complex fluid systems (e.g., two-dimensional flow) that are not equivalent to linear systems have higher-order symmetries.

We now use the theory of Hamiltonian pde's. The Hamiltonian structure[17] of (14) is given by

$$\begin{pmatrix} w \\ \rho \end{pmatrix}_t = \begin{pmatrix} 0 & -D_x \\ -D_x & 0 \end{pmatrix}\begin{pmatrix} \partial/\partial w \\ \partial/\partial\rho \end{pmatrix}\Big[\frac{1}{2}\rho w^2 + \frac{1}{\gamma-1}\rho^\gamma\Big]. \quad (16)$$

In fact, (16) is a one-dimensional restriction of the Hamiltonian structure of three-dimensional compressible fluid flow considered by others.[18–20] In a future paper, we show how this compressible Hamiltonian structure passes over to the incompressible one.[21]

A pdf T is a conserved density of (14) if

$$D_t T = D_x X$$

for some other pdf X. In the formal variational calculus, all functions are assumed to vanish rapidly at infinity, so the above is equivalent to

$$D_t \int_{-\infty}^{\infty} T\,dx = \int_{-\infty}^{\infty} D_x X\,dx = X(\infty) - X(-\infty) = 0.$$

Note that if P is a symmetry of an evolution equation $u_t = K$, and if T is a conserved density, then $V(P)T$ is also a conserved density. For P to be a symmetry means that if $u_s = P$, then $D_t D_s = D_s D_t$. Thus $D_t V(P)T = D_t D_s T = D_s D_t T = D_s D_x X = V(P)D_x X = D_x V(P)X$.

The Hamiltonian $\frac{1}{2}\rho w^2 + (1/(\gamma-1))\rho^\gamma$ is a conserved density of (14), the energy density, in fact. By Noether's theorem[22] for Hamiltonian pde's, every conserved density T (possibly depending on higher derivatives) gives rise to a symmetry (called Hamiltonian) according to

$$\begin{pmatrix} w \\ \rho \end{pmatrix}_s = \begin{pmatrix} 0 & -D_x \\ -D_x & 0 \end{pmatrix}\begin{pmatrix} E_w \\ E_\rho \end{pmatrix}[T],$$

where E_w and E_ρ are Frechet derivatives with respect to w and ρ. We have discovered a new first-order conserved density of compressible flow:

$$T = \rho_x/(w_x{}^2 - \gamma\rho^{\gamma-3}\rho_x{}^2). \quad (17)$$

In fact, we can give the flux X explicitly:

$$D_t T = D_x((w_x\rho - w\rho_x)/(w_x{}^2 - \gamma\rho^{\gamma-3}\rho_x{}^2)),$$

so the proof is an exercise in differentiation. T was originally found by solving $D_t T = D_x X$ in the u,v coordinates of (1) for all T, X of first order, independent of x and t. The unique solution was then transformed to the w,ρ coordinates using (15). Note that although T is formally a conserved density, it may not be meaningful on an arbitrary solution of (14). For example, the denominator of T vanishes identically on simple waves. However, there are many smooth (for a short time, at least) solutions of (14) where T will be a genuine conserved density. More about this unusual conserved density will be given in a future publication.

The density (17) gives rise to a third-order symmetry of (14) corresponding to the third-order symmetry generator of (1) with highest-order terms

$$\begin{pmatrix} u_{xxx}/u_x{}^3 \\ v_{xxx}/v_x{}^3 \end{pmatrix}.$$

Since density (17) is the only first-order density of (14) that is independent of x and t, the other third-order symmetries of (14) that are independent of x and t must be non-Hamiltonian (do not come from conserved densities via Noether's theorem). As E_w and E_ρ double the differential order, and D_x adds one to it, an nth-order conserved density leads to a $(2n+1)$th-order Hamiltonian symmetry. So, only an odd-order symmetry can be Hamiltonian. Which of the higher-odd-order symmetries actually are Hamiltonian and what their corresponding densities are is an open question which we will address in a subsequent publication. If the analog with the KdV equation is correct, there would be a conserved density of each positive-integer order leading to a Hamiltonian symmetry of each odd order.

There is another way infinitely many conserved densities might be produced. If $u_s = P$ is a symmetry and T is a conserved density of $u_t = K$ then $V(P)T$ is also a conserved density. For if $D_t T = D_x X$ then $D_t V(P)T = D_t D_s T = D_s D_t T = D_s D_x X = V(P)D_x X = D_x V(P)X$. We know that (14) has infinitely many symmetries to use in this manner. It is possible, however, that the new conserved densities would be trivial (of the form $T = D_x V$), and the computations are prohibitively lengthy.

ACKNOWLEDGMENTS

I would like to thank S. I. Rosencrans for his help and J. E. Marsden for sending references on Hamiltonian structures of fluid equations. Among these was Ref. 20.

[1]P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[2]N. Kh. Ibragimov and A. B. Shabat, Functional Anal. Appl. **14**, 19 (1980).
[3]N. Kh. Ibragimov and A. B. Shabat, Functional Anal. Appl. **14**, 313 (1980).
[4]G. Wilson, Math. Proc. Cambridge Philos. Soc. **86**, 131 (1979).
[5]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).
[6]See Ref. 2, p. 19.
[7]P. J. Olver, *Application of Lie Groups to Differential Equations, Oxford University Lecture Note Series* (Oxford U.P., New York, 1980).
[8]L. Ovsjannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[9]S. I. Rosencrans, J. Differential Equations **43**, 305 (1982).
[10]See Ref. 9, p. 312.
[11]P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964), p. 512.
[12]See Ref. 7, p. 143.
[13]See Ref. 1, Examples 3–6.
[14]See Ref. 1, Examples 3–6.
[15]Y. Akyildiz, J. Math.Phys. **23**, 1723 (1982).
[16]See Ref. 11, p. 512.
[17]P. J. Olver, Math. Proc. Cambridge Philos. Soc. **88**, 71 (1980).
[18]J. E. Marsden, T. Ratin, and A. Weinstein Preprint PAM-96, University of California, Berkeley, 1982.
[19]J. E. Marsden and A. Weinstein, Phys. D **4**, 394 (1982).
[20]P. J. Morrison and J. M. Greene, Phys. Rev. Lett. **45**, 790 (1980).
[21]P. J. Olver, J. Math. Anal. Appl. **89**, 233 (1982).
[22]See Ref. 17, p. 83.

A Galerkin method to cylindrical KdV equations

Fengsu Chen[a)]

Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

(Received 20 July 1982; accepted for publication 8 April 1983)

The variational method has been used to obtain approximate solutions of the nonlinear Klein–Gordon equation and KdV equation by Hsieh, and the numerical solitary wave solutions of cylindrical nonlinear KdV equations by Chwang and Wu. The author has already used the Galerkin method to obtain approximate solutions of some highly nonlinear KdV equations and Schrödinger equations. In the present paper, we shall employ the Galerkin method to obtain approximate solutions of cylindrical nonlinear equations.

PACS numbers: 03.40.Kf, 02.30.Hq, 47.35. + i

I. INTRODUCTION

Great interest in the propagation of one-dimensional long waves has been generated by the exact solutions of the KdV equation found by Miura *et al.*[1] Recently, growing attention has been devoted to the study of solutions of highly nonlinear KdV equations and cylindrical nonlinear KdV equations.

The variational method has been used to obtain approximate solutions of the nonlinear Klein–Gordon equation and KdV equation by Hsieh,[2] and the numerical solitary wave solutions of cylindrical nonlinear KdV equations by Chwang and Wu.[3]

The author has already used the Galerkin method to obtain approximate solutions of some highly nonlinear KdV equations and Schrödinger equations.[4] In the present paper we shall employ the Galerkin method to obtain approximate solutions of cylindrical nonlinear equations.

II. GENERAL SCHEME OF THE GALERKIN METHOD

The method to be developed is a generalization of the Galerkin method[5] from the perspective of the variational method. The essence of the Galerkin method may be described as follows. Take the differential equation

$$L[x(t)] = 0. \tag{1}$$

A trial solution is taken in the form

$$x = \sum_{i=1}^{N} C_i x_i(t), \tag{2}$$

where $\{x_i(t)\}$ is a set of given functions. Then choose a set of weighting functions $\{W_i(t)\}$. The parameters $\{C_i\}$ are to be determined by the following set of algebraic equations:

$$\int \left(L\left[\sum_{i=1}^{N} C_i x_i\right]\right) W_j \, dt = 0, \quad j = 1,...,N. \tag{3}$$

The weighting functions $\{W_i(t)\}$ were originally chosen by Galerkin to be identical to $\{x_i(t)\}$. Now there is some ambiguity in the choice of suitable form of L, since, for example, $f(x)L[x(t)] = 0$ will be equivalent to (1) for any $f(x)$. The scheme proposed will remove this ambiguity.

Consider a differential equation schematically represented by

$$L[x(t),\alpha] = 0, \tag{4}$$

where α is a parameter. As examples, take the following differential equations representing oscillations with linear and nonlinear dampings:

$$x'' + 2\alpha x' + x = 0 \tag{5}$$

and

$$x'' + \alpha(x')^3 + x = 0 \tag{6}$$

when $\alpha = 0$; a functional J can be found and a proper variational formulation exists. Then $\Delta J = 0$ will lead to the relation

$$\int L[x(t);0]\Delta x \, dt = 0. \tag{7}$$

This L is now unequivocally determined by the variational formulation. Then for α sufficiently small, we expect that a similar relation

$$\int L[x(t);\alpha]\Delta x \, dt = 0 \tag{8}$$

is also valid. This is the essence of the general scheme of the proposed Galerkin method.

III. APPROXIMATE SOLUTIONS OF THE CYLINDRICAL KdV EQUATION

We consider the following equations which govern the propagation of three-dimensional long waves in water of constant depth:

$$\eta_t + \tfrac{1}{2}\eta_r + (1 + \tfrac{3}{2}\epsilon\eta)\eta_r + (\mu^2/6)\eta_{rrr} = 0. \tag{9}$$

The fluid is assumed to be incompressible and inviscid and the motion irrotational. The vertical z axis is measured from the still water level. The displacement of the free surface from the still water level is $\eta(r,t)$ and the solid bottom is at $z = -h$. The fluid is supposed to be unbounded in the radial (r) direction. $\epsilon = a_0/h$, $\mu = k_0 h$, $k_0 = 2\pi/\lambda_0$, λ_0 is the wavelength and a_0 is the characteristic wave amplitude. ϵ and μ^2 are assumed to be of the same order.

Equation (9) is the cylindrical KdV equation for waves moving in the positive r direction. This equation was first formulated by Maxon and Viecelli[6] for an ion-acoustic wave propagating in a collisionless plasma, and it was applied later to long waves propagating in water of constant depth by Miles.[7]

[a)] On leave from Wuhan Institute of Hydraulic and Electric Engineering, Wuhan, People's Republic of China.

 0022-2488/84/040889-03$02.50

The two-dimensional version of Eq. (9) is the well-known KdV equation:

$$\eta_t + (1 + \tfrac{3}{2}\epsilon\eta)\eta_x + (\mu^2/6)\eta_{xxx} = 0\,. \tag{10}$$

Thus Eq. (9) is the KdV equation plus an additional term $\eta/2r$.

Now we consider the cylindrical KdV equation

$$\eta_t + \tfrac{1}{2}\eta_r + (1 + \tfrac{3}{2}\epsilon\eta)\eta_r + (\mu^2/6)\eta_{rrr} = 0\,. \tag{9}$$

Let

$$\sigma = r - t\,, \quad \tau = \epsilon t; \tag{11}$$

then (9) becomes

$$\eta_\tau + \tfrac{1}{2}\eta/\tau + \tfrac{3}{2}\eta\eta_\sigma + \tfrac{1}{6}(\mu^2/\epsilon)\eta_{\sigma\sigma\sigma} = 0\,. \tag{12}$$

In order to eliminate the term η/τ, suppose

$$u = \eta(\tau/\tau_0)^{1/2}\,, \quad \hat{t} = \tau_0^{1/2}(\tau^{1/2} - \tau_0^{1/2})\,. \tag{13}$$

Then (12) becomes

$$u_{\hat{t}} + 3uu_\sigma + (\mu^2/3\epsilon)(\tau/\tau_0)^{1/2}u_{\sigma\sigma\sigma} = 0\,. \tag{14}$$

We use the Galerkin method:

$$\int_0^{\hat{t}}\int_{-\infty}^{+\infty}\left[u_{\hat{t}} + 3uu_\sigma + \frac{\mu^2}{3\epsilon}\left(\frac{\tau}{\tau_0}\right)^{1/2}u_{\sigma\sigma\sigma}\right]\Delta u\, d\sigma\, d\hat{t} = 0\,. \tag{15}$$

Let us take the trial solution of the form

$$u = A(\sigma,\hat{t})\varphi\left[\frac{(\sigma - \sigma_0) - \hat{t}}{(\tau/\tau_0)^{1/4}}\right],$$

where A is a slowly varying function of $(\sigma,\hat{t})$:

$$\begin{aligned}
u_{\hat{t}} &= A_{\hat{t}}\varphi - A\dot{\varphi}(\tau/\tau_0)^{-1/4}\,,\\
u_\sigma &= A_\sigma\varphi + A\dot{\varphi}(\tau/\tau_0)^{-1/4}\,,\\
uu_\sigma &= A\varphi\,[A_\sigma\varphi + A\dot{\varphi}(\tau/\tau_0)^{-1/4}]\\
&= AA_\sigma\varphi^2 + A^2\varphi\dot{\varphi}(\tau/\tau_0)^{-1/4}\,,\\
u_{\sigma\sigma\sigma} &= A_{\sigma\sigma\sigma}\varphi + 3A_{\sigma\sigma}\dot{\varphi}(\tau/\tau_0)^{-1/4}\\
&\quad + 3A_\sigma\ddot{\varphi}(\tau/\tau_0)^{-1/2} + A\dddot{\varphi}(\tau/\tau_0)^{-3/4}\,,\\
\Delta u &= \varphi\Delta A + A\Delta\varphi\,.
\end{aligned}$$

Substituting these expressions into (15), we obtain

$$(\mu^2/3\epsilon)\dddot{\varphi} + 3A\varphi\dot{\varphi} - \dot{\varphi} = 0 \tag{16}$$

if $A = \text{const}$. This integrates immediately to

$$(\mu^2/3\epsilon)\ddot{\varphi} + 3A\varphi^2/2 - \varphi + C_1 = 0 \tag{17}$$

with a further integration to

$$(\mu^2/6\epsilon)\dot{\varphi}^2 + A\varphi^3/2 - \varphi^2/2 + C_1\varphi + C_2 = 0\,. \tag{18}$$

In the special case when φ and its derivatives tend to zero at ∞, then $C_1 = C_2 = 0$. Therefore the equation may be written

$$(\mu^2/3\epsilon)\dot{\varphi}^2 + A\varphi^3 - \varphi^2 = 0\,, \tag{19}$$

$$\varphi = \operatorname{sech}^2\frac{\sqrt{3\epsilon}}{2\mu}\left[\frac{(\sigma - \sigma_0) - \hat{t}}{(\tau/\tau_0)^{1/4}}\right],$$

$$\varphi = \operatorname{sech}^2(\sqrt{3\epsilon}/2\mu)(\tau_0/\tau)^{1/4}[(\sigma - \sigma_0) - \hat{t}\,] \tag{20}$$

if $A = 1$. From (11) and (13) we obtain

$$\begin{aligned}
\eta &= (t_0/t)^{1/2}\operatorname{sech}^2\{\sqrt{3\epsilon}/2\mu(t_0/t)^{1/4}\\
&\quad\times[(r - r_0) - (t - t_0) - \epsilon t_0^{1/2}(t^{1/2} - t_0^{1/2})]\}\,.
\end{aligned} \tag{21}$$

The position for the maximum η at a given time $t_{\max}$ is found at

$$r_{\max} = r_0 + (t_{\max} - t_0) + \epsilon t_0^{1/2}(t_{\max}^{1/2} - t_0^{1/2})\,, \tag{22}$$

and the phase velocity of the wave is

$$C = \left.\frac{dr}{dt}\right|_{r = r_{\max}} = 1 + \frac{\epsilon}{2}\left(\frac{t_0}{t_{\max}}\right)^{1/2}. \tag{23}$$

For simplification, we let $r_0 = t_0$. Therefore the maximum value of the wave amplitude becomes

$$\eta_{\max} = (r_0/r_{\max})^{1/2} + O(\epsilon)\,, \tag{24}$$

as is expected from the leading order solution. The phase velocity can be written as

$$C = 1 + (\epsilon/2)(r_0/r_{\max})^{1/2} + O(\epsilon^2)\,. \tag{25}$$

The behavior of the wave predicated by Eqs. (24) and (25) is in agreement with the numerical solution of Chwang and Wu.[3]

IV. FAMILY OF APPROXIMATE SOLUTIONS OF THE CYLINDRICAL KdV EQUATION

In the past, almost all the analytical solutions for cylindrical solitary waves contained a sech^2 function; the main difference between them is how the wave amplitude and wave width change during propagation.

A cylindrical outgoing solitary wave in water of constant depth is governed by the equation

$$\eta_\tau + \tfrac{1}{2}\eta/\tau + \tfrac{3}{2}\eta\eta_\sigma + \tfrac{1}{6}(\mu^2/\epsilon)\eta_{\sigma\sigma\sigma} = 0\,. \tag{12}$$

In order to eliminate the term η/τ and obtain a general solution, let

$$u = \eta(\tau/\tau_0)^{1/2}\,, \quad \hat{t} = \tau_0^n(\tau^{(1-n)} - \tau_0^{(1-n)})\,. \tag{26}$$

Then (12) becomes

$$u_{\hat{t}} + \frac{3}{2(1-n)}\left(\frac{\tau}{\tau_0}\right)^{n-1/2}uu_\sigma + \frac{1}{6(1-n)}\frac{\mu^2}{\epsilon}\left(\frac{\tau}{\tau_0}\right)^n u_{\sigma\sigma\sigma} = 0\,. \tag{27}$$

We use the Galerkin method

$$\int_0^{\hat{t}}\int_{-\infty}^{+\infty}\left[u_{\hat{t}} + \frac{3}{2(1-n)}\left(\frac{\tau}{\tau_0}\right)^{n-1/2}uu_\sigma + \frac{1}{6(1-n)}\frac{\mu^2}{\epsilon}\left(\frac{\tau}{\tau_0}\right)^n u_{\sigma\sigma\sigma}\right]\Delta u\, d\sigma\, d\hat{t} = 0\,. \tag{28}$$

In order to obtain a general solution, let us take the trial solution of the form

$$u = A(\sigma,\hat{t})\varphi\left[\frac{(\sigma - \sigma_0) - g(\hat{t})}{(\tau/\tau_0)^{n/2}}\right], \tag{29}$$

where A and the derivative of g are slowly varying functions. Substituting (29) into (28) as before, we obtain

$$\frac{1}{6(1-n)}\frac{\mu^2}{\epsilon}\dddot{\varphi} + \frac{3A}{2(1-n)}\left(\frac{\tau}{\tau_0}\right)^{n-1/2}\varphi\dot{\varphi} - \dot{\varphi}\dot{g} = 0\,. \tag{30}$$

Let

$$A = (\tau_0/\tau)^{n-1/2}\,, \quad \dot{g} = \tfrac{1}{2}(1-n)\,;$$

then (30) becomes

$$(\mu^2/3\epsilon)\dddot{\varphi} + 3\varphi\dot{\varphi} - \dot{\varphi} = 0 .$$

Integrating twice, we obtain

$$(\mu^2/6\epsilon)\dot{\varphi}^2 + \varphi^3/2 - \varphi^2/2 + C_1\varphi + C_2 = 0 . \quad (31)$$

In the special case when φ and its derivatives tend to zero at ∞, then $C_1 = C_2 = 0$. Therefore the equation may be written

$$(\mu^2/3\epsilon)\dot{\varphi}^2 + \varphi^3 - \varphi^2 = 0 , \quad (32)$$

$$\varphi = \operatorname{sech}^2 \frac{\sqrt{3\epsilon}}{2\mu}\left[\frac{(\sigma-\sigma_0)-g(\hat{t})}{(\tau/\tau_0)^{n/2}}\right],$$

$$\varphi = \operatorname{sech}^2 \{\sqrt{3\epsilon}/2\mu)(\tau_0/\tau)^{n/2}[(\sigma-\sigma_0)-g(\hat{t})]\} . \quad (33)$$

From (26) we obtain

$$u = (\tau_0/\tau)^{n-1/2}\operatorname{sech}^2\{\sqrt{3\epsilon}/2\mu)(\tau_0/\tau)^{n/2}[(\sigma-\sigma_0)-g(\tau)]\} ,$$

$$\eta = (\tau_0/\tau)^n \operatorname{sech}^2\{(\sqrt{3\epsilon}/2\mu)(\tau_0/\tau)^{n/2}[(\sigma-\sigma_0)-g(\tau)]\} , \quad (34)$$

where

$$g(\tau) = (1/2(1-n))\tau_0^n(\tau^{(1-n)} - \tau_0^{(1-n)}) + g(\tau_0) . \quad (35)$$

Therefore from (34) and (35), we note that with $n = \frac{1}{2}$ the result is our solution, i.e., Eq. (21), and with $n = \frac{2}{3}$ the result is Cumberbatch's solutions.[8]

As in previous sections, we shall let $t_0 = r_0$ for simplicity; then the position of the maximum value of η is found at

$$r_{\max} = t_{\max} + (\epsilon t_0^n/2(1-n))[t^{(1-n)} - t^{(1-n)}] .$$

Therefore the maximum value of η is

$$\eta_{\max} = (r_0/r_{\max})^n + O(\epsilon) ,$$

and the phase velocity is

$$C = 1 + (\epsilon/2)(r_0/r_{\max})^n + O(\epsilon^2) .$$

V. DISCUSSION

By means of a Galerkin method we have obtained approximate solutions of cylindrical nonlinear KdV equations. It may be remembered, in contrast to many other asymptotic methods, that the perturbation method is not easily adapted to solve highly nonlinear equations, and the variational method is not intrinsically a perturbation method and therefore is well adapted to treat a certain class of nonlinear equations.

The method presented above is simple in concept and straightforward in application, yet it yields a great deal of information. It is clear that much work is still needed to answer the many questions raised by the proposed Galerkin method and to explore its wide ranging potentials.

ACKNOWLEDGMENTS

The author wishes to thank Professor D. Y. Hsieh for many stimulating discussions and thank the Division of Applied Mathematics of Brown University for their hospitality.

[1]R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204–09 (1968).
[2]D. Y. Hsieh, J. Math. Phys. **16**, 1630–36 (1975).
[3]A. T. Chwang and T. Y. Wu, Proc. IUTAM Symp. (1976).
[4]F. Chen, "A Galerkin method to strongly nonlinear KdV equations and Schrödinger equations" in *Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems*, AIP Conference Proceedings No. 88 (AIP, New York, 1982), pp. 319–326.
[5]F. Chen and D. Y. Hsieh, "A Galerkin method and nonlinear oscillations and waves", J. Math. Phys. **24**, 1041–1046 (1983).
[6]S. Maxon and J. Viecelli, Phys. Rev. Lett. **17**, 1614–1616 (1974).
[7]J. Miles, J. Fluid Mech. **84**, 181–191 (1978).
[8]E. Cumberbatch, Phys. Fluids **21**, 374–376 (1978).

On the limit from the intermediate long wave equation to the Benjamin–Ono equation

P. M. Santini,[a] M. J. Ablowitz, and A. S. Fokas
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York, 13676

(Received 26 April 1983; accepted for publication 29 June 1983)

The intermediate long wave equation is a physically important singular integrodifferential equation containing a parameter, referred to here as δ. For $\delta \to \infty$ it reduces to the Benjamin–Ono equation. It has been recently shown that the inverse scattering transform schemes of the above equations have certain significant differences. Here it is shown that for $\delta \to \infty$, the inverse scattering transform scheme of the intermediate long wave equation reduces to that of the Benjamin–Ono equation.

PACS numbers: 03.40.Kf

I. INTRODUCTION

The intermediate long wave (ILW) equation arises in the context of long internal gravity waves in a stratified fluid with finite depth.[1–6] Moreover, it arises in other circumstances as well (e.g., long waves in a stratified shear flow.[7,1] The ILW equation can be taken in the dimensionless form

$$u_t + (1/\delta)u_x + 2uu_x + Tu_{xx} = 0;$$
$$(Tv)(x) \doteq \frac{1}{2\delta}\int_{-\infty}^{\infty} \coth\left(\frac{\pi(\xi - x)}{2\delta}\right)v(\xi)\,d\xi, \tag{1}$$

where Cauchy principal-value integrals are assumed if needed. In the internal gravity waves problem, the parameter δ can be thought as the ratio of depth to wavelength; Eq. (1) reduces to the Korteweg–deVries (KdV) equation[8] as $\delta \to 0$ (shallow-water limit)

$$u_t + 2uu_x + (\delta/3)u_{xxx} = 0, \tag{2}$$

and to the Benjamin–Ono (BO) equation[9] as $\delta \to \infty$ (deep-water limit)

$$u_t + 2uu_x + Hu_{xx} = 0; \quad (Hv)(x) \doteq \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{v(\xi)}{\xi - x}\,d\xi. \tag{3}$$

Equations (1)–(3) are special cases of an equation discussed by Whitham.[10] N-soliton solutions, an infinite number of conserved quantities, Bäcklund transformations, and Lax pairs for the ILW and BO equations have been established in Refs. 3, 4, 6, 11, and in 12–16, respectively.

The inverse scattering transform (IST) scheme, a method for solving suitable initial-value problems for certain nonlinear equations, was discovered in connection with the KdV equation.[17] The IST schemes for the ILW and BO equations have been recently established in Refs. 11, 18, and 19, respectively. The limit of the IST scheme of the ILW equation to that of the KdV equation ($\delta \to 0$) is rather straightforward and was given in Refs. 6 and 18. However, the limit of the IST scheme of the ILW equation to that of the BO equation ($\delta \to \infty$) presents certain difficulties. This is a reflection of the fact that the IST schemes of the ILW and BO equations have significant differences. Actually the IST scheme of the ILW equation is conceptually similar to that of the KdV equation (see subsec. IIA below); on the other hand, the IST scheme of the BO equation is similar to that of the Kadomtsev–Petviashvili equation (a two-dimensional analog of the KdV equation).[20] Hence the limit process $\delta \to \infty$ in a sense provides a limit between two different types of IST formalisms, appropriate for one and two dimensional problems, respectively.

In this paper, it is established that as $\delta \to \infty$, the IST scheme of the ILW equation reduces to that of the BO equation.

II. REVIEW OF THE IST FOR THE ILW AND BO EQUATIONS

A. The ILW equation

The following results can be found in Ref. 18.

1. The direct scattering problem

The direct scattering problem of the ILW equation is based on the x-part (4a) of the "Lax pair"

$$L_\delta W \doteq iW_x^+ + [\zeta_+(\lambda) + 1/2\delta](W^+ - W^-) = -uW^+, \tag{4a}$$

$$iW_t^\pm - 2i\zeta_+ W_x^\pm + W_{xx}^\pm + [\mp iu_x - Tu_x + \rho_\delta]W^\pm = 0, \tag{4b}$$

where

$$\zeta_\pm(\lambda) \doteq \pm\frac{\lambda e^{\pm\lambda\delta}}{e^{\lambda\delta} - e^{-\lambda\delta}} \mp \frac{1}{2\delta}$$
$$\rho_\delta(\lambda) \doteq \lambda\zeta_+ + (\lambda/2)^2 + \nu, \ \nu$$

is an arbitrary constant, λ is a constant and is interpreted as a spectral parameter. Given u, Eq. (4a) defines a ***Riemann–Hilbert problem in a strip of the complex x plane***; $W^\pm(x)$ represent the boundary values of functions [i.e., $\psi^\pm(x) = \lim_{y\to 0}\psi(x \pm iy)$] analytic in the horizontal strips between $\operatorname{Im} z = 0$ and $\operatorname{Im} z = \pm 2\delta$, and periodically extended vertically. Importantly, Eq. (4a) can also be solved

[a] Permanent address: Istituto di Fisica, Università di Roma, 00183 Roma, Italy.

without appealing to Riemann–Hilbert theory since it can be viewed as a differential-difference equation. This follows from the periodicity condition $\psi^-(x) = \psi^+(x + 2i\delta)$.

Let us concentrate on the $(+)$ functions and let $M, \overline{M}$ denote the $(+)$ "left" eigenfunctions, while $\overline{N}, N$ denote the $(+)$ "right" eigenfunctions. These eigenfunctions, in addition to solving (4), also satisfy the following boundary conditions

$$M \to 1, \quad \overline{M} \to e^{i\lambda x + \lambda\delta} \quad \text{as } x \to -\infty; \quad \overline{N} \to 1, \quad N \to e^{i\lambda x + \lambda\delta} \quad \text{as } x \to +\infty. \tag{5}$$

The eigenfunctions $M, \overline{M}, N, \overline{N}$ satisfy the following Fredholm integral equations

$$\begin{pmatrix} M(x,\lambda) \\ \overline{M}(x,\lambda) \end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x + \lambda\delta} \end{pmatrix} + \int_{-\infty}^{\infty} G_+(x,y,\zeta_+(\lambda)) u(y) \begin{pmatrix} M(y,\lambda) \\ \overline{M}(y,\lambda) \end{pmatrix} dy, \tag{6a}$$

$$\begin{pmatrix} N(x,\lambda) \\ \overline{N}(x,\lambda) \end{pmatrix} = \begin{pmatrix} e^{i\lambda x + \lambda\delta} \\ 1 \end{pmatrix} + \int_{-\infty}^{\infty} G_-(x,y,\zeta_+(\lambda)) u(y) \begin{pmatrix} N(y,\lambda) \\ \overline{N}(y,\lambda) \end{pmatrix} dy, \tag{6b}$$

where

$$G_\pm(x,y,\zeta_+(\lambda)) \doteq \frac{1}{2\pi} \int_{C_\pm} dp \frac{e^{ip(x-y)}}{p - [\zeta_+(\lambda) + 1/2\delta](1 - e^{-2p\delta})}, \tag{7}$$

where $C_\pm$ are the contours $\operatorname{Re}(p \mp i0)$.

The eigenfunctions $M, \overline{N}, N$ are related through the scattering equation

$$M(x,\lambda) = a(\lambda)\overline{N}(x,\lambda) + \theta(\zeta_+(\lambda) + 1/2\delta) b(\lambda) N(x,\lambda), \tag{8}$$

where

$$a(\lambda) \doteq 1 + \frac{1}{2i\zeta_+(\lambda)} \int_{-\infty}^{\infty} dy\, u(y) M(y,\lambda);$$

$$b(\lambda) \doteq -\frac{1}{2i\zeta_-(\lambda)\delta} \int_{-\infty}^{\infty} dy\, u(y) M(y,\lambda) e^{-i\lambda y - \lambda\delta} \tag{9}$$

and

$$\theta(\lambda) = 1 \quad \text{for } \lambda > 0, \quad \theta(\lambda) = 0 \quad \text{for } \lambda < 0.$$

The "bound states" correspond to those $\hat{\lambda}_l$ for which

$$a_l \doteq a(\hat{\lambda}_l) = 0, \quad l = 1,2,\ldots,n, \tag{10a}$$

$$M_l(x) \doteq M(x,\hat{\lambda}_l) = b(\hat{\lambda}_l) N(x,\hat{\lambda}_l) \doteq b_l N_l(x). \tag{10b}$$

2. The inverse scattering problem

The solution of the inverse problem is based on Eq. (8). Given $a(\lambda)$, $b(\lambda)$, and appropriate information about the bound states, find $M, \overline{N}, N$. In order to view (8) as a *Riemann–Hilbert problem in the complex $\zeta_+(\lambda)$ plane*, one needs to establish analyticity properties in $\zeta_+(\lambda)$ for the eigenfunctions $M, \overline{N}, N$. The Kernels of the integral equations satisfied by $M, \overline{N}$ are $(+)$ and $(-)$ functions, respectively, in $\zeta_+(\lambda)$, i.e., they are analytically extendable in the appropriate regions of the $\zeta_+(\lambda)$ plane. Since the forcing in both cases is unity, Fredholm theory implies that the solutions M and $\overline{N}$ are also $(+)$ and $(-)$ functions in $\zeta_+(\lambda)$, provided that there exist no solutions to Eqs. (6) (when δ is finite, it can be shown that for suitable potentials this is actually the case). Furthermore, Eq. (9a) implies that $a(\zeta_+)$ is a $(+)$ function in λ.

In order to solve (8), one needs to establish an *analytic connection* or symmetry condition between $\overline{N}$ and N. This follows from the relationship

$$N(x,\lambda) = \overline{N}(x,-\lambda) e^{i\lambda x + \lambda\delta}, \tag{11}$$

which is a consequence of

$$G_\pm(x,y,\lambda) = G_\pm(x,y,-\lambda) e^{i\lambda(x-y)}. \tag{12}$$

Equation (8), using the above analytic properties of M, $\overline{N}$, and a, as well as Eq. (11), defines a Riemann–Hilbert problem in $\zeta_+(\lambda)$. From this, the following integral equation is obtained (see Appendix A):

$$\overline{N}(\zeta_+) - \frac{1}{2\pi i} \int_{-1/2\delta}^{\infty} \frac{\rho(\zeta'_+) N(\zeta'_+)}{\zeta'_+ - (\zeta_+ - i0)} d\zeta'_+ = 1 + i \sum_{j=1}^{n} \frac{C_j N_j}{\zeta_+ - \zeta_{+j}}, \tag{13}$$

where

$$C_j \doteq -i \frac{b(\lambda)}{a_{\zeta_+}(\zeta_+(\lambda))} \bigg|_{\lambda = \hat{\lambda}_j}, \quad \rho \doteq \frac{b(\lambda)}{a(\lambda)}. \tag{14}$$

The Gel'fand–Levitan–Marcenko equation given in Ref. 18 can be easily obtained by taking an appropriate Fourier transform of (13), supplied by the analytic information (11).

We shall also need the following relationship, which is obtained from (13) asymptotically as $\zeta_+ \to \infty$ (see Appendix A):

$$u^+(x) = \frac{1}{2\pi i} \int_{-1/2\delta}^{\infty} \rho(\zeta_+) N(\zeta_+) d\zeta_+ - i \sum_{j=1}^{n} C_j N_j;$$

$$u^+(x) \doteq \frac{1}{4i\delta} \int_{-\infty}^{\infty} \cosh\left(\frac{\pi(y - x - i0)}{2\delta}\right) u(y) dy. \tag{15}$$

And finally, the reality of $u(x)$ implies that $u(x) = u^+(x) - u^-(x) = u^+(x) + (u^+(x))^*$.

Equation (13) defines N in terms of $\rho, C_j, \hat{\lambda}_j$, and Eq. (15) defines u^+ in terms of ρ, C_j, N. Hence Eqs. (13) and (15) define $u^+(x)$ in terms of ρ, C_j, $\hat{\lambda}_j$, the so-called scattering data. However, the scattering data need only be evaluated at time $t = 0$ [i.e., in terms of the initial data $u(x, 0)$ only] since their evolution is known.

$$\{\hat{\lambda}_j(t) = \hat{\lambda}_j(0),$$

$$C_j(t) = C_j(0) \exp\left(i\hat{\lambda}_j [\hat{\lambda}_j \coth(\hat{\lambda}_j\delta) - 1/\delta] t\right)\}_{j=1}^{n},$$

$$\rho(\lambda, t) = \rho(\lambda, 0) \exp\left[i\lambda(\lambda \coth \lambda\delta - 1/\delta) t\right]. \tag{16}$$

The above evolution of the scattering data follows easily from the t-part (4b) of the Lax pair.

B. The BO equation

The following results can be found in Ref. 19.

1. The direct scattering problem

The direct scattering problem of the BO equation is based on the x part (17a) of the "Lax pair":

$$Lw \doteq iw_x^+ + \lambda(w^+ - w^-) = -uw^+, \tag{17a}$$

$$iw_t^\pm - 2i\lambda w_x^\pm + w_{xx}^\pm - 2i[u]_x^\pm w^\pm = -\rho w^\pm, \tag{17b}$$

where

$$[u]^\pm \doteq \pm u/2 + (1/2i)Hu. \tag{17c}$$

Given u, Eq. (17a) defines a differential *Riemann–Hilbert problem in the complex x-plane*; $w^\pm(x)$ represent the boundary values of functions analytic in the upper (+) and lower (−) half x-plane, i.e., $w^\pm(x) = \lim_{y\to 0} w(x \pm iy)$.

The eigenfunctions $m, \bar m, n, \bar n$ satisfy the boundary conditions

$$m \to 1, \quad \bar m \to e^{i\lambda x} \quad \text{as } x \to -\infty;$$

$$\bar n \to 1, \quad n \to e^{i\lambda x} \quad \text{as } x \to +\infty, \tag{18}$$

and are characterized through the following Fredholm integral equations:

$$\begin{pmatrix} m(x,\lambda) \\ \bar m(x,\lambda)\end{pmatrix} = \begin{pmatrix} 1 \\ e^{i\lambda x}\end{pmatrix} + \int_{-\infty}^{\infty} g_+(x,y,\lambda)u(y)\begin{pmatrix} m(y,\lambda) \\ \bar m(y,\lambda)\end{pmatrix}dy, \tag{19a}$$

$$\begin{pmatrix} n(x,\lambda) \\ \bar n(x,\lambda)\end{pmatrix} = \begin{pmatrix} e^{i\lambda x} \\ 1\end{pmatrix} + \int_{-\infty}^{\infty} g_-(x,y,\lambda)u(y)\begin{pmatrix} n(y,\lambda) \\ \bar n(y,\lambda)\end{pmatrix}dy, \tag{19b}$$

where g_+, g_- are the (+) and (−) parts of sectionally holomorphic function

$$g(x,y,\Lambda) = \frac{1}{2\pi}\int_0^\infty dp\, \frac{e^{i(x-y)p}}{p-\Lambda}, \tag{20}$$

and Λ denotes the complex extension of λ, i.e.,

$$g_\pm(x,y,\lambda) = \frac{1}{2\pi}\int_0^\infty dp\,\frac{e^{i(x-y)p}}{p-(\lambda \pm i0)}. \tag{21}$$

The eigenfunctions $m, \bar n, n$ are related through the scattering equation

$$m(x,\lambda) = \bar n(x,\lambda) + \theta(\lambda)\beta(\lambda)n(x,\lambda), \tag{22}$$

where

$$\beta(\lambda) = i\int_{-\infty}^{\infty} u(y)m(y,\lambda)e^{-i\lambda y}dy. \tag{23}$$

2. The inverse scattering problem

The solitons of the ILW equation correspond to "bound states" which are generated from the zeros of $a(\lambda)$. However, in Eq. (22) [which is the analog of (8)], the coefficient of $\bar n$ is 1. Hence the solitons of the BO equation are generated through a different mechanism; the integral equations (19), in contrast to the integral equations (6), may have homogeneous solutions Φ_j, for some λ_j, where $\lambda_j < 0$, i.e.,

$$\Phi_j(x) = \int_{-\infty}^{\infty} g(x,y,\lambda_j)u(y)\Phi_j(y)dy, \quad \lambda_j < 0. \tag{24}$$

The kernels of the integral equations for $m, \bar n$ are (+) and (−) functions respectively in λ. Hence

$$m(x,\lambda) = 1 + \sum_1^n \frac{C_j\Phi_j(x)}{\lambda-\lambda_j} + m_+(x,\lambda), \tag{25}$$

$$\bar n(x,\lambda) = 1 + \sum_1^n \frac{\bar C_j\Phi_j(x)}{\lambda-\lambda_j} + \bar n_-(x,\lambda), \tag{26}$$

where $m_+, \bar n_-$ are (+) and (−) functions, respectively, in λ. It turns out that

$$C_j = \bar C_j = -i, \; j = 1,2,\ldots,n.$$

In order to view Eq. (22) as a Riemann–Hilbert problem in the complex λ plane, one needs to establish analytic information about n and $\bar n$. This follows from

$$\frac{\partial}{\partial\lambda}(n(x,\lambda)e^{-i\lambda x}) = f(\lambda,t)e^{-i\lambda x}\bar n(x,\lambda);$$

$$f(\lambda) \doteq -\frac{1}{2\pi\lambda}\int_{-\infty}^{\infty} u(y)n(y,\lambda)dy. \tag{27}$$

Equation (27) is a consequence of

$$\frac{\partial}{\partial\lambda} g_\pm(x,y,\lambda) = -\frac{1}{2\pi\lambda} + i(x-y)g_\pm(x,y,\lambda). \tag{28}$$

Using (28), one also finds that

$$\lim_{\lambda\to\lambda_j}[\bar n(x,\lambda) - i\Phi_j(x)/(\lambda-\lambda_j)] = (x+\gamma_j)\Phi_j. \tag{29}$$

Equation (22), using (25)–(27), and (29), defines a *nonlocal Riemann–Hilbert problem in the complex λ plane* which is equivalent to the integral equation

$$n(x,\lambda,t) = \frac{1}{2\pi}\int_0^\infty h(x,t,\lambda,l)\beta(l,t)n(x,t,l)dl + \sum_{l=1}^n \Phi_l(x,t)h(x,t,\lambda,\lambda_l) = v(x,t,\lambda), \tag{30}$$

$$(x+\gamma_j(t))\Phi_j(x,t) - \frac{1}{2\pi i}\int_0^\infty \frac{\beta(l,t)n(x,t,l)}{\lambda-\lambda_j}dl + i\sum_{\substack{l=1\\ l\neq j}}^n \frac{\Phi_l(x)}{\lambda_j-\lambda_l} = 1, \tag{31}$$

where

$$v(x,t,\lambda) \doteq \int_0^\lambda (f(l,t)e^{ix(\lambda-l)} + f_s(l)e^{ix\lambda})dl; \quad f_s(\lambda) = \frac{1}{\lambda\ln\lambda}, \tag{32a}$$

$$h(x,t,\lambda,l) \doteq e^{i(\lambda-l)x}\int_\infty^x v(\xi,\lambda)e^{-i(\lambda-l)\xi}d\xi, \quad l>0, \tag{32b}$$

$$h(x,t,\lambda,\lambda_j) \doteq e^{i(\lambda-l)x}\int_\alpha^x v(\xi,\lambda)e^{-i(\lambda-\lambda_j)\xi}d\xi + e^{i(\lambda-\lambda_j)(x-\alpha)}\int_0^\lambda\left(\frac{f(l,t)e^{i\alpha(\lambda-l)}}{\lambda_j-l} + \frac{f_s(l)}{\lambda_j}\right)dl. \tag{32c}$$

The following equation is also valid:

$$[u]^+ = \frac{1}{2\pi i}\int_0^\infty \beta(\lambda)n(x,\lambda)d\lambda + i\sum_1^n\Phi_j(x), \tag{33}$$

and assuming u real, $u(x) = u^+(x) + (u^+(x))^*$.

Equations (30)–(33) define $[u]^+$ in terms of $\lambda_j, \gamma_j, \beta(\lambda), f(\lambda)$. However, the scattering data need only be evaluated at time $t=0$, since their evolution is known from (17b):

$$\{\lambda_j(t)=\lambda_j(0),\quad \gamma_j(t)=2\lambda_j t+\gamma_j(0)\}_{j=1}^{n},$$
$$\beta(\lambda,t)=\beta(\lambda,0)e^{i\lambda^2 t},\quad f(\lambda,t)=f(\lambda,0)e^{i\lambda^2 t}. \tag{34}$$

III. THE LIMIT FROM THE ILW EQUATION TO THE BO EQUATION

In this section, we will show how the IST scheme of the BO equation can be obtained from the IST scheme of the ILW equation.

A. The direct scattering problem

As noticed in Ref. 18, the limit $\delta\to\infty,\lambda>0$, of the Lax pair (4) goes directly to the system (19); the strips between $\operatorname{Im} z=0$ and $\operatorname{Im} z=\pm 2\delta$ become the upper and lower half z-plane, and then $w^{\pm}(x)=\lim_{\delta\to\infty}W^{\pm}(x)$ are nothing but the boundary values of functions analytic in the upper $(+)$ and lower $(-)$ half z-plane.

It is straightforward to show [see (B2)] that $\lim_{\delta\to\infty}G_{\pm}(x,y,\zeta_+(\lambda))=g_{\pm}(x,y,\lambda),\lambda>0$, where the Green functions $G_{\pm}$ and $g_{\pm}$ are defined in (7) and (21), respectively, then the Jost functions of (17a) are solutions of the Fredholm equations (19) and can be obtained, for $\lambda>0$, in the following way:

$$m(x,\lambda)=\lim_{\delta\to\infty}M(x,\lambda) \tag{35a}$$

$$\bar m(x,\lambda)=\lim_{\delta\to\infty}\overline{M}(x,\lambda)e^{-\lambda\delta}, \tag{35b}$$

$$n(x,\lambda)=\lim_{\delta\to\infty}N(x,\lambda)e^{-\lambda\delta}, \tag{35c}$$

$$\bar n(x,\lambda)=\lim_{\delta\to\infty}\overline{N}(x,\lambda). \tag{35d}$$

The analytic information about $G_{\pm}$ (and, consequently, about N) contained in (11) and (12) are apparently lost in the limit $\delta\to\infty$, from which we find the identity $g_{\pm}=g_{\pm}$. Nevertheless, one may show that taking the derivative of (12) with respect to λ,

$$G_{\pm,\lambda}(x,y,\lambda)=i(x-y)G(x,y,\lambda)+G_\lambda(x,y,-\lambda)e^{i\lambda(x-y)}, \tag{36}$$

and *then* taking the limit $\delta\to\infty$ of this equation, one gets [see, (B3)–(B5)] the nontrivial equation (28). Analogously, by taking the λ derivative of Eq. (6b) for $\overline{N}(x,-\lambda)$, enriched by the property (11),

$$\begin{aligned}(N(x,\lambda)e^{-i\lambda x-\lambda\delta})_\lambda &=\int_{-\infty}^{\infty}G_\lambda(x-y,-\lambda)u(y)N(y,\lambda)e^{-i\lambda y-\lambda\delta}dy\\&+\int_{-\infty}^{\infty}G(x-y,\lambda)e^{i\lambda(x-y)}u(y)(N(y,\lambda)e^{-i\lambda y-\lambda\delta})dy,\end{aligned} \tag{37}$$

and *then* taking the limit $\delta\to\infty$, one gets the analytic connection formula (27). This highly nontrivial formula is derived at this stage as a consequence of the noncommutativity of the two operators $\lim_{\delta\to\infty}$ and $\partial/\partial\lambda$. It will be rederived later (perhaps in a more satisfactory way) from the scattering problem.

Using Eqs. (35) for $\lambda>0$, together with the symmetry condition (11) for the case $\lambda<0$, one can take the limit of the scattering equation (8). Specifically,

$$\lim_{\delta\to\infty}a(\lambda)=\begin{cases}1, & \lambda>0, & (38a)\\ d(\lambda), & \lambda<0, & (38b)\end{cases}$$

$$\lim_{\delta\to\infty}b(\lambda)=e^{-\lambda\delta}\beta(\lambda),\quad \lambda>0, \tag{39a}$$

$$\lim_{\delta\to\infty}\delta e^{\lambda\delta}b(\lambda)=\frac{i}{2\lambda}\int_{-\infty}^{\infty}u(y)\bar m(y,-\lambda)dy,\quad \lambda<0, \tag{39b}$$

where $d(\lambda)\doteq 1+i\int_{-\infty}^{\infty}u(y)\bar m(y,-\lambda)e^{i\lambda y}dy$, and $\beta(\lambda)$ is defined in (23). Then in the limit $\delta\to\infty$, Eq. (8) goes to Eq. (22) for $\lambda>0$, and it goes to

$$\bar m(x,-\lambda)=d(\lambda)n(x,-\lambda) \tag{40}$$

for $\lambda<0$.

Finally, (38) and (39), together with (14) and (40), imply that

$$\lim_{\delta\to\infty}\rho(\lambda)=e^{-\lambda\delta}\beta(\lambda),\quad \lambda>0, \tag{41a}$$

$$\lim_{\delta\to\infty}\delta e^{\lambda\delta}\rho(\lambda)=i\pi f(-\lambda),\quad \lambda<0, \tag{41b}$$

with $f(\lambda)$ defined in (27).

The solution of the inverse problem for the BO equation will be obtained taking the limit $\delta\to\infty$ of Eq. (13). However, in order to do that, we must still characterize the asymptotics of the bound states $\hat\lambda_j=\hat\lambda_j(\delta), j=1,2,\ldots,n$ of the ILW equation.

B. The bound states

As shown in Ref. 18 for every finite δ the $\hat\lambda_l$'s are simple zeros of $a(\lambda)$ and lie on that portion of the imaginary axes contained in the fundamental sheet of the λ plane: $\hat\lambda_j=ik_j$, $0<k_j<\pi/\delta$.

In order to establish the asymptotics of $\hat\lambda_j$, we will study the equation $a_j\doteq a(\hat\lambda_j,\delta)=0$ for large δ with the following ansatz:

$$\hat\lambda_j=(i\pi/\delta)(\alpha_0^{(j)}+\alpha_1^{(j)}/\delta+\alpha_2^{(j)}/\delta^2+O(\delta^{-3})), \tag{42}$$

and the restriction $0\leqslant\alpha_0^{(j)}\leqslant 1$, which is a direct consequence of the property $0<k_j<\pi/\delta$.

Substituting ansatz (42) into the equation $a_j=0$ evaluated for large δ, one gets an equation in inverse powers of δ. In order to equate to zero the coefficients of the $O(1)$ term, the following conditions must be satisfied (see Appendix C):

$$\alpha_0^{(j)}=1, \tag{43}$$

$$M_j(x)=\delta\mu_0^{(j)}(x)+\mu_1^{(j)}(x)+O(\delta^{-1}), \tag{44}$$

$$\int_{-\infty}^{\infty}u(y)\mu_0^{(j)}(y)dy=-2i\lambda_j, \tag{45a}$$

where

$$\lambda_j\doteq 1/2\alpha_1^{(j)}; \tag{45b}$$

while, equating to zero the coefficient of the $O(\delta^{-1})$ term, one gets

$$\int_{-\infty}^{\infty} u(y)\mu_1^{(j)}(y)dy = -2i\nu_{+j}, \tag{46a}$$

where

$$\nu_{\pm_j} \doteq i(\pi/2) \mp 2\lambda_j^2 \alpha_2^{(j)}. \tag{46b}$$

As a consequence of result (43), the property $k_j < \pi/\delta$ implies that $\alpha_1^{(j)}$ (and then λ_j) is negative, otherwise arbitrary.

Moreover

$$\zeta_{\pm_j} = \zeta_\pm(\hat{\lambda}_j) = \pm\lambda_j + \nu_{\pm j}/\delta + O(\delta^{-2}), \quad \delta \gg 1. \tag{47}$$

Substituting the expansion (44) into Eq. (6a) evaluated for large δ and equating to zero the coefficients of the first two terms, we get the integral equations satisfied by $\mu_0^{(j)}(x)$ and $\mu_1^{(j)}(x)$:

$$(K\mu_0^{(j)})(x) = 0, \tag{48a}$$

$$(K\mu_1^{(j)})(x) = 1 + \int_{-\infty}^{\infty} g_1(x, y, \lambda_j)u(y)\mu_0^{(j)}(y)dy, \tag{48b}$$

where

$$(K\eta)(x) \doteq \eta(x) - \int_{-\infty}^{\infty} g(x\ y, \lambda_j)u(y)\eta(y)dy, \tag{49a}$$

and

$$g_1(x, y, \lambda_j) \doteq (\nu_{+_j} + \tfrac{1}{2})g_\lambda(x, y, \lambda_j) - i/4\lambda_j \tag{49b}$$

is the coefficient of the $O(\delta^{-1})$ term in the expansion of $G_+(x, y, \hat{\lambda}_j)$ when $\delta \gg 1$ [see (B6)]:

$$G_+(x, y, \hat{\lambda}_j) = g(x, y, \lambda_j) + (1/\delta)g_1(x, y, \lambda_j) + O(\delta^{-2}). \tag{49c}$$

Equation (48a) shows that the leading term of the expansion of M_j is a solution of the homogeneous equation (24), corresponding to the eigenvalue λ_j. $\mu_1^{(j)}(x)$ is the solution of the inhomogeneous equation (48b) and the necessary and sufficient condition for such a solution to exist in [see (D5)] $\alpha_2^{(j)} = 1/(4\lambda_j^2)$, whereupon then $\nu_{\pm_j} = \frac{1}{2}(i\pi \mp 1)$. So Eq. (48b) becomes

$$(\hat{K}\mu_1^{(j)})(x) = \frac{i\pi}{2}\int_{-\infty}^{\infty} (x-y)g(x, y, \lambda_j)u(y)\mu_0^{(j)}(y)dy. \tag{50}$$

In both Eqs. (48a) and (50), the solutions $\mu_0^{(j)}(x)$ and $\mu_1^{(j)}(x)$ are defined up to a multiplicative constant that can be determined using Eqs. (45a) and (46a).

Formulas (42), (44), (45a), (47) allow us to evaluate the limit of $c_j \doteq -i(b_j/a_j') \doteq -i(b(\hat{\lambda}_j)/a_{\zeta_+}(\hat{\lambda}_j))$.

$$\begin{aligned} b_j &= -\frac{1}{2i\delta\zeta_{-j}}\int_{-\infty}^{\infty} u(y)M_j(y)e^{-i\lambda_j y - \lambda_j\delta}dy \\ &= -\frac{1}{2i\lambda_j}\int_{-\infty}^{\infty} u(y)\mu_0^{(j)}(y)dy \\ &\quad + O(\delta^{-1}) = 1 + O(\delta^{-1}), \quad \delta \gg 1, \end{aligned} \tag{51}$$

so, as a consequence,

$$N_j = M_j/b_j = \delta\mu_0^{(j)}(x) + O(1), \quad \delta \gg 1. \tag{52}$$

For definition (9),

$$\begin{aligned} a_j' &= -\frac{1}{2i\delta\zeta_{+_j}^2}\int_{-\infty}^{\infty} u(y)M_j(y)dy \\ &\quad + \frac{1}{2i\delta\zeta_{+_j}}\int_{-\infty}^{\infty} u(y)M_j'(y)dy. \end{aligned} \tag{53}$$

$M_j'(y) \doteq M_+(y, \zeta_+(\lambda))|_{\lambda=\lambda_j}$ satisfies the equation

$$\begin{aligned} M_j'(x) &- \int_{-\infty}^{\infty} G_+(x, y, \hat{\lambda}_j)u(y)M_j'(y)dy \\ &= \int_{-\infty}^{\infty} G'_{+_j}(x, y)u(y)M_j(y)dy, \end{aligned} \tag{54}$$

with $G'_{+_j}(x, y) = G_{+_\zeta}(x, y, \zeta_+(\lambda))|_{\lambda=\lambda_j}$. The asymptotics of G'_{+_j} (see Appendix B)

$$G'_{+_j}(x, y) = g_\lambda(x, y, \lambda_j) + O(\delta^{-1}), \quad \delta \gg 1 \tag{55}$$

and the condition (D5) of the existence of solutions of equation (54) for large δ, suggest the following ansatz:

$$M_j'(x) = \delta^2\tilde{\mu}_0(x) + \delta\tilde{\mu}_1(x) + O(1), \quad \delta \gg 1. \tag{56}$$

Substituting (56) into (54), using (55) and (49c) we get, for the first two orders in δ,

$$(K\tilde{\mu}_0^{(j)})(x) = 0, \tag{57a}$$

$$\begin{aligned} (K\tilde{\mu}_1^{(j)})(x) &= \int_{-\infty}^{\infty} g_\lambda(x, y, \lambda_j)u(y)\mu_0^{(j)}(y)dy \\ &\quad + \int_{-\infty}^{\infty} g_1(x, y, \lambda_j)u(y)\tilde{\mu}_0^{(j)}(y)dy. \end{aligned} \tag{57b}$$

$\tilde{\mu}_0^{(j)}(x)$ is the solution of the homogeneous equation (24) with eigenvalue λ_j, then $\tilde{\mu}_0^{(j)} = \mu\mu_0^{(j)}(x)$, μ constant. Using this relationship, together with (49b), (28), and (D5), one gets $\mu = i/\pi$; then,

$$\begin{aligned} a_j' &= \frac{\delta}{2\pi\lambda_j}\int_{-\infty}^{\infty} u(y)\mu_0^{(j)}(y)dy + O(1) \\ &= -\frac{i\delta}{\pi} + O(1), \quad \delta \gg 1, \end{aligned} \tag{58}$$

and finally,

$$c_j = \pi/\delta + O(\delta^{-2}), \quad \delta \gg 1. \tag{59}$$

C. The inverse scattering problem

We are now ready to take the limit $\delta \to \infty$ of Eq. (13) which is the inverse scattering scheme for the ILW equation.

Let us analytically extend Eq. (13) to $\lambda = -\hat{\lambda}_j$:

$$\begin{aligned} &\left(e^{-i\lambda_j x - \lambda_j\delta} + \frac{iC_j}{\zeta_{-_j} + \zeta_{+_j}}\right)N_j(x)_- \\ &\quad \times \frac{1}{2\pi i}\int_{-1/2\delta}^{\infty} \frac{\rho(\zeta'_+)N(x, \zeta'_+)}{\zeta'_+ + \zeta_{-_j} + i0}d\zeta'_+ \\ &\quad + i\sum_{\substack{l=1 \\ l\neq j}}^{n} \frac{C_l N_l}{\zeta_{-_j} + \zeta_{+_l}} = 1. \end{aligned} \tag{60}$$

Its limit when $\delta \to \infty$, evaluated using (42), (47), (59), (41), and (35), is

$$- \pi[x + \gamma_j(t)]\mu_0^{(j)}(x) - \frac{1}{2\pi i} \times \int_0^\infty \frac{\beta(\lambda)n(\lambda)}{\lambda - \lambda_j} d\lambda - i\pi \sum_{\substack{l=i \\ l \neq j}}^{n} \frac{\mu_0^{(l)}(x)}{\lambda_l - \lambda_j} = 1, \tag{61}$$

where

$$\gamma_j(t) \doteq -\tilde{C}_j(t)/\pi + 1/2\pi\lambda_j - i/2\lambda_j, \tag{62}$$

and $\tilde{C}_j(t)$ characterizes the second term of the expansion of C_j for Large δ:

$$C_j(t) = (\pi/\delta)(1 + \tilde{C}_j(t)/\delta + O(\delta^{-2})), \quad \delta \gg 1. \tag{63}$$

Asymptotically in x, Eq. (61) reads $-\pi x \mu_0^{(j)}(x) = 1$, then

$$\mu_0^{(j)}(x) = -1/\pi\Phi_j(x), \tag{64}$$

where $\Phi_j(x)$ is the solution of (24) with the property

$$x\Phi_j(x) \xrightarrow[x \to \infty]{} 1.$$

So in terms of $\Phi_j(x)$, Eq. (61) becomes Eq. (31) and Eq. (45a) becomes

$$\int_{-\infty}^{\infty} u(y)\Phi_j(y)dy = 2\pi i \lambda_j. \tag{65}$$

Let us consider now $\lambda \in \mathbb{R}^+$; in this case Eq. (13) goes directly to

$$\bar{n}(x, \lambda) - \frac{1}{2\pi i} \int_0^\infty \frac{\beta(l)n(x, l)}{l - (\lambda - i0)} dl = 1 - i \sum_{m=1}^{n} \frac{\Phi_m(x)}{\lambda - \lambda_m} \tag{66}$$

(see Appendix C), which implies that

$$- \frac{1}{2\pi i} \int_0^\infty \frac{\beta(l)n(x, l)}{l + i\epsilon} dl = 1 + i \sum_{m=1}^{n} \frac{\Phi_m(x)}{\lambda_m}. \tag{67}$$

The last choice for λ is $\lambda \in \mathbb{R}^-$; in this case,

$$n(x, -\lambda)e^{i\lambda x} - \frac{1}{2\pi i} \int_0^\infty \frac{\beta(l)n(x,l)}{l + i\epsilon} dl - 2\pi i \int_{-\infty}^{0} \frac{lf(-l)\bar{n}(x, -l)}{-le^{2l\delta} + \lambda e^{2\lambda\delta} + i\epsilon} e^{ilx + 2l\delta} dl = 1 + i \sum_{m=1}^{n} \frac{\phi_m(x)}{\lambda_m} + O(\delta^{-1}), \quad \delta \gg 1.$$

Making use of (67), splitting $\int_{-\infty}^{0} dl$ into $\int_{-\infty}^{\lambda} dl + \int_{\lambda}^{0} dl$ and expanding the corresponding integrands, we get

$$n(x,\lambda)e^{-i\lambda x} - \int_0^\lambda f(l)\bar{n}(x, l)e^{-ilx} dl = 0, \quad \lambda > 0 \tag{68}$$

(see Appendix C), which is nothing but the integral form of the analytic connection formula (27).

Formulas (66) and (68) are equivalent to the integral equation (30) and together with Eq. (31) they determine, in principle, the $n(x)$, $\bar{n}(x)$, and Φ_j's, and thus contain all the information one needs to solve the inverse scattering problem associated with the BO equation.

It is remarkable that these three equations are derived from the same Eq. (13) when $\delta \to \infty$, in the three different situations $\lambda = -\hat{\lambda}_j$, $\lambda \in \mathbb{R}^+$, and $\lambda \in \mathbb{R}$.

Finally, the limit of (15) goes directly to formulas (33), showing how to reconstruct $u^+(x)$, and then $u(x) = u^+(x) + (u^+(x))^*$, the solution of the BO equation, from the scattering data.

D. Time evolution

In order to obtain the time evolution of the scattering data, we notice first of all that $\hat{\lambda}_j(t) = \hat{\lambda}_j(0), j = 1,2,...,n$, then $\lambda_j(t) = \lambda_j(0)$ too. Moreover from (16), we get $C_j(t) = C_j(0)(1 - 2\pi\lambda_j t/\delta + O(\delta^{-2}))$, $\delta \gg 1$, while, from (63) at $t = 0$, we get $C_j(0) = (\pi/\delta)(1 + \tilde{C}(0)/\delta + 0(\delta^{-2}))$, $\delta \gg 1$.

Comparing these results with formula (63), we infer the time evolution of $\tilde{C}_j(t)$ and, through (62), the time evolution of $\gamma_j(t)$:

$$\gamma_j(t) = \gamma_j(0) + 2\lambda_j t, \ \gamma_j(0) \doteq -\tilde{C}_j(0)/\delta + (1/2\pi\lambda_j)(1 - i\pi). \tag{69}$$

$\beta(\lambda,t)$ and $f(\lambda,t)$ originate from two different limits ($\lambda > 0$ and $\lambda < 0$) of $\rho(\lambda, t)$. Comparing the limit of Eq. (16),

$$\rho(\lambda,t) = \rho(\lambda,0)e^{\pm i\lambda^2 t}(1 + O(\delta^{-1})), \lambda \gtrless 0, \quad \delta \gg 1,$$

with formulas (41a) and (41b), we infer that

$$\beta(\lambda,t) = \beta(\lambda,0)e^{i\lambda^2 t}, \quad f(\lambda,t) = f(\lambda,0)e^{i\lambda^2 t} \tag{70}$$

APPENDIX A

In this Appendix we will derive formulas (13) and (15) that characterize an alternative approach for solving the inverse problem of the ILW equation, to that given in Ref. 18, which is in terms of a Gelfand–Levitan–Marchenko equation. While the two approaches are equivalent for the ILW equation, it turns out that the one presented here is the most appropriate to describe the limit to the BO equation. Let us divide the scattering equation (8) by $a(\zeta_+)$; the function $M(\zeta_+)/a(\zeta_+)$ is analytic in the upper half ζ_+ plane except for poles (the zeros ζ_{+_j} of a); then

$$\frac{M(x,\zeta_+)}{a(\zeta_+)} = 1 + \mu_+(x,\zeta_+) + i \sum_{j=1}^{n} \frac{C_n N_j}{\zeta_+ - \zeta_{+_j}}, \tag{A1}$$

where $\mu_+(x, \zeta_+)$ is analytic in the upper ζ_+ half plane and N_j and C_j are defined in (10b) and (14), respectively.

Expressing $\theta(\zeta_+ + 1/2\delta)\rho(\zeta_+)N(x,\zeta_+)$ in terms of its (+) and (−) parts,

$$\theta(\zeta_+ + 1/2\delta)\rho(\zeta_+)N(x,\zeta_+) = U^+(x,\zeta_+) - U^-(x,\zeta_-), \tag{A2}$$

$$U^{\pm}(x,\zeta_+) = \frac{1}{2\pi i} \int_{-1/2\delta}^{\infty} \frac{\rho(\zeta'_+)N(x,\zeta'_+)}{\zeta'_+ - (\zeta_+ \pm i0)} d\zeta'_+, \tag{A3}$$

and substituting all of this information into (8), we get $\mu_+(x, \zeta_+) = U^+(x,\zeta_+)$, and Eq. (13).

Equation (15) is obtained by considering Eq. (13) for large ζ_+. In order to do that, we must evaluate the asymptotics of $\bar{N}(x, \zeta_+)$ for large ζ_+.

$$G_-(x, y, \zeta_+) = \frac{1}{4i\delta} \int_{-\infty}^{\infty} \coth\left[\frac{\pi(y' - x - i0)}{2\delta}\right] \times \left(\frac{1}{2\pi} \int_C \frac{e^{ip(y'-y)}(1 - e^{-2p\delta})}{p - (\zeta_+ + 1/2\delta)(1 - e^{-2p\delta})} dp\right) dy'$$

$$= -\frac{1}{4i\delta\zeta_+} \coth[(\pi/2\delta)(y - x - i0)](1 + O(\zeta_+^{-1})), \quad \zeta_+ \gg 1. \tag{A4}$$

Then, using Eq. (6b), we get $\overline{N}(x, \zeta_+) \sim 1 - (1/\zeta_+)u^+(x)$ as $\zeta_+ \to \infty$, and Eq. (13) yields Eq. (15).

APPENDIX B

We will briefly discuss here the procedure used to evaluate certain asymptotic calculations. As a prototype example, consider the integral

$$G_\pm(x,y,\lambda) = \frac{1}{2\pi}\int_{C_\pm} \frac{e^{i(x-y)p}}{p-(\zeta_+ + 1/2\delta)(1-e^{-2p\delta})}\,dp. \tag{B1}$$

Using Cauchy's theorem, we may evaluate the order of magnitude of the contributions about 0 and λ, the two singularities of the integrand. Asymptotically in δ they are $(i/2)e^{i\lambda(x-y)}$ and $-i/4\lambda\delta$, respectively. Then we split the integral $\int_{-\infty}^{\infty} dp = \int_{-\infty}^{0} dp + \int_{0}^{\infty} dp$ and we expand the corresponding integrands; the first term gives a $O(\delta^{-1})$ contribution and the second one gives $g_\pm(x,y,\lambda)$. So

$$G_\pm(x,y,\lambda) = g_\pm(x,y,\lambda) + O(\delta^{-1}), \quad \delta \gg 1. \tag{B2}$$

Exactly the same procedure yields formula (55) and

$$G_{\pm_\lambda}(x,y,\lambda) = g_{\pm_\lambda}(x,y,\lambda) + O(\delta^{-1}), \quad \delta \gg 1, \tag{B3}$$

used in (36). The evaluation of $G_\pm(x,y,-\lambda)e^{i\lambda(x-y)}$ in (36) requires more attention:

$$\begin{aligned} G_{\pm_\lambda}(x,y,-\lambda)e^{i\lambda(x-y)} &= \frac{-2\lambda\delta + 1 - e^{-2\delta\lambda}}{2\pi(e^{\lambda\delta} - e^{-\lambda\delta})^2} \\ &\times \int_{C_\pm} \frac{e^{ip(x-y)}(1-e^{-2\delta(p-\lambda)})}{[p-(\zeta_+ + 1/2\delta)(1-e^{-2p\delta})]^2}\,dp \\ &= \frac{\lambda\delta}{\pi}\int_{C_\pm} \frac{e^{ip(x-y)-2p\delta}}{[p-(\zeta_+ + 1/2\delta)(1-e^{-2p\delta})]^2} \\ &\times dp(1+O(\delta^{-1})). \end{aligned} \tag{B4}$$

Replacing $\zeta_+(\lambda) + 1/2\delta$ with λ (with an exponentially small error) and rescaling p with $p\delta$ we finally get

$$\begin{aligned} G_{\pm_\lambda}(x,y,-\lambda)e^{i\lambda(x-y)} &= \frac{1}{\pi\lambda}\int_{C_\pm} \frac{e^{-2p}}{(1-e^{2p})^2}\,dp(1+O(\delta^{-1})) \\ &= -\frac{1}{2\pi\lambda} + O(\delta^{-1}). \end{aligned} \tag{B5}$$

The evaluation of $G_+(x,y,\hat{\lambda}_j)$ up to terms of order δ^{-1} is performed as follows:

$$\begin{aligned} G_+(x,y,\hat{\lambda}_j) &= \frac{1}{2\pi}\left(\frac{\pi i}{1-2\delta(\zeta_{+_j} + 1/2\delta)}\right. \\ &\left. + \int_{-\infty}^{\infty} \frac{e^{ip(x-y)}}{p-(\zeta_{+_j} + 1/2\delta)(1-e^{-2p\delta})}\,dp\right) \\ &= -\frac{i}{4\lambda_j\delta} + \frac{1}{2\pi}\int_0^\infty \frac{e^{ip(x-y)}}{p-\lambda_j}\left(\frac{\nu_{+_j} + \frac{1}{2}}{\delta(p-\lambda_j)}\right. \\ &\left. + \sum_{n=0}(-1)^n \frac{\lambda_j^n e^{-2np\delta}}{(p-\lambda_j)^n}\right)dp \\ &+ \frac{1}{2\pi\lambda_j}\int_{-\infty}^{0} e^{ip(x-y)+2p\delta} \\ &\times\left(\sum_{n=0}\frac{(-1)^n(p-\lambda_j)^n e^{2np\delta}}{\lambda_j^n}\right)dp + O(\delta^{-2}) \\ &= g(x,y,\lambda_j) + (1/\delta)g_1(x,y,\lambda_j) + O(\delta^{-2}), \quad \delta \gg 1, \end{aligned} \tag{B6}$$

APPENDIX C

In this Appendix, we will discuss the asymptotic behavior of $n(x,\lambda)$, $m(x,\lambda)$, $\bar{n}(x,\lambda)$ when $\lambda \to 0$. The same ideas will also be used to obtain equations (43)–(45).

Let us consider function $n(x,\lambda)$, solution of

$$n_x - i\lambda n = i[un]^+, \quad n \xrightarrow[x\to\infty]{} e^{i\lambda x} \tag{C1}$$

($[h]^+$ indicates the $(+)$ projection of h), or, equivalently, the solution of

$$n(x,\lambda) = e^{i\lambda x} + \int_{-\infty}^{\infty} g_-(x,y,\lambda)u(y)n(y,\lambda)\,dy. \tag{C2}$$

Noticing that $g_-(x,y,\lambda) \sim -1/2\pi \ln\lambda$ as $\lambda \to 0$, Eq. (C2) will be satisfied at the $O(1)$ iff

$$n(x,\lambda) \sim n_0(x)/\ln\lambda, \quad \lambda \to 0, \tag{C3}$$

where $n_0(x)$ satisfies the normalization condition

$$\int_{-\infty}^{\infty} u(y)n_0(y)\,dy = 2\pi. \tag{C4}$$

Substituting (C3) into (C1), we get

$$n_{0_x} = i[un_0]^+. \tag{C5}$$

Then Eq. (C5) and the normalization condition (C4) define the coefficient $n_0(x)$ of the leading term in the asymptotic expansion of $n(x,\lambda)$ when $\lambda \to 0$. In particular, it is easy to show that (C5) and (C4) imply that $n_0(x) \xrightarrow[|x|\to\infty]{} -\ln x$.

In exactly the same way, it is possible to show that

$$m(x,\lambda),\ \bar{n}(x,\lambda) \sim n_0(x)/\ln\lambda, \quad \text{as } \lambda \to 0. \tag{C6}$$

Moreover, using (C6) and (C4), we can easily get

$$\beta(\lambda) \sim 2\pi i/\ln\lambda, \quad f(\lambda) \sim -1/\lambda\ln\lambda, \quad \lambda \to 0. \tag{C7}$$

Formulas (C7), as well as (C3) and (C6), are implicitly used to prove the validity of (67) and to show that the integrals contained in formulas (66) and (68) are well defined.

Formulas (C3)–(C6) supercede the formulas (24) in Ref. 18 (the first of which is incorrect; however, only the order of magnitude of the limit $\lambda \to 0$ was used in Ref. 18. This indicates that $\int_{-\infty}^{\infty} u\,dx = 0$ is not special in the limit $\lambda \to 0$.

Let us now prove by contradiction that $\alpha_0^{(j)} = 1$. Then let us suppose that $0 < \alpha_0^{(j)} < 1$; it follows that

$$\zeta_+(\hat{\lambda}_j) = (\hat{C}_j - \frac{1}{2})\frac{1}{\delta} + O(\delta^{-2}); \quad \hat{C}_j \doteq \frac{\pi\alpha_0^{(j)} e^{i\pi\alpha_0^{(j)}}}{2\sin(\pi\alpha_0^{(j)})},$$

and

$$G_+(x, y, \hat{\lambda}_j) \sim \frac{1}{2\pi}\int_{C_+} \frac{e^{i(x-y)p}}{p - (\hat{C}_j/\delta)(1 - e^{-2p\delta})}\,dp$$
$$\sim -\frac{1}{2\pi}\ln\left(\frac{\hat{C}_j}{\delta}\right).$$

Consequently using the same kind of arguments leading to (C3) and (C4), one can show that Eq. (6a) implies

$$M_j(x, \delta) \sim \frac{M_{0j}(x)}{\ln(\hat{C}_j/\delta)}, \quad \text{with} \int_{-\infty}^{\infty} u(y)M_{0j}(y)dy = 2\pi.$$

Then

$$a_j \sim 1 - \frac{i\pi}{(\hat{C}_j - \frac{1}{2})\ln(\hat{C}_j/\delta)} \sim 1 \neq 0,$$

which contradicts the hypothesis.

Analogously, if $\alpha_0^{(j)} = 0$, $\zeta_+(\hat{\lambda}_j) = (i\pi/2\delta^2)\alpha_1^{(j)} + O(\delta^{-3})$ and $G_+(x, y, \hat{\lambda}_j) \sim -(1/2\pi)\ln(1/2\delta)$. Consequently Eq. (6a) implies

$$M_j(x,\delta) \sim \frac{M_{0j}(x)}{\ln(1/2\delta)}, \quad \text{with} \int_{-\infty}^{\infty} u(y)M_{0j}(y)dy = 2\pi.$$

Then

$$a_j \sim 1 - \frac{2\delta}{\alpha_1^{(j)}\ln(1/2\delta)} \sim -\frac{2\delta}{\alpha_1^{(j)}\ln(1/2\delta)} \neq 0,$$

which again contradicts the hypothesis. So we are left with the only choice $\alpha_0^{(j)} = 1$. In this case,

$$\zeta_{+j} = \lambda_j + v_{+j}/\delta + O(\delta^{-2}),$$

so

$$a_j = 1 + \frac{1}{2i\delta\lambda_j}\int_{-\infty}^{\infty} u(y)M_j(y)dy(1 + O(\delta^{-1}))$$

will be zero only if (44) and (45a) hold.

APPENDIX D

Given the following equation,

$$(Kh)(x) = 1 + C_1\int_{-\infty}^{\infty} u(y)\mu_0^{(j)}(y)dy + C_2\int (x-y)u(y)g(x,y,\lambda_j)\mu_0^{(j)}(y)dy \doteq F(x) \quad \text{(D1)}$$

[the operator K is defined in (49a)], Fredholm theory says that a solution exists iff

$$\int_{-\infty}^{\infty} \phi^*(x)F(x)dx = 0, \quad \text{(D2)}$$

where ψ satisfies the equation $(K^+\psi)(x) = 0$, where K^+ is the adjoint operator of K:

$$(K^+h)(x) = h(x) - u^*(x)\int_{-\infty}^{\infty} g^*(y - x, \lambda_j)h(y)dy. \quad \text{(D3)}$$

As a consequence of the equation $(K^+\psi)^*(x) = 0$, we have that

$$\int_{-\infty}^{\infty} \psi^*(x)F(x)dx = \left[1 + C_1\int_{-\infty}^{\infty} u(y)\mu_0^{(j)}(y)dy\right]\int_{-\infty}^{\infty} \psi^*(x)dx. \quad \text{(D4)}$$

Then the condition $\int_{-\infty}^{\infty} \psi^*(x)dx \neq 0$ implies

$$1 + C_1\int_{-\infty}^{\infty} u(y)\mu_0^{(j)}(y)dy = 0. \quad \text{(D5)}$$

ACKNOWLEDGMENTS

This work was partially supported by the Office of Naval Research under Grant Number N00014-76-C-0867, the National Science Foundation under Grant Number MCS-8202117, and the Air Force Office of Scientific Research under Grant Number 78-3674-D.

[1] R. I. Joseph, J. Phys. A **101**, L225 (1977).
[2] T. Kubota, D. R. S. Ko, and D. Dobbs, J. Hydronaut. **12**, 157 (1978).
[3] R. I. Joseph and R. Egri, J. Phys. A **11**, L97 (1978).
[4] H. H. Chen and Y. C. Lee, Phys Rev. Lett. **43**, 264 (1979).
[5] J. Satsuma and M. J. Ablowitz, in *Nonlinear Partial Differential Equations in Engineering and Applied Science*, edited by R. L. Sternberg, A. J. Kalinowski, and J. S. Papadakis (Marcel Dekker, New York, 1980), p. 397.
[6] J. Satsuma, M. J. Ablowitz, and Y. Kodama, Phys. Lett. A **73**, 283 (1979).
[7] S. A. Maslowe, L. G. Redekopp, J. Fluid Mech. **101**, 321 (1980).
[8] D. J. Korteweg and G. deVries, Philos. Mag. **39**, 422 (1895).
[9] T. B. Benjamin, J. Fluid Mech. **29**, 559 (1967); H. Ono, J. Phys. Soc. Jpn. **39**, 1082 (1975).
[10] G. B. Whitham, Proc. R. Soc. London, Ser. A **229**, 6 (1967); G. B. Whitham, in *Linear and Nonlinear Waves* (Wiley, New York, 1974).
[11] Y. Kodama, J. Satsuma, and M. J. Ablowitz, Phys. Rev. Lett. **46**, 687 (1981).
[12] K. M. Case, Proc. Natl. Acad. Sci. USA **75**, 3563 (1978); **76**, 1 (1979).
[13] H. H. Chen, Y. C. Lee, N. R. Pereira, Phys. Fluids **22**, 187 (1979).
[14] A. Nakamura, J. Phys. Soc. Jpn. **47**, 1701 (1979).
[15] T. L. Boch and M. D. Kruskal, Phys. Lett. A **74**, 173 (1979).
[16] A. S. Fokas, B. Fuchssteiner, Phys. Lett. A **86** 341 (1981).
[17] C. S. Gardner, J. M. Green, M. D. Kruskal, and R. M. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[18] Y. Kodama, M. J. Ablowitz, and J. Satsuma, J. Math. Phys. **23**, 564 (1982).
[19] A. S. Fokas and M. J. Ablowitz, "The Inverse Scattering Transform for the Benjamin–Ono Equation—A Pivot to Multidimensional Problems" (preprint), 1982 (to appear in Stud. in Appl. Math.)
[20] A. S. Fokas and M. J. Ablowitz, "On the Inverse Scattering and Direct Linearizing Transforms of the Kadomtsev–Petviashvili Equation" (preprint), 1982 (to appear in Phys. Lett.).

On a Bäcklund transformation and scattering problem for the modified intermediate long wave equation

Junkichi Satsuma,[a)] Thiab R. Taha, and Mark J. Ablowitz
Department of Mathematics and Computer Science, Clarkson College of Technology, Potsdam, New York 13676

(Received 6 August 1982; accepted for publication 10 June 1983)

In this paper, we give a Bäcklund transformation, an associated linear scattering problem, and a method for finding the conservation laws for the so-called modified intermediate long wave equation. This equation reduces to the modified Korteweg–deVries equation in the shallow water limit.

PACS numbers: 03.40.Kf, 47.10. + g

I. INTRODUCTION

It is well known in the theory of solitons (see, for example, Ablowitz and Segur[1]) that multisoliton solutions of certain nonlinear evolution equations can be obtained by more than one method. Some examples are the inverse scattering transform (IST),[2–4] Bäcklund transformation (BT),[5] and the so-called bilinear theory.[6] The IST scheme gives a method of solving the initial-value problem for a broad class of nonlinear evolution equations.[4] Nevertheless, if we are given a particular evolution equation, it can be difficult to set up an appropriate IST. On the other hand, if we can establish a BT associated with a given evolution equation, then by suitable transformations, the BT usually may be reduced to the IST. In this paper, we shall introduce a BT, an IST, and a method for finding the conservation laws for the so-called modified intermediate long wave (MILW) equation. In the literature, the MILW equation is related to the intermediate long wave (ILW) equation[7,8] in the analogous way that the MKdV is to the KdV.

The ILW equation describes long internal gravity waves in a stratified fluid with finite depth. It is written in a simplified form as

$$u_t + 2uu_x + Tu_{xx} = 0, \tag{1.1}$$

where

$$T(u_x) = \hat{T}(u_x) + (1/\delta)u, \tag{1.2}$$

and

$$\hat{T}(u) = \fint_{-\infty}^{\infty} \left(-\frac{1}{2\delta} \coth \frac{\pi(x-\xi)}{2\delta} \right) u(\xi)\, d\xi \tag{1.3}$$

($\fint_{-\infty}^{\infty}$ denotes the Cauchy principal value integral), and δ is a parameter representing the distance between the boundary and the internal wave layer. In the shallow water limit, $\delta \to 0$, Eq. (1.1) reduces to the Korteweg–deVries (KdV) equation

$$u_t + 2uu_x + (\delta/3)u_{xxx} = 0, \tag{1.4}$$

and in the deep water limit, $\delta \to \infty$, reduces to the Benjamin–Ono (BO) equation

$$u_t + 2uu_x + H(u_{xx}) = 0, \tag{1.5}$$

where H denotes the Hilbert transform operator

$$H(u) = \frac{1}{\pi} \fint_{-\infty}^{\infty} \frac{u(x')}{x' - x}\, dx'. \tag{1.6}$$

Equations (1.1) (finite δ) and (1.5) have recently been solved by IST.[9,10] The scattering problem, and hence solutions by IST, is open for the MILW equation [see (2.4) below].

II. THE DERIVATION OF THE MILW EQUATION

Introducing $\omega = \int^x u\, dx$, Eq. (1.1) may be written as

$$\omega_t + (\omega_x)^2 + T(\omega_{xx}) = 0. \tag{2.1}$$

The BT of Eq. (1.1) is expressed as[8]

$$(\omega' + \omega)_x = \lambda + iT(\omega' - \omega)_x - i\delta^{-1}(\omega' - \omega) + \mu e^{i(\omega' - \omega)}, \tag{2.2a}$$

$$\begin{aligned}(\omega' - \omega)_t = {} & -(\delta^{-1} + \lambda)(\omega' - \omega)_x + i(\omega' + \omega)_{xx} \\ & - i(\omega' - \omega)_x T(\omega' - \omega)_x \\ & + i\delta^{-1}(\omega' - \omega)(\omega' - \omega)_x,\end{aligned} \tag{2.2b}$$

where λ and μ are arbitrary parameters. (See Appendix B for the BT in bilinear form for the ILW equation.) We note that if ω satisfies Eq. (2.1), ω', defined by Eq. (2.2), also satisfies Eq. (2.1). Substituting Eq. (2.2a) into (2.2b) and introducing $V = \omega - \omega'$, we have[8]

$$V_t + T(V_{xx}) + \lambda V_x + V_x[\mu e^{iV} + i(T(V_x) - (1/\delta)V)] = 0. \tag{2.3}$$

By setting $V \to -iV$, $t \to (3/\delta)t$, $\lambda = -1/\delta$ and $\mu = 1/\delta$, Eq. (2.3) takes the following form:

$$(\delta/3)V_t + \hat{T}(V_{xx}) + V_x((1/\delta)e^V + \hat{T}(V_x)) = 0. \tag{2.4}$$

We refer to Eq. (2.4) as the MILW equation, which in the limit $\delta \to 0$, using the expansion

$$\hat{T}(V_x) = -(1/\delta)V + \tfrac{1}{3}\delta V_{xx} + \tfrac{1}{45}\delta^3 V_{xxxx} + O(\delta^5) \tag{2.5a}$$

and

$$V = 2\delta u - \tfrac{1}{6}\delta^3 u_{xx} + O(\delta^5), \tag{2.5b}$$

yields the modified Korteweg–deVries (MKdV) equation,

$$u_t + 6u^2 u_x + u_{xxx} = 0. \tag{2.6}$$

III. THE BT OF THE MILW EQUATION

Introducing the following dependent variable transformation:

[a)] Permanent address: General Education, Miyazaki Medical College, Kiyotake, Miyazaki 889-16, Japan.

 0022-2488/84/040900-05$02.50

$$V = \log\,(f^+g^-/f^-g^+), \tag{3.1}$$

Eq. (2.4) can be expressed in bilinear form as

$$(\tfrac{1}{3} i\delta D_t + D_x^2) f^{\pm}\cdot g^{\pm} = 0, \tag{3.2a}$$

$$2i\delta D_x f^-\cdot g^+ = -f^-g^+ + f^+g^-, \tag{3.2b}$$

where we have used the abbreviations

$$f^{\pm}(x) = f(x \pm i\delta), \tag{3.3}$$

and introduced bilinear differential operators defined by Refs. 6 and 11 as

$$D_t^n D_x^m a\cdot b = \Big(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\Big)^n\Big(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\Big)^m \times a(x,t)b(x',t')\big|_{x=x',\,t=t'} \tag{3.4}$$

(see Appendix A for some properties of these operators). An (exact) N-soliton solution of Eq. (3.2) has been obtained by Nakamura.[12] In the limit $\delta \to 0$, these soliton solutions tend to the soliton solutions of the MKdV Eq. (2.6).

As discussed above, the MKdV equation plays the same role with respect to the MILW equation as does the KdV equation with respect to the ILW equation[9] in Fig. 1.

$$\begin{tikzcd}
\boxed{\text{ILW}} \arrow[rr, leftrightarrow, "\text{A Miura Transform}"] \arrow[d, leftrightarrow, "\delta\to 0"] & & \boxed{\text{MILW}} \arrow[d, leftrightarrow, "\delta\to 0"] \\
\boxed{\text{KdV}} \arrow[rr, leftrightarrow, "\text{A Miura Transform}"] & & \boxed{\text{MKdV}}
\end{tikzcd}$$

FIG. 1. The relations of the ILW, KdV, MILW, and MKdV equations.

By analogy to the KdV–MKdV case,[13–17] we have found the following BT of Eq. (2.4) in bilinear form:

$$\big[\tfrac{1}{3} i\delta D_t + i(1/\delta - c)D_x + D_x^2 - \tfrac{1}{4}k^2\big] f^{\pm}\cdot f'^{\pm} = 0, \tag{3.5a}$$

$$\big[\tfrac{1}{3} i\delta D_t + i(1/\delta - c)D_x + D_x^2 - \tfrac{1}{4}k^2\big] g^{\pm}\cdot g'^{\pm} = 0, \tag{3.5b}$$

$$(2iD_x + c) f^-\cdot f'^+ = \mu f^+\cdot f'^-, \tag{3.5c}$$

$$(2iD_x + c) g^-\cdot g'^+ = \mu g^+\cdot g'^-, \tag{3.5d}$$

where

$$c = k\cot k\delta, \tag{3.5e}$$

and

$$\mu = k\csc k\delta. \tag{3.5f}$$

Equation (3.5) relates a solution (f, g) of (3.2) with another solution (f', g') of (3.2) (for more details, see Ref. 16).

It can be shown that the limit, $\delta \to 0$, of the bilinear form of the MILW equation [which is given in Eq. (3.2)] tends to the bilinear form of the MKdV Eq. (2.6), which is

$$(D_t + D_x^3) f\cdot f^* = 0, \tag{3.6a}$$

$$D_x^2 f\cdot f^* = 0, \tag{3.6b}$$

where $u = i(\log f^*/f)_x$, and the limit of the bilinear BT (3.5) reduces to the bilinear BT of the MKdV equation (see Ref. 17):

$$(D_t + \tfrac{3}{4}k^2 D_x + D_x^3) f\cdot f' = 0, \tag{3.7a}$$

$$(D_t + \tfrac{3}{4}k^2 D_x + D_x^3) f^*\cdot f'^* = 0, \tag{3.7b}$$

$$D_x^2 f\cdot f' = \tfrac{1}{4}k^2 f\cdot f', \tag{3.7c}$$

$$D_x^2 f'^*\cdot f^* = \tfrac{1}{4}k^2 f^*\cdot f'^* \tag{3.7d}$$

(where * implies a complex conjugate). Using the following relations (see Ref. 8):

$$(\log f^{\pm})_x = \tfrac{1}{4}(i\hat{T} \pm 1)\{(i\hat{T} + 1)V_x + (1/\delta)(e^V - 1)\}, \tag{3.8}$$

$$(\log g^{\pm})_x = \tfrac{1}{4}(i\hat{T} \pm 1)\{(i\hat{T} - 1)V_x + (i/\delta)(e^V - 1)\}, \tag{3.9}$$

the bilinear BT (3.5) transforms to the BT in physical variables:

$$\tfrac{1}{3} i\delta(V - V')_t + i(1/\delta - c)(V - V')_x + (V + V')_{xx} - \tfrac{1}{2}\hat{T}(V - V')_x\{\hat{T}(V - V')_x + (1/\delta)(e^V - e^{V'})\} - \tfrac{1}{2}(V - V')_x\hat{T}\{\hat{T}(V - V')_x + (1/\delta)(e^V - e^{V'})\} = 0, \tag{3.10a}$$

$$\frac{i}{3}\delta\Big[\hat{T}(V - V') + \frac{1}{\delta}\int_{-\infty}^{x}(e^V - e^{V'})dx\Big]_t - \Big(\frac{1}{\delta} - c\Big)\Big[\hat{T}(V - V')_x + \frac{1}{\delta}(e^V - e^{V'})\Big] + i\hat{T}(V + V')_{xx} + \frac{i}{\delta}(e^V + e^{V'})_x - \tfrac{1}{2}i\big[\hat{T}\{\hat{T}(V - V')_x + (1/\delta)(e^V - e^{V'})\}\big]\big[\hat{T}(V - V')_x + (1/\delta)(e^V - e^{V'})\big] + \tfrac{1}{2}i(V - V')_x\hat{T}(V - V')_x = 0, \tag{3.10b}$$

$$-\hat{T}(V - V')_x - i(V + V')_x = 2\mu\exp\frac{i}{2}\Big[(\hat{T}(V - V') + \frac{1}{\delta}\int^x(e^V - e^{V'})dx\Big]\sinh\frac{1}{2}(V - V'), \tag{3.10c}$$

$$-i\hat{T}\big[\hat{T}(V - V')_x + (1/\delta)(e^V - e^{V'})\big] + \hat{T}(V + V')_x + (1/\delta)(e^V + e^{V'} - 2) = -2c + 2\mu\exp\frac{i}{2}\Big[\hat{T}(V - V') + \frac{1}{\delta}\int^x(e^V - e^{V'})dx\Big]\cosh\frac{1}{2}(V - V'). \tag{3.10d}$$

We note that either set of equations (3.10a), (3.10c) or (3.10b), (3.10d) is adequate to represent the BT of the MILW equation.

IV. THE CONSERVED QUANTITIES OF THE MILW EQUATION

One may derive the conserved quantities from the BT (see, for example, Refs. 7 and 18). Introduce $W = V - V'$ and rewrite Eq. (3.10) as

$$\tfrac{1}{3} i\delta W_t + i(1/\delta - c)W_x + (2V - W)_{xx} - \tfrac{1}{2}\hat{T}(W_x)I_x - \tfrac{1}{2}W_x\hat{T}(I_x) = 0, \tag{4.1a}$$

$$\tfrac{1}{3} i\delta I_t - (1/\delta - c)I_x - i\{I_x - 2\hat{T}(V_x) - (2/\delta)e^V\}_x - \tfrac{1}{2}iI_x\hat{T}(I_x) + \tfrac{1}{2}iW_x\hat{T}(W_x) = 0, \tag{4.1b}$$

$$-\hat{T}(W_x) + i(W-2V)_x = 2\mu(\exp \tfrac{1}{2} iI)\sinh \tfrac{1}{2} W, \tag{4.1c}$$

$$-i\hat{T}[I_x] + \hat{T}(2V-W)_x + (1/\delta)(e^V + e^{V-W} - 2) = -2c + 2\mu(\exp \tfrac{1}{2} iI)\cosh \tfrac{1}{2} W,$$

where

$$I = \hat{T}(W) + \frac{1}{\delta}\int_{-\infty}^{x} e^V(1-e^{-W})dx. \tag{4.2}$$

Imposing the boundary conditions $V(\pm\infty) = 0$, and using

$$\int_{-\infty}^{\infty} [u\hat{T}(v) + v\hat{T}(u)]dx = 0, \tag{4.3}$$

we have, from Eq. (4.1a), that

$$\left[\int_{-\infty}^{\infty} W\,dx\right]_t = 0, \tag{4.4}$$

and, from Eq. (4.1b), that

$$\left[\int_{-\infty}^{\infty} I\,dx\right]_t = 0, \tag{4.5}$$

which means that $\int_{-\infty}^{\infty} W\,dx$ and $\int_{-\infty}^{\infty} I\,dx$ are conserved quantities.

Following Miura,[19] for $\mu \to \infty$, if we expand

$$W = \sum_{n=1}^{\infty} \epsilon^n W_n, \tag{4.6}$$

where

$$\epsilon = 1/2\mu \ll 1, \tag{4.7}$$

and equating the same power of ϵ in Eq. (4.1c), we obtain

$$W_1 = -4iV_x, \tag{4.8a}$$

$$W_2 = 8i[(\hat{T}-i)V_{xx} + V_x\hat{T}(V_x) + (1/\delta)V_x(e^V-1)], \tag{4.8b}$$

$$\begin{aligned} W_3 = {} & -2(\hat{T}-i)W_{2x} - \frac{1}{24}W_1^3 - W_1 \\ & \times\left(\frac{i}{2}\hat{T}(W_2) + \frac{i}{2\delta}\int_{-\infty}^{x} e^V(W_2 - \tfrac{1}{2}W_1^2)dx \right. \\ & \left. - \frac{1}{8}(\hat{T}(W_1) + \int_{-\infty}^{x}\frac{1}{\delta}e^V W_1\,dx)^2\right) \\ & - \frac{i}{2}W_2\left(\hat{T}(W_1) + \frac{1}{\delta}\int_{-\infty}^{x} e^V W_1\,dx\right). \end{aligned} \tag{4.8c}$$

Unfortunately, the integrals $\int_{-\infty}^{\infty} W_1\,dx$, $\int_{-\infty}^{\infty} W_2\,dx$, $\int_{-\infty}^{\infty} W_3\,dx$, etc., are trivial conserved quantities. But from the expansion of I,

$$I = \hat{T}(W) + \frac{1}{\delta}\int_{-\infty}^{x} e^V(1-e^{-W})dx = \sum_{n=1}^{\infty}\epsilon^n I_n, \tag{4.9}$$

we obtain the following [using (4.6)]:

$$\begin{aligned} I_1 = {} & -4i\left(\hat{T}(V_x) + \frac{1}{\delta}\int_{-\infty}^{x} e^V V_x\,dx\right), \\ I_2 = {} & 8i\hat{T}[(\hat{T}-i)V_{xx} + V_x\hat{T}(V_x) + (1/\delta)V_x(e^V-1)] \\ & + \frac{1}{\delta}\int_{-\infty}^{x} e^V[8i\{(\hat{T}-i)V_{xx} + V_x\hat{T}(V_x) \\ & + (1/\delta)V_x(e^V-1)\} + 8V_x^2]\,dx, \end{aligned} \tag{4.10}$$

from which are deduced the following conserved quantities:

$$(1)\ (1/\delta)(e^V-1), \tag{4.11a}$$

$$(2)\ \hat{T}(V_x\hat{T}(V_x)) + (1/\delta)e^V\hat{T}(V_x) + (1/2\delta^2)(e^{2V}-1). \tag{4.11b}$$

These may be verified directly.

In the limit as $\delta \to 0$, Eqs. (4.11a) and (4.11b), using (2.5), give the first and the second conserved quantities of the MKdV equation (at the lowest nontrivial order of δ), which are u and u^2, respectively.[20]

V. THE IST PROBLEM FOR THE MILW EQUATION

In a manner analogous to Wadati[20] and Satsuma *et al.*,[7] we obtain the IST problem of Eq. (2.4) from the BT given in Eq. (4.1). By using the following dependent variable transformation,

$$(\log \psi_1^\pm)_x = (\log(f'/f))_x = \tfrac{1}{4}(i\hat{T} \pm 1)\{W_x + iI_x\}, \tag{5.1a}$$

and

$$(\log \psi_2^\pm)_x = (\log(g'/g))_x = -\tfrac{1}{4}(i\hat{T} \pm 1)\{-W_x + iI_x\} \tag{5.1b}$$

(see Appendix C for useful relations which can be derived from the above equation), and by substituting (5.1) and Appendix C relations into Eq. (4.1), adding and substracting the resulting equations yield

$$-2i\psi_{1x}^+ + [c - iV_x + \hat{T}(V_x) + (1/\delta)(e^V-1)]\psi_1^+ = \mu\psi_1^-, \tag{5.2a}$$

$$-2i\psi_{2x}^+ + [c + iV_x + \hat{T}(V_x) + (1/\delta)(e^V-1)]\psi_2^+ = \mu\psi_2^-, \tag{5.2b}$$

$$-\tfrac{1}{3}i\delta\psi_{1t}^\pm - i((1/\delta)-c)\psi_{1x}^\pm + [\tfrac{1}{2}(i\hat{T}\pm 1)\{V_{xx} + i\hat{T}(V_{xx}) + (i/\delta)(e^V)_x\} - \tfrac{1}{4}k^2]\psi_1^\pm + \psi_{1xx}^\pm = 0, \tag{5.2c}$$

$$-\tfrac{1}{3}i\delta\psi_{2t}^\pm - i((1/\delta)-c)\psi_{2x}^\pm + [\tfrac{1}{2}(i\hat{T}\pm 1)\{-V_{xx} + i\hat{T}(V_{xx}) + (i/\delta)(e^V)_x\} - \tfrac{1}{4}k^2]\psi_2^\pm + \psi_{2xx}^\pm = 0. \tag{5.2d}$$

We expect that $\psi_i^\pm$ have the following analytical meaning: $\psi^\pm$ represent the boundary values of functions [i.e., $\psi^\pm(x) = \lim_{\operatorname{Im} z\to 0}\psi^\pm(z)$] analytic in the horizontal strips between $\operatorname{Im} z = 0$ and $\operatorname{Im} z = \mp 2\delta$, and periodically extended thereafter.

By defining

$$\psi_1 = \tfrac{1}{2}(v_1 - iv_2), \tag{5.3a}$$

$$\psi_2 = \tfrac{1}{2}(v_1 + iv_2), \tag{5.3b}$$

and substituting into Eqs. (5.2), we get the following set of equations (by adding and substracting):

$$v_{1x}^+ + \bar{W}v_1^+ + Uv_2^+ - \tfrac{1}{2}i\mu v_1^- = 0, \tag{5.4a}$$

$$v_{2x}^+ - Uv_1^+ + \bar{W}v_2^+ - \tfrac{1}{2}i\mu v_2^- = 0, \tag{5.4b}$$

$$\tfrac{1}{3}i\delta v_{1t}^\pm + i(1/\delta - c)v_{1x}^\pm - v_{1xx}^\pm - A^\pm v_1^\pm - B^\pm v_2^\pm = 0, \tag{5.4c}$$

$$\tfrac{1}{3} i\delta v_{2t}^{\pm} + i(1/\delta - c)v_{2x}^{\pm} - v_{2xx}^{\pm} + B^{\pm} v_1^{\pm} - A^{\pm} v_2^{\pm} = 0, \tag{5.4d}$$

where

$$U = \tfrac{1}{2} i V_x, \tag{5.5}$$

$$\overline{W} = \tfrac{1}{2} i\{\widehat{T}(V_x) + (1/\delta)(e^V - 1) + c\}, \tag{5.6}$$

$$A^{\pm} = \tfrac{1}{2}(i\widehat{T} \pm 1)\{i\widehat{T}(V_x) + (i/\delta)(e^V - 1)\}_x - \tfrac{1}{4}k^2, \tag{5.7}$$

$$B^{\pm} = \tfrac{1}{2} i(i\widehat{T} \pm 1)V_{xx}, \tag{5.8}$$

$$c = k \cot k\delta, \tag{5.9}$$

$$\mu = k \csc k\delta. \tag{5.10}$$

It is readily shown by a cross differentiation that Eqs. (5.4a)–(5.4d) constitute an IST problem associated with Eq. (2.4). Assuming $V = O(\delta)$, expanding $v_i^{\pm}$ $(i = 1,2)$ in terms of δ and substituting them into Eqs. (5.4), we obtain, at the lowest nontrivial order of δ,

$$v_{1xx} + (u^2 - \tfrac{1}{4}k^2)v_1 + v_2 u_x = 0, \tag{5.11a}$$

$$v_{2xx} + (u^2 - \tfrac{1}{4}k^2)v_2 - v_1 u_x = 0, \tag{5.11b}$$

$$v_{1t} + 2v_{2x} u_x - v_2 u_{xx} - v_1(u^2)_x + v_{1x}\{k^2 + 2u^2\} = 0, \tag{5.11c}$$

$$v_{2t} - 2v_1 u_x + v_1 u_{xx} - v_2(u^2)_x + v_{2x}\{k^2 + 2u^2\} = 0, \tag{5.11d}$$

which is one of the known IST problems associated with the MKdV Eq. (2.6). (See Ref. 20 and also 21.) It is to be noted that Eq. (2.4) is a necessary compatibility condition for Eqs. (5.2a) and (5.2c). (See Appendix D for still another form of IST problem for the MILW equation.) Hence we now have a natural (and novel) extension of the inverse scattering transform for the modified KdV equation (5.11) (Refs. 20, 21) to the modified intermediate long wave equation (5.2).

APPENDIX A: PROPERTIES OF THE D OPERATOR

We have

$$D_t^n D_x^m a \cdot b = \left(\frac{\partial}{\partial t} - \frac{\partial}{\partial t'}\right)^n \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right)^m \times a(t,x)\, b(t',x')|_{t=t',\, x=x'}.$$

The following properties are easily seen from the definition.

(1) $D_x^m a \cdot l = \left(\dfrac{\partial}{\partial x}\right)^m a,$

(2) $D_x^m a \cdot b = (-1)^m D_x^m b \cdot a,$

(3) $D_x^m a \cdot a = 0$ for odd m,

(4) $D_x^m a \cdot b = D_x^{m-1}(a_x \cdot b - a \cdot b_x),$

(5) $D_x^m a \cdot a = 2D_x^{m-1} a_x \cdot a$ for even m,

(6) $D_x\, d_t\, a \cdot a = 2D_x\, a_t \cdot a = 2D_t\, a_x \cdot a,$

(7) $D_x^m \exp(k_1 x) \cdot \exp(k_2 x) = (k_1 - k_2)^m \exp(k_1 + k_2)x.$

Let $\phi = \log(a/b)$ and $\rho = \log ab$; we have

(8) $(D_x\, a \cdot b)/ab = \phi_x,$

(9) $(D_x^2\, a \cdot b)/ab = \rho_{xx} + (\phi_x)^2,$

(10) $(D_x^3\, a \cdot b)/ab = \phi_{xxx} + 3\phi_x \rho_{xx} + (\phi_x)^3.$

APPENDIX B: ON THE ILW EQUATION

The ILW equation is given by

$$u_t + 2uu_x + T(u_{xx}) = 0. \tag{B1}$$

Introducing the following dependent variable transform

$$u = i(\log f^-/f^+)_x, \tag{B2}$$

Eq. (B1) can be expressed in bilinear form as

$$(iD_t + (i/\delta)D_x + D_x^2)f^+ \cdot f^- = 0 \tag{B3}$$

(see Appendix A for some properties of these operators). The BT of Eq. (B3) in the bilinear form is given by[8]

$$(iD_t + i(1/\delta + \lambda)D_x + D_x^2 - \tfrac{1}{4}k^2)f^{\pm} \cdot g^{\pm} = 0, \tag{B3a}$$

$$(2iD_x - \lambda)f^- \cdot g^+ = \nu f^+ \cdot g^-, \tag{B3b}$$

where

$$\lambda = -k \cot k\delta, \tag{B3c}$$

$$\nu = k \csc k\delta. \tag{B3d}$$

APPENDIX C: USEFUL RELATIONS DERIVED FROM EQUATION (5.1)

(1) $(\log(\psi_1^+/\psi_1^-))_x = -\tfrac{1}{2}\{W_x + iI_x\},$

(2) $(\log(\psi_2^+/\psi_2^-))_x = -\tfrac{1}{2}[-W_x + iI_x],$

(3) $(\log(\psi_2^+/\psi_1^+))_x = \tfrac{1}{2}(i\widehat{T} + 1)W_x,$

(4) $(\log(\psi_2^-/\psi_1^-))_x = \tfrac{1}{2}(i\widehat{T} - 1)W_x,$

(5) $(\log(\psi_2^+\psi_1^-/\psi_1^+\psi_2^-))_x = W_x,$

(6) $(\log(\psi_2^+\psi_2^-/\psi_1^+\psi_1^-))_x = i\widehat{T}W_x.$

APPENDIX D: ANOTHER IST PROBLEM FOR THE MILW EQUATION

In this Appendix, we will give an alternative IST problem for the MILW equation (2.4). It is to be noted that we need only equations for one of the ψ_i, say ψ_1, to obtain this.

Define

$$Q^{\pm} = (\widehat{T} \pm i)V + \int_{-\infty}^{x} \frac{1}{\delta}(e^V - 1)dx. \tag{D1}$$

Rewrite Eqs. (5.1a) and (5.1b) as

$$(\log \psi_1^{\pm}) = -\tfrac{1}{4}(\widehat{T} \mp i)(Q'^- - Q^-), \tag{D2a}$$

$$(\log \psi_2^{\pm}) = -\tfrac{1}{4}(\widehat{T} \mp i)(Q'^+ - Q^+), \tag{D2b}$$

respectively. For convenience, we denote $\psi_1^{\pm}$ by $\xi_1^{\mp}$ and $\psi_2^{\pm}$ by $\xi_2^{\mp}$. Equations (5.2a) and (5.2c) can be rewritten as

$$-2i\xi_{1x}^- + (Q_x^- + c)\xi_1^- = -\mu\xi_1^+, \tag{D3a}$$

$$\tfrac{1}{3} i\delta\xi_{1t}^{\mp} + i(1/\delta - c)\xi_{1x}^{\mp} + \{\tfrac{1}{2}(\widehat{T} \mp i)Q_{xx}^- + \tfrac{1}{4}k^2\}\xi_1^{\mp} - \xi_{1xx}^{\mp} = 0, \tag{D3b}$$

respectively, and from the analytical property, note that

$$(Q^{\pm})^* = Q^{\mp}, \quad (\xi_1^{\mp})^* = \xi_2^{\pm}. \tag{D4}$$

In analogy to the MKdV case, if we take Eqs. (D3) together with their complex conjugates, this will yield the IST problem of the MILW equation. Indeed, the following set of equations for ξ_1 and ξ_2, which are

$$-2i\xi_{1x}^{-}+(Q_x^{-}+c)\xi_1^{-}=-\mu\xi_1^{+}, \tag{D5a}$$

$$2i\xi_{2x}^{+}+(Q_x^{+}+c)\xi_2^{+}=-\mu\xi_2^{-}, \tag{D5b}$$

$$\tfrac{1}{3}i\delta\xi_{1t}^{\mp}+i(1/\delta-c)\xi_{1x}^{\mp}+\{\tfrac{1}{2}(\hat{T}\mp i)Q_{xx}^{-}+\tfrac{1}{4}k^2\}\xi_1^{\mp}-\xi_{1xx}^{\mp}=0, \tag{D5c}$$

$$-\tfrac{1}{3}i\delta\xi_{2t}^{\pm}-i(1/\delta-c)\xi_{2x}^{\pm}+\{\tfrac{1}{2}(\hat{T}\pm i)Q_{xx}^{+}+\tfrac{1}{4}k^2\}\xi_2^{\pm}-\xi_{2xx}^{\pm}=0, \tag{D5d}$$

yields an alternative IST problem for Eq. (2.4). By defining

$$\begin{aligned}\xi_1^{-}&=-\tfrac{1}{2}(1+i\hat{T})A,\\ \xi_2^{+}&=\tfrac{1}{2}(1-i\hat{T})A,\\ \xi_1^{+}&=\tfrac{1}{2}(1-i\hat{T})\phi,\\ \xi_2^{-}&=-\tfrac{1}{2}(1+i\hat{T})\phi,\end{aligned} \tag{D6}$$

Eqs. (D5a)–(D5d) give the following set of equations (by adding and subtracting):

$$2iA_x-\overline{W}(i\hat{T}A)+iV_xA=-i\mu\hat{T}\phi, \tag{D7a}$$

$$-2\hat{T}A_x-\overline{W}A+iV_x\hat{T}A=\mu\phi, \tag{D7b}$$

$$\tfrac{1}{3}i\delta A_t+i(1/\delta-c)A_x-U^{+}A-G^{+}(i\hat{T}A)-i\hat{T}A_{xx}=0, \tag{D7c}$$

$$-\tfrac{1}{3}\delta\hat{T}A_t-(1/\delta-c)\hat{T}A_x-U^{+}(i\hat{T}A)-G^{+}A-A_{xx}=0, \tag{D7d}$$

$$-\tfrac{1}{3}i\delta\phi_t-i(1/\delta-c)\phi_x+U^{-}\phi-G^{-}(i\hat{T}\phi)-i\hat{T}\phi_{xx}=0, \tag{D7e}$$

$$-\tfrac{1}{3}\delta T\phi_t-(1/\delta-c)\hat{T}\phi_x-U^{-}(i\hat{T}\phi)+G^{-}\phi+\phi_{xx}=0, \tag{D7f}$$

where

$$\overline{W}=c-i[i\hat{T}(V_x)+(i/\delta)(e^V-1)], \tag{D8a}$$

$$U^{\pm}=\tfrac{1}{2}i\hat{T}(V_{xx})\pm\tfrac{1}{2}[i\hat{T}V_{xx}+(i/\delta)(e^V-1)], \tag{D8b}$$

$$G^{\pm}=\tfrac{1}{2}i\hat{T}[i\hat{T}(V_{xx})+(i/\delta)(e^V)_x]\pm\tfrac{1}{2}V_{xx}-\tfrac{1}{4}k^2. \tag{D8c}$$

ACKNOWLEDGMENTS

This research was partially supported by the Office of Naval Research, under Grant No. 00014-76-C-0867, the Air Force Office of Scientific Research, under Grant No. 78-3674-C, and the National Science Foundation, under Grant No. MCX79-03498-01. We are grateful for useful conversations with Professor A. S. Fokas and Dr. A. Nakamura.

[1] M. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[2] C. Gardner, J. Greene, M. Kruskal, and R. Miura, Phys. Rev. Lett. **19**, 1095 (1967).
[3] V. Zakharov and A. Shabat, Zh. Eksp. Teor. Fiz. **61**, 118 (1971) [Sov. Phys. JETP **34**, 62 (1972)].
[4] M. Ablowitz, D. Kaup, A. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[5] G. L. Lamb, *Elements of Soliton Theory* (Wiley, New York, 1980).
[6] R. Hirota, "Direct Method of Finding Exact Solutions of Nonlinear Evolution Equations, Bäcklund Transformations," in *Lecture Notes in Mathematics*, Vol. 515, edited by R. M. Miura (Springer-Verlag, Berlin, 1974), pp. 40–68.
[7] J. Satsuma, M. J. Ablowitz, and Y. Kodama, "On an Internal Wave Equation Describing a Stratified Fluid With Finite Depth," Phys. Lett. **73A**, 283–286 (1979).
[8] J. Satsuma and M. Ablowitz, "Solutions of an Internal Wave Equation Describing a Stratified Fluid With Finite Depth," in *Nonlinear Partial Differential Equations in Engineering and Applied Science*, edited by R. L. Sternberg, A. J. Kalinowki, and J. S. Papadakis (Marcel Dekker, New York, 1980), p. 397.
[9] Y. Kodama, M. J. Ablowitz, and J. Satsuma, "Direct and Inverse Scattering Problems of the Nonlinear Intermediate Long Wave Equation," J. Math. Phys. **23**, 564 (1982).
[10] M. J. Ablowitz, A. S. Fokas, and R. L. Anderson, "The Direct Linearizing Transform and the Benjamin-Ono Equation" (preprint); A. S. Fokas and M. J. Ablowitz, "The Inverse Scattering Transform for the Benjamin–Ono Equation—A Pivot To Multidimensional Problems" (preprint).
[11] R. Hirota and J. Satsuma, "A Variety of Nonlinear Network Equations Generated from the Bäcklund Transformations for the Toda Lattice," Prog. Theor. Phys. Suppl. **59**, 64–100 (1976).
[12] A. Nakamura, J. Phys. Soc. Jpn. **47**, 2043–2046 (1979).
[13] H. D. Wahlquist and F. B. Estabrook, Phys. Rev. Lett. **31**, 1386 (1973).
[14] M. Wadati, J. Phys. Soc. Jpn **36**, 1498 (1974).
[15] G. L. Lamb, Jr., J. Math. Phys. **15**, 2157–2165 (1974).
[16] T. Taha, Ph.D. thesis (Clarkson College, Potsdam, NY, 1982).
[17] R. Hirota and J. Satsuma, J. Phys. Soc. Jpn. **45**, 1741–1750 (1978).
[18] J. Satsuma, Prog. Theor. Phys. **52**, 1396 (1974); **53**, 585 (1975).
[19] R. M. Miura, J. Math. Phys. **9**, 1202 (1968); R. M. Miura, C. S. Gardner, and M. D. Kruskal, J. Math. Phys. **9**, 1204 (1968).
[20] M. Wadati, H. Sanki, and K. Konno, Prog. Theor. Phys. **53**, 419–436 (1975).
[21] F. Calogero and A. Degasperis, J. Math. Phys. **22**, 23–31 (1981).

Propagators from integral representations of Green's functions for the N-dimensional free-particle, harmonic oscillator and Coulomb problems

S. M. Blinder
Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109

(Received 16 June 1983; accepted for publication 26 August 1983)

The radial Green's functions for the N-dimensional free-particle, isotropic harmonic oscillator and Coulomb problems all contain a product of two Bessel or Whittaker functions. After integral representations for these respective products are introduced, each Green's function exhibits the structure of a Fourier transform. One obtains thereby the Feynman propagators $K(r_1,r_2,t)$ for the free particle and harmonic oscillator. In the Coulomb case, the Fourier transform involves the quantum number variable and leads instead to the recently defined Sturmian propagator. The well-known connection between Coulomb and oscillator eigenstates of various dimensionality is manifested in a new way by the structure of the propagators derived here.

PACS numbers: 03.65. — w

1. GREEN'S FUNCTIONS AND PROPAGATORS

Consider a particle moving in an N-dimensional Euclidian space in a "central" potential $V(r)$, where r represents the N-dimensional radius $[\Sigma_i x_i^2]^{1/2}$. The L th "partial-wave" Green's function satisfies the radial equation[1]

$$\left[E+\frac{1}{2r^{N-1}}\frac{\partial}{\partial r}r^{N-1}\frac{\partial}{\partial r}-\frac{L(L+N-2)}{2r^2}-V(r)\right]\times G_L^{(N)}(r,r',E)=\delta(r-r')/(rr')^{N/2-1/2} \tag{1.1}$$

$$(L=0,1,2,\cdots).$$

Atomic units will be used throughout, with $\hbar=m=e=1$. We will consider, in turn, the free particle, isotropic harmonic oscillator and Coulomb system, corresponding to $V(r)=0$, $\frac{1}{2}\omega^2r^2$, and $-Z/r$, respectively. Equation (1.1), rearranged to standard Sturm–Liouville form, becomes

$$\left[\frac{\partial}{\partial r}\left(\frac{1}{2}r^{N-1}\right)\frac{\partial}{\partial r}+r^{N-1}\left(E+\frac{L(L+N-2)}{2r^2}-V(r)\right)\right]\times G_L^{(N)}(r,r',E)=\delta(r-r'). \tag{1.2}$$

The usual procedure for constructing Green's functions then gives[2]

$$G(r,r',E)=u(r_<)v(r_>)/\tfrac{1}{2}r^{N-1}W[u,v]. \tag{1.3}$$

Here, $u(r)$ and $v(r)$ are solutions of the homogeneous equation (1.2), when $r\neq r'$, appropriate to the boundary conditions at $r=0$ and $r=\infty$, respectively, while $W[u,v]$ is the Wronskian

$$W[u,v]\equiv u(r)v'(r)-v(r)u'(r). \tag{1.4}$$

The particular solution $G^+(r,r',E)$ behaves like an outgoing spherical wave as $r_>\to\infty$. It is associated with the contour along the E axis such that $\operatorname{Im}E>0$. This Green's function is related by a Fourier transform to the Feynman propagator $K(r,r',t)$, as follows[3]:

$$G^+(r,r',E)=-i\int_0^\infty K(r,r',t)e^{iEt}\,dt. \tag{1.5}$$

This can be shown, most readily, from the respective spectral representations

$$G^+(r,r',E)=\sum_n\frac{R_n(r)R_n^*(r')}{E+i\epsilon-E_n} \tag{1.6}$$

and

$$K(r,r',E)=\sum_n R_n(r)R_n^*(r')e^{-iE_nt}. \tag{1.7}$$

2. N-DIMENSIONAL FREE PARTICLE

Applying the procedure outlined above to the free particle, we write Eq. (1.2) with $V(r)=0$ and $E=k^2/2$. The solution of the homogeneous equation analytic at $r=0$ is readily shown to be

$$u(r)=r^{1-N/2}J_{L+N/2-1}(kr). \tag{2.1}$$

For the outgoing wave Green's function G^+, the appropriate form of the outer solution is

$$v(r)=r^{1-N/2}H^{(1)}_{L+N/2-1}(kr). \tag{2.2}$$

The Wronskian of (2.1) and (2.2) is given by[4]

$$W[u,v]=r^{2-N}W\left[J_{L+N/2-1}(kr),H^{(1)}_{L+N/2-1}(kr)\right]=(2i/\pi)r^{1-N}. \tag{2.3}$$

Thus, by (1.3)

$$G_L^{(N)+}(r_1,r_2,k)=-i\pi(r_1r_2)^{1-N/2}J_{L+N/2-1}(kr_<)H^{(1)}_{L+N/2-1}(kr_>). \tag{2.4}$$

For odd dimension ($N=3,5,\ldots$), one can introduce the corresponding spherical Bessel functions, to give

$$G_L^{(N)+}(r_1,r_2,k)=-2ik(r_1r_2)^{3/2-N/2}j_{L+N/2-3/2}(kr_<)\times h^{(1)}_{L+N/2-3/2}(kr_>). \tag{2.5}$$

We now make use of an integral representation for a product of two Bessel functions[5]

$$J_\nu(z)H_\nu^{(1)}(Z)=\frac{1}{i\pi}\int_0^{c+i\infty}\exp\left[\frac{1}{2}t-\frac{Z^2+z^2}{2t}\right]I_\nu\left(\frac{zZ}{t}\right)\frac{dt}{t}. \tag{2.6}$$

 0022-2488/84/040905-05$02.50

With the substitutions $t \to ik^2 t$, $c=0$, $\nu = L + N/2 - 1$, $z = kr_1$, $Z = kr_2$, we obtain

$$J_{L+N/2-1}(kr_<)H^{(1)}_{L+N/2-1}(kr_>) = \frac{(-i)^{L+N/2}}{\pi}\int_0^\infty e^{ik^2t/2}e^{i(r_1^2+r_2^2)/2t} \times J_{L+N/2-1}\left(\frac{r_1 r_2}{t}\right)\frac{dt}{t}. \tag{2.7}$$

Therefore

$$G_L^{(N)+}(r_1,r_2,k) = (-i)^{L+N/2+1}(r_1r_2)^{1-N/2} \times \int_0^\infty e^{ik^2t/2}e^{i(r_1^2+r_2^2)/2t}J_{L+N/2-1}\left(\frac{r_1r_2}{t}\right)\frac{dt}{t}. \tag{2.8}$$

We note that the energy parameter $E = k^2/2$ is now isolated in the first exponential factor of the integrand. Thus (2.8) can be identified with the Fourier transform (1.5), which immediately gives the free-partial propagator

$$K_L^{(N)}(r_1,r_2,t) = (-i)^{L+N/2}(r_1r_2)^{1-N/2}t^{-1} \times e^{i(r_1^2+r_2^2)/2t}J_{L+N/2-1}(r_1r_2/t). \tag{2.9}$$

From the derivative formula[6] $J_{\nu+1}(z) = -z^\nu(\partial/\partial z)[z^{-\nu}J_\nu(z)]$, the following recursive relation for N with $L=0$ can be demonstrated

$$K_0^{(N+2)} = \frac{it}{\eta}\left(\frac{\partial K_0^{(N)}}{\partial \eta}\right)_\xi,$$
$$\eta \equiv r_1 r_2, \quad \xi \equiv (r_1^2 + r_2^2)/2. \tag{2.10}$$

It is readily verified that (2.9) is a solution of the partial differential equation

$$i\frac{\partial K}{\partial t} + \frac{1}{2r^{N-1}}\frac{\partial}{\partial r}r^{N-1}\frac{\partial K}{\partial r} - \frac{L(L+N-2)}{2r^2}K = 0, \tag{2.11}$$

with the initial condition

$$K_L^{(N)}(r_1,r_2,0) = \delta(r_1 - r_2)/(r_1r_2)^{N/2-1/2}. \tag{2.12}$$

If the Bessel function in (2.9) is expressed in terms of a confluent hypergeometric function,[7] viz.,

$$J_\nu(z) = \frac{(z/2)^\nu e^{-iz}}{\Gamma(\nu+1)}M\left(\nu + \frac{1}{2};2\nu+1;2iz\right), \tag{2.13}$$

then the propagator (2.9) exhibits the structure[8]

$$K = F e^{iS}, \tag{2.14}$$

with S representing the one-dimesional free-particle action

$$S(r_1,r_2,t) = (r_1 - r_2)^2/2t. \tag{2.15}$$

We cite, in particular, the two- and three-dimensional cases. For $N=2$, with the customary notation $L = m$ and $r = \rho$:

$$K_m^{(2)}(\rho_1,\rho_2,t) = (-i)^{m+1}t^{-1}e^{i(\rho_1^2+\rho_2^2)/2t}J_m(\rho_1\rho_2/t) \quad (m = 0, \pm 1, \pm 2,\cdots). \tag{2.16}$$

Summation over m gives

$$K^{(2)}(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,t) = \frac{1}{2\pi}\sum_{m=-\infty}^{\infty}K_m^{(2)}(\rho_1,\rho_2,t)e^{im(\phi_1-\phi_2)}. \tag{2.17}$$

Making use of the generating function

$$e^{z(u-1/u)/2} = \sum_{m=-\infty}^{\infty}u^m J_m(z), \tag{2.18}$$

with $u = -ie^{i(\phi_1-\phi_2)}$, $z = \rho_1\rho_2/t$, we obtain

$$K^{(2)}(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,t) = (2\pi it)^{-1}e^{i(\rho_1^2+\rho_2^2)/2t}e^{-i(\rho_1\rho_2/t)\cos(\phi_1-\phi_2)} = (2\pi it)^{-1}e^{i(\boldsymbol{\rho}_1-\boldsymbol{\rho}_2)^2/2t}. \tag{2.19}$$

For $N=3$ with $L = l$:

$$K_l^{(3)}(r_1,r_2,t) = (-1)^{l+3/2}(2/\pi t^3)^{1/2} \times e^{i(r_1^2+r_2^2)/2t}j_l(r_1r_2/t). \tag{2.20}$$

The sum over partial waves gives

$$K^{(3)}(\mathbf{r}_1,\mathbf{r}_2,t) = \sum_{l=0}^\infty \frac{2l+1}{4\pi}P_l(\cos\theta)K_l^{(3)}(r_1,r_2,t). \tag{2.21}$$

With use of the addition theorem[9]

$$e^{iz\cos\theta} = \sum_{l=0}^\infty (2l+1)i^l j_l(z)P_l(\cos\theta), \tag{2.22}$$

we obtain the familiar three-dimensional free-particle propagator[10]

$$K^{(3)}(\mathbf{r}_1,\mathbf{r}_2,t) = (2\pi it)^{-3/2}e^{i(r_1^2+r_2^2)/2t}e^{-ir_1r_2\cos\theta/t} = (2\pi it)^{-3/2}e^{i(\mathbf{r}_1-\mathbf{r}_2)^2/2t}. \tag{2.23}$$

3. N-DIMENSIONAL HARMONIC OSCILLATOR

For the N-dimensional isotropic harmonic oscillator, Eq. (1.1) takes the form

$$\left[E + \frac{1}{2}\left(\frac{1}{r^{N-1}}\frac{\partial}{\partial r}r^{N-1}\frac{\partial}{\partial r} - \frac{L(L+N-2)}{r^2} - \omega^2r^2\right)\right] \times G_L^{(N)}(r,r',E) = \delta(r-r')/(rr')^{N/2-1/2}. \tag{3.1}$$

All results in this section reduce to the corresponding free-particle formulas in the limit $\omega \to 0$. The following solutions of the homogeneous equation can be demonstrated[11]

$$u(r) = r^{-N/2}M_{E/2\omega}^{(L+N/2-1)/2}(\omega r^2) \tag{3.2}$$

and

$$v(r) = r^{-N/2}W_{E/2\omega}^{(L+N/2-1)/2}(\omega r^2), \tag{3.3}$$

where M and W are Whittaker functions as defined by Buchholz. However, for compactness of notation, we write $M_\kappa^{\mu/2}(z)$ in place of $\mathscr{M}_{\kappa,\mu/2}(z)$ and $W_\kappa^{\mu/2}(z)$ in place of $W_{\kappa,\mu/2}(z)$.[12] Using the Wronskian[13]

$$W[M_\kappa^{\mu/2}(z),W_\kappa^{\mu/2}(z)] = \frac{-1}{\Gamma[(\mu+1)/2-\kappa]}, \tag{3.4}$$

we obtain

$$W[u,v] = \frac{-2\omega r^{1-N}}{\Gamma(L/2+N/4-E/2\omega)} \tag{3.5}$$

and thereby, by (1.3),

$$G_L^{(N)}(r_1,r_2,E) = -\omega^{-1}(r_1r_2)^{-N/2}\Gamma(L/2+N/4-E/2\omega) \times M_{E/2\omega}^{(L+N/2-1)/2}(\omega r_<^2)W_{E/2\omega}^{(L+N/2-1)/2}(\omega r_>^2). \tag{3.6}$$

Note that this Green's function is nonpropagating, as is indeed expected for a purely discrete spectrum. The eigenvalues for this system follow simply from the poles of the gamma function, viz.,

$$E_{n,L}^{(N)} = (2n + L + N/2)\omega \quad (n = 0,1,2,\cdots). \tag{3.7}$$

We next make use of an integral representation for a product of two Whittaker functions given by Buchholz[14]:

$$\Gamma\left(\frac{\mu+1}{2} - \kappa\right) W_\kappa^{\mu/2}(a_1 t) M_\kappa^{\mu/2}(a_2 t) = t\sqrt{a_1 a_2}\int_0^\infty \exp\, -\tfrac{1}{2}(a_1 + a_2)t \cosh v \times I_\mu(t\sqrt{a_1 a_2}\sinh v)\coth^{2\kappa}(v/2)dv \tag{3.8}$$

restricted, however, by the condition that

$$\mathrm{Re}\left(\frac{\mu+1}{2} - \kappa\right) > 0. \tag{3.9}$$

In order to make this representation applicable, we temporarily turn ω into a pure imaginary

$$\omega = -i\sigma. \tag{3.10}$$

With the substitutions in (3.8): $\mu = L + N/2 - 1$, $t = -i\sigma$, $\kappa = iE/2\sigma$, $a_1 = r_>^2$, $a_2 = r_<^2$ and the variable transformation $\sinh v = \operatorname{csch} \sigma t$, $\cosh v = \coth \sigma t$, $\coth(v/2) = e^{\sigma t}$, $dv = -\sigma \operatorname{csch} \sigma t\, dt$, we obtain

$$G_L^{(N)}(r_1,r_2,E)|_{\omega=-i\sigma} = -(-i)^{L+N/2-1}\sigma(r_1 r_2)^{1-N/2} \times \int_0^\infty dt\, e^{iEt} \operatorname{csch} \sigma t \times e^{(1/2)i\sigma(r_1^2 + r_2^2)\coth \sigma t} \times J_{L+N/2-1}(\sigma r_1 r_2 \operatorname{csch} \sigma t). \tag{3.11}$$

Again, this can be identified with the Fourier transform (1.5). After reverting back to real $\omega(\sigma = i\omega)$, we obtain the harmonic-oscillator propagators

$$K_L^{(N)}(r_1,r_2,t) = (-i)^{L+N/2}\omega(r_1 r_2)^{1-N/2}\csc \omega t \times e^{(1/2)i\omega(r_1^2 + r_2^2)\cot \omega t} J_{L+N/2-1}(\omega r_1 r_2 \csc \omega t). \tag{3.12}$$

Again, using (2.13), we find that (3.12) shows the structure (2.14) with the one-dimensional harmonic-oscillator action[15]

$$S(r_1,r_2,t) = \tfrac{1}{2}\omega(r_1^2 + r_2^2)\cot \omega t - \omega r_1 r_2 \csc \omega t. \tag{3.13}$$

The spectral representation of the propagator (3.12) follows from the Hille–Hardy formula[16,17]:

$$\frac{e^{(x+y)h/(1+h)}}{1+h}\, \frac{J_\mu[2\sqrt{xyh}^{\,1/2}/(1+h)]}{(xyh)^{\mu/2}} = \sum_{\lambda=0}^\infty \frac{\lambda!}{\Gamma(\lambda+\mu+1)}(-h)^\lambda L_\lambda^{(\mu)}(x) L_\lambda^{(\mu)}(y). \tag{3.14}$$

Expressing the Laguerre functions $L_\lambda^{(\mu)}$ in terms of Whittaker functions[18] and rearranging, we obtain

$$h^{-\mu/2}(1+h)^{-1}\sqrt{xy}e^{(x+y)(1+h)/(1-h)/2} \times J_\mu[2\sqrt{xy}h^{1/2}/(1+h)] = \sum_{\lambda=0}^\infty (-h)^\lambda \frac{\Gamma(\lambda+\mu+1)}{\lambda!} \times M_{\lambda+(\mu+1)/2}^{\mu/2}(x) M_{\lambda+(\mu+1)/2}^{\mu/2}(y). \tag{3.15}$$

With the substitutions: $\mu = L + N/2 - 1$, $x = \omega r_1^2$, $y = \omega r_2^2$, $h = -e^{-2i\omega t}$, $\lambda = n$, we obtain

$$K_L^{(N)}(r_1,r_2,t) = \sum_{n=0}^\infty R_{n,L}^{(N)}(r_1) R_{n,L}^{(N)}(r_2) e^{-iE_{n,L}^{(N)}t} \tag{3.16}$$

with the eigenvalues $E_{n,L}^{(N)}$ given by (3.7) and the radial eigenfunctions by

$$R_{n,L}^{(N)}(r) = \left[\frac{2\Gamma(n+L+N/2)}{n!}\right]^{1/2} r^{-N/2} M_{n+L/2+N/4}^{(L+N/2-1)/2}(\omega r^2) \quad (n = 0,1,2,\cdots). \tag{3.17}$$

Again we note the special cases $N = 2$ and 3. For $N = 2$,[19]

$$K_m^{(2)}(\rho_1,\rho_2,t) = (-i)^{m+1}\omega \csc \omega t \times e^{(1/2)i\omega(\rho_1^2 + \rho_2^2)\cot \omega t} J_m(\omega\rho_1\rho_2 \csc \omega t) \quad (M = 0, \pm 1, \pm 2,\cdots). \tag{3.18}$$

The summation analogous to (2.17) results in

$$K^{(2)}(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2, t) = (-i/2\pi)\omega \csc \omega t \times \exp[\tfrac{1}{2}i\omega(\rho_1^2 + \rho_2^2)\cot \omega t - i\omega \boldsymbol{\rho}_1 \cdot \boldsymbol{\rho}_2 \csc \omega t]. \tag{3.19}$$

For $N = 3$:

$$K_l^{(3)}(r_1,r_2,t) = (-i)^{l+3/2}(2/\pi)^{1/2}(\omega \csc \omega t)^{3/2} \times e^{(1/2)i\omega(r_1^2 + r_2^2)\cot \omega t} j_l(\omega r_1 r_2 \csc \omega t). \tag{3.20}$$

The sum over partial waves as in (2.21) gives[20]

$$K^{(3)}(\mathbf{r}_1,\mathbf{r}_2,t) = (\omega \csc \omega t/2\pi i)^{3/2} \times \exp[\tfrac{1}{2}i\omega(r_1^2 + r_2^2)\cot \omega t - i\omega \mathbf{r}_1 \cdot \mathbf{r}_2 \csc \omega t]. \tag{3.21}$$

4. N-DIMENSIONAL COULOMB PROBLEM

We consider finally the N-dimensional hydrogenic radial equation

$$\frac{1}{2}\left[k^2 + \frac{1}{r^{N-1}}\frac{\partial}{\partial r} r^{N-1}\frac{\partial}{\partial r} - \frac{L(L+N-2)}{r^2} + \frac{2Z}{r}\right] \times G_L^{(N)}(r,r',k) = \delta(r-r')/(rr')^{N/2-1/2}. \tag{4.1}$$

The appropriate solutions to the homgeneous equation are, in this case[21]

$$u(r) = r^{1/2-N/2} M_{i\nu}^{L+N/2-1}(-2ikr), \qquad v(r) = r^{1/2-N/2} W_{i\nu}^{L+N/2-1}(-2ikr), \tag{4.2}$$

where

$$\nu \equiv Z/k. \tag{4.3}$$

Using (3.4) and (1.3) once again, we obtain the Coulomb Green's functions

$$G_L^{(N)}(r_1,r_2,k) = (ik)^{-1}\Gamma(L+N/2-1/2-i\nu)\times(r_1r_2)^{1/2-N/2}M_{i\nu}^{L+N/2-1}(-2ikr_<)\times W_{i\nu}^{L+N/2-1}(-2ikr_>), \tag{4.4}$$

a result previously given by Hostler.[22] The poles of the gamma function at $\nu = -i(L+N/2-1/2+n')$, $n' = 0,1,2,\ldots$, determine the N-dimensional hydrogenic spectrum:

$$E_{n,L}^{(N)} = Z^2/2\nu^2 = \frac{-Z^2}{2(n+N/2-3/2)^2}, \quad n = L+1, L+2,\cdots. \tag{4.5}$$

The integral representation (3.8) is again applicable, now with $\mu/2 = L+N/2-1$, $\kappa = i\nu$, $t = -2ik$, $a_1 = r_>$, $a_2 = r_<$, and the variable transformation $\sinh v = \operatorname{csch} q$. We obtain

$$G_L^{(N)}(r_1,r_2,k) = -2(-i)^{2L+N-2}(r_1r_2)^{1-N/2}\times\int_0^\infty dq\, e^{2i\nu q}\operatorname{csch} q\, e^{ik(r_1+r_2)\coth q}\times J_{2L+N-2}(2k\sqrt{r_1r_2}\operatorname{csch} q). \tag{4.6}$$

Unfortunately, this is not a Fourier transform wrt time and energy variables. In fact, no closed form for the Coulomb propagator is known, as yet. Equation (4.6) does, however, represent a Fourier transform wrt the quantum number variable ν. In a similar instance, we have introduced the *Sturmian propagator*,[23] defined by the transform

$$G(r_1,r_2,\nu) = \frac{-2i}{k}\int_0^\infty dq\, e^{2i\nu q}S(r_1,r_2,q). \tag{4.7}$$

For the Green's function (4.6), we identify the corresponding Sturmian propagator

$$S_L^{(N)}(r_1,r_2,q) = (-i)^{2L+N-1}k(r_1r_2)^{1-N/2}\operatorname{csch} q\times e^{ik(r_1+r_2)\operatorname{csch} q}J_{2L+N-2}(2k\sqrt{r_1r_2}\operatorname{csch} q). \tag{4.8}$$

By substituting $Z = k\nu$ [cf. Eq. (4.3)] in (4.1) and using the Fourier transform (4.7), we obtain a partial differential equation for S:

$$\frac{1}{2}\left[k^2 + \frac{1}{r^{N-1}}\frac{\partial}{\partial r}r^{N-1}\frac{\partial}{\partial r} - \frac{L(L+N-2)}{r^2} + \frac{ik}{r}\frac{\partial}{\partial q}\right]S = 0, \quad r = r_1, r_2 \tag{4.9}$$

subject to the boundary condition

$$S_L^{(N)}(r_1,r_2,0) = \delta(r_1-r_2)/(r_1r_2)^{N/2-1}. \tag{4.10}$$

The propagator (4.8), with k and ν real, pertains to the Coulomb continuum. Of more significance is the discrete spectrum Sturmian propagator, obtained by the substitutions: $k \to ik$, $q \to iq$, viz.,

$$S_L^{(N)}(r_1,r_2,q) = (-i)^{2L+N-1}k(r_1r_2)^{1-N/2}\csc q\times e^{ik(r_1+r_2)\cot q}J_{2L+N-2}(2k\sqrt{r_1r_2}\csc q). \tag{4.11}$$

The spectral representation of (4.11) follows again from (3.15), with $\mu = 2L+N-2$, $x = 2kr_1$, $y = 2kr_2$, $h = -e^{-2iq}$, $\lambda = n-L-1$. The result is

$$S_L^{(N)}(r_1,r_2,q) = \sum_{n=L+1}^{\infty} R_{n,L}^{(N)}(r_1)R_{n,L}^{(N)}(r_2)\times e^{-2iq(n+N/2-3/2)}, \tag{4.12}$$

with

$$R_{n,L}^{(N)}(r) = \left[\frac{(n+L+N-3)!}{(n-L-1)!}\right]^{1/2}\times r^{1/2-N/2}M_{n+N/2-3/2}^{L+N/2-1}(2kr), \quad n = L+1, L+2,\cdots. \tag{4.13}$$

If $k = Z/n$ then (4.13) gives the N-dimensional Coulomb radial eigenfunctions. For k arbitrary, as is the case here, the $R_{n,L}^{(N)}(r)$ represent Sturmian functions,[24] hence our designation for the propagator $S_L^{(N)}$.

Relationships between Coulomb eigenstates and those of harmonic oscillators of various dimension have been known for a long time.[25] This connection manifests itself in the similarity of the propagator $S_L^{(N)}(r_1,r_2,q)$ to the harmonic-oscillator propagator $K_\lambda^{(\nu)}(\rho_1,\rho_2,t)$. Specifically, under the substitutions

$$r = \rho^2/2, \quad k = \omega, \quad q = \omega t \tag{4.14}$$

and

$$2L+N-2 = \lambda + \nu/2 - 1 \tag{4.15}$$

the two propagators are related by

$$(r_1r_2)^{N/2-1}S_L^{(N)}(r_1,r_2,q) = (\rho_1\rho_2)^{\nu/2-1}K_\lambda^{(\nu)}(\rho_1,\rho_2,t). \tag{4.16}$$

A formula equivalent to (4.15) was found by Giovannini and Tonietti.[26] Two realizations of (4.15) for the three-dimensional Coulomb problem ($N = 3$, $L = l$) have been given. Schwinger[27] set $\nu = 2$, so that $2l+1 = \lambda = |m|$, thus connecting hydrogenic states to those of a two-dimensional oscillator. Bergmann and Frishman[28] set $\lambda = 0$, so that $4l+4 = \nu$, thus establishing a connection with states of a ν-dimensional oscillator. Further, comparison of (4.13) with (3.17) shows that the hydrogenic principal quantum number n corresponds to the oscillator quantum number $n' = n-L-1$.

For the three-dimensional case, the sum over partial waves according to

$$S(\mathbf{r}_1,\mathbf{r}_2,q) = \sum_{l=0}^{\infty}\frac{2l+1}{4\pi}P_l(\cos\theta)S_l^{(3)}(r_1,r_2,q) \tag{4.17}$$

can be evaluated using Neumann's formula[29]

$$\tfrac{1}{2}zJ_0(z\cos\theta/2) = \sum_{l=0}^{\infty}(-)^l(2l+1)P_l(\cos\theta)J_{2l+1}(z). \tag{4.18}$$

The result is the Coulomb Sturmian propagator[23]

$$S(\mathbf{r}_1,\mathbf{r}_2,q) = -(4\pi)^{-1}k^2\csc^2 q\times e^{ik\xi\cot q}J_0(k\eta\csc q), \tag{4.19}$$

$$\xi \equiv r_1+r_2, \quad \eta \equiv 2\sqrt{r_1r_2}\cos\theta/2.$$

Substituting (4.19) back into (4.7), we recover an integral representation for the Coulomb Green's function first derived by Hostler.[30]

[1] The form of the N-dimensional Laplacian is given by J. D. Louck, J. Mol. Spectrosc. **4**, 298 (1960).
[2] See, for example, S. I. Vetchinkin and V. L. Bachrach, Int. J. Quantum Chem. **6**, 143 (1972).
[3] See, for example, S. M. Blinder, *Foundations of Quantum Dynamics* (Academic, New York, 1974), p. 148ff.
[4] M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions* (National Bureau of Standards, Washington, D.C., 1972), p. 360.
[5] G. N. Watson, *Theory of Bessel Functions*, 2nd ed. (Cambridge University, England, 1966), p. 439, Eq. (2).
[6] Reference 4, p. 361, Eq. (9.1.30).
[7] Reference 4, p. 362, Eq. (9.1.69).
[8] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw–Hill, New York, 1965), p. 58ff; Ref. 3, p. 152.
[9] Reference 4, p. 440, Eq. (10.1.47).
[10] Reference 3, p. 155, Eq. (6.5.56).
[11] H. Buchholz, *The Confluent Hypergeometric Function* (Springer, New York, 1969), pp. 32–33, Eqs. (3a), (3b) with $z = r, \lambda = 2, \alpha_\lambda = 0, \beta = -N/2, A = \omega, \kappa = E/2\omega, \mu = L + N/2 - 1$.
[12] S. M. Blinder, J. Math. Phys. **22**, 306 (1981).
[13] Reference 11, p. 25, Eq. (33).
[14] Reference 11, p. 86, Eq. (5c).
[15] Reference 8, p. 63, Eq. (3.59); Ref. 3, p. 11, Eq. (1.4.29).
[16] Reference 11, p. 139, Eq. (12a).
[17] A. Erdelyi, ed., *Higher Transcendental Functions* (McGraw–Hill, New York, 1953), Vol. 2, p. 189.
[18] Reference 11, p. 212.
[19] A very similar formula for a harmonic oscillator perturbed by an inverse quadratic potential has been given by D. C. Khandekar and S. V. Lawande, J. Math. Phys. **16**, 384 (1975).
[20] Reference 3, p. 159, Eq. (6.5.87).
[21] Reference 11, with $z = r, \lambda = 1, \beta = (1 - N)/2, A = -2ik, \kappa = iZ/k, \mu/2 = L + N/2 - 1$.
[22] L. C. Hostler, J. Math. Phys. **11**, 2966 (1970).
[23] S. M. Blinder, "Sturmian propagator for the nonrelativistic Coulomb problem," Phys. Rev. A (in press).
[24] M. Rotenberg, Ann. Phys. (NY) **19**, 262 (1962); Adv. At. Mol. Phys. **6**, 233 (1970).
[25] E. Schrödinger, Proc. R. Irish Acad. A **46**, 183 (1941).
[26] A. Giovannini and T. Tonietti, Nuovo Cimento A **54**, 1 (1968).
[27] J. Schwinger, unpublished lecture notes; see also G. Bahm, *Lectures on Quantum Mechanics* (Benjamin, Reading, MA, 1969), p. 179.
[28] D. Bergmann and Y. Frishman, J. Math. Phys. **6**, 1855 (1965).
[29] Reference 5, p. 140, Eq. (3) with $\nu = 0, \mu = 1, k = \cos(\theta/2), P_l(\cos\theta) = (-)^l {}_2F_1[-l, l+1; 1; \cos^2(\theta/2)]$.
[30] L. Hostler, J. Math. Phys. **5**, 591 (1964).

Analytic treatment of the $\alpha x^2 + \beta x^4$ oscillator

George P. Flessas and R. R. Whitehead
Department of Natural Philosophy, The University, Glasgow G12 8QQ, Scotland, United Kingdom
A. Rigas[a)]
Department of Physiology, The University, Glasgow G12 8QQ, Scotland, United Kingdom

(Received 3 December 1982; accepted for publication 17 June 1983)

We carry out an investigation of the Schrödinger equation with the $\alpha x^2 + \beta x^4$ interaction for $0 < \mathrm{Re}(x) < \infty$, and $\beta = |\beta| \exp(i(\pi + \lambda))$, $-\pi < \lambda < \pi$. In the sectors $0 \leqslant \arg(x) < \pi/3 - \lambda/6$, $0 \leqslant \lambda < \pi$, and $-\lambda/6 < \arg(x) < \pi/3 - \lambda/6$, $-\pi < \lambda < 0$, a subdominant solution is constructed essentially in terms of the Laplace transform of a function $f(s)$, $0 \leqslant s < \infty$, which is expressible as a converging power series in s. The solution, $y(x)$, thus obtained, is compared and contrasted with the one valid in the usual case where $\mathrm{Re}(x) \leqslant 0$ is included and it is argued that $y(x)$ may not be accessible to perturbative approaches.

PACS numbers: 03.65. — w

I. INTRODUCTION

In recent years the quantum mechanics of the anharmonic oscillator has been the subject of a large number of publications. Interest in such a model stems primarily from two factors. First, this system provides a simple model for field theories with polynomial interactions.[1,2] Second, and likewise important, the fact that, until now, the Schrödinger equation of the general anharmonic oscillator has not been solved exactly has stimulated the development of many analytic and numerical methods which, besides being used for the calculation of the energy eigenvalues, are of a much wider applicability and usefulness.[1–9]

In the course of the investigations on the anharmonic interactions, exact analytic solutions and eigenvalues have been found for a class of such potentials.[10–13] Such solutions are of importance since they can serve as a testing ground for various approximate approaches. Indeed, some of these solutions may not be accessible to Rayleigh–Schrödinger perturbation theory.[14] Furthermore, exact results may provide us with a better insight into the underlying physics, as such results are free of any assumptions that would be needed in the case of an approximate treatment of the problem in question.

A striking feature, however, of the general anharmonic oscillator defined by

$$V_N(x) = \sum_{i=1}^{N} \beta_i x^{2i}, \quad -\infty < x < \infty, \quad N = 1,2,3,... \tag{1}$$

is the following. All the methods so far yield exact results only when $N=$ odd, and completely fail to do so if $N=$ even. More specifically, the solution y_{2n+1} corresponding to $N = 2n+1$, $n = 0, 1,...$, is a function of the parameters β_i, the dependence on β_{2n+1} being of the form $1/\beta_{2n+1}$. Hence if we attempt, by letting $\beta_{2n+1} \to 0$, $\beta_{2n} \neq 0$, to generate, from y_{2n+1}, the solution y_{2n} relevant to $N = 2n$, we obtain infinities which do not cancel. This can be most easily seen in the case of the doubly anharmonic oscillator $(N = 3)$[10,11] where $\beta_3 \to 0$ does *not* lead to an exact solution for $\beta_1 x^2 + \beta_2 x^4$. In this respect, an interesting observation can be made. In fact, a basic set S of the Schrödinger equation with the $\beta_1 x^2 + \beta_2 x^4$ potential consists of a subdominant, y_s, and a dominant, y_d, solution[3,15–16] and, as we have argued in a recent work,[16] due to the appearance of $1/\beta_2$ in y_s and y_d, their structure and analytic properties do not seem to be obtainable by means of standard perturbation methods based on the $\beta_1 x^2$ harmonic oscillator. Since, of course, the physical solution pertaining to the $\beta_1 x^2 + \beta_2 x^4$ interaction is a linear combination of y_s and y_d, it appears that there is a nonconventional perturbative transition from $\beta_1 x^2$ to $\beta_1 x^2 + \beta_2 x^4$, and indeed it is well known that the Raleigh–Schrödinger perturbation theory diverges for the $\beta_1 x^2 + \beta_2 x^4$ oscillator. In this context we also note the important remark by Simon[3] that, namely, $\beta_2 x^4$ is a singular perturbation in the sense that a radical change in the physics is taking place when the perturbation is switched on. Now in the case of Eq. (1) and for $N = 2n+1$, the exact solutions are generalizations[10–13] of the solution for $\beta_1 x^2$. If we add to $V_{2n+1}(x)$ the term $\beta_{2n+2} x^{2(2n+2)}$, all the techniques that work for $V_{2n+1}(x)$ break down as previously noted when they are applied to $V_{2n+2}(x) = V_{2n+1}(x) + \beta_{2n+2} x^{2(2n+2)}$ and thus we arrive at a situation which may be taken as equivalent to or a generalization of the situation prevalent in the transition $\beta_1 x^2 \to \beta_1 x^2 + \beta_2 x^4$, which corresponds to passing from $N = 1$ to $N = 2$ in Eq. (1). To put it otherwise, a basic system of the Schrödinger equation corresponding to $N = 2n$ may not be derivable from a basic system relevant to $N = 2n-1$ by means of a straightforward perturbative approach. We feel that this signifies, yet again, a radical change in the underlying physics. This point of view is corroborated by the discovery of some rather unexpected and exciting features of the $\beta_1 x^2 + \beta_2 x^4$ potential.[1,3,17–19] Therefore it might be worthwhile to investigate $V_N(x)$ further. A way to realize such an investigation could be by trying to generalize the analytic results known for the $N = 2$ case to higher even N's, much the same as it has been done for $N=$ odd, as we pointed out above (cf. Refs. 10–13).

In this contribution, therefore, we attempt to find some further analytic properties of the interaction

$$V(x) = \alpha x^2 + \beta x^4, \quad 0 < \mathrm{Re}(x) < \infty, \tag{2}$$

[a)] Author of Appendix.

 0022-2488/84/040910-08$02.50

where α is an arbitrary complex number and β is complex such that

$$\beta = |\beta| \exp(i(\pi + \lambda)), \quad -\pi < \lambda < \pi, \tag{3}$$

in the hope of further elucidating the nonconventional perturbative nature of βx^4. Equation (2) represents the simplest N-case in Eq. (1), that is, $N = 2$. The restriction of $\mathrm{Re}(x)$ to the positive half-axis implies that we are considering $\alpha x^2 + \beta x^4$ in the context of a cut-off theory. In fact, if α = real and $\lambda = 0$, that is, $\beta < 0$, the constraint $x > 0$ means that we introduce a small cut-off distance $\epsilon > 0$ and set $V(x) = 0$ for $-\infty < x \leqslant \epsilon$ and $V(x) = \alpha x^2 + \beta x^4$ for $x > \epsilon$. Since ϵ is an arbitrary positive number, the solution valid for $x > \epsilon$ holds for all $x > 0$, hence Eq. (2) which refers to the case of a complex x. The reason that ϵ must be positive stems from the following. Our aim is to obtain analytic results for $\alpha x^2 + \beta x^4$ and, consequently, we must make some appropriate ansatz for the solution of the Schrödinger equation in question. In so doing we are forced by the very nature of our ansatz to exclude $\mathrm{Re}(x) \leqslant 0$ or, equivalently, $x \leqslant 0$ for real x, and, hence, the introduction of ϵ as described above is an attempt to attribute a physical meaning to the condition $x \leqslant 0$. A possible justification for the constraint $\mathrm{Re}(x)$ may be the fact that, in this way, we obtain a subdominant solution of probably nonperturbative character. Thus in Sec. II, we carry out the solution of the Schrödinger equation pertaining to Eq. (2) and obtain the aforementioned subdominant solution. In Sec. III we discuss our result and compare it with that relevant to the usual $\alpha x^2 + \beta x^4$, $-\infty < \mathrm{Re}(x) < \infty$, oscillator. In the Appendix we present a numerical investigation related to Sec. I.

II. SOLUTION OF THE SCHRÖDINGER EQUATION

We consider

$$y''(x) + (E - \alpha x^2 - \beta x^4)y(x) = 0, \quad \mathrm{Re}(x) > 0, \tag{4}$$

and represent $y(x)$ as

$$y(x) = \exp(ax + bx^3)\int_0^\infty \exp(-sx) f(s) ds, \quad \mathrm{Re}(x) > 0. \tag{5}$$

Then we observe that Eq. (5) solves Eq. (4) if the function $f(s)$ fulfills

$$I(\infty) - I(0) = 0, \quad I(s) = 6b\,(\exp(-sx)s[x f(s) + f'(s)]), \tag{6}$$

$$6bsf''(s) + 6bf'(s) + [-s^2 + 2as - (a^2 + E)] f(s) = 0 \tag{7}$$

and the parameters a, b are given by

$$9b^2 = \beta, \quad 6ab = \alpha. \tag{8}$$

From the indicial equation of Eq. (7) it follows that

$$f(s) = \sum_{n=0}^\infty c_n s^n, \quad c_0 \neq 0, \tag{9}$$

which converges for $s \in [0, \infty)$ according to the general theory of differential equations[20] and

$$6b(n+1)^2 c_{n+1} - (a^2 + E)c_n + 2ac_{n-1} - c_{n-2} = 0,$$
$$c_{-1} = c_{-2} = 0, \quad n = 0,1,2,\ldots . \tag{10}$$

Deferring the justification of Eq. (5) to Sec. III of this paper, we present our results in the form of a theorem.

Theorem 1: The Schrödinger equation (4) can be solved by Eqs. (5)–(10) which constitute a subdominant solution valid in the sectors $0 \leqslant \arg(x) < \pi/3 - \lambda/6$, $0 \leqslant \lambda < \pi$, and $-\lambda/6 < \arg(x) < \pi/3 - \lambda/6$, $-\pi < \lambda < 0$.

Proof: The main part of the proof involves the verification of the relation

$$\lim_{s \to \infty} [f(s) \exp(-sx)] = 0, \quad \mathrm{Re}(x) > 0. \tag{11}$$

If Eq. (11) holds, then the integral in Eq. (5) exists.[21] To ascertain the behavior of $f(s)$ for $s \to \infty$, we first observe that Eq. (10) gives

$$c_{n+1}/c_n = 1/[(6b)^{1/3} n^{2/3}], \quad n \gg 1. \tag{12}$$

Since, from Eq. (3), we get, with $L = (4|\beta|)^{-1/6} > 0$ and Eq. (8),

$$(6b)^{1/3} = L^{-1} \exp[i((\pi + \lambda)/6 + 2\pi k/6)], \; k = 0,1,\ldots,5, \tag{13}$$

Eq. (12) yields

$$\frac{c_{n+1}}{c_n} = \frac{iL \exp(-i\lambda/6)}{n^{2/3}}, \quad n \gg 1, \tag{14}$$

where in Eq. (13) we have taken $k = 4$. Equation (14) shows that $f(s)$ for $s \gg 1$ behaves as[22]

$$\begin{aligned} F(s) &= \sum_{n=0}^\infty i^n \exp\left(\frac{-i\lambda n}{6}\right)\frac{z_1^n}{(n!)^{2/3}} \\ &= \sum_{n=0}^\infty (-1)^n \frac{\cos(\lambda n/6)}{[(2n)!]^{2/3}} z^n \\ &\quad - \frac{i}{z_1} \sum_{n=1}^\infty (-1)^n \frac{\cos(\lambda n/6)}{[(2n-1)!]^{2/3}} z^n \\ &\quad - i \sum_{n=0}^\infty (-1)^n \frac{\sin(\lambda n/6)}{[(2n)!]^{2/3}} z^n \\ &\quad - \frac{1}{z_1} \sum_{n=1}^\infty (-1)^n \frac{\sin(\lambda n/6)}{[(2n-1)!]^{2/3}} z^n \\ &= \Sigma_1 - \frac{i}{z_1}\Sigma_2 - i\Sigma_3 - \frac{1}{z_1}\Sigma_4, \quad z_1 = Ls, \quad z = z_1^2. \end{aligned} \tag{15}$$

The meaning of the abbreviations $\Sigma_1,\ldots,\Sigma_4$ is obvious from Eq. (15). On comparing now Σ_1 with

$$\exp(-z) = \sum_{n=0}^\infty (-1)^n \frac{z^n}{n!}, \tag{16}$$

and by application of a method from the theory of analytic continuation,[23] we obtain

$$\Sigma_1 = \exp(-z) \sum_{n=0}^\infty S_n \frac{z^n}{n!} = \exp(-z)\Sigma_1', \tag{17}$$

$$S_n = n! \sum_{k=0}^n (-1)^k \frac{\cos(\lambda k/6)}{[(2k)!]^{2/3}(n-k)!}. \tag{18}$$

It will be shown in Lemma 2.1 that

$$|S_n| < n, \quad n \geqslant N, \quad N = \text{finite}. \tag{19}$$

Thus from Eqs. (17)–(19) we deduce

$$|\Sigma_1'| < 1 + \sum_{n=1}^{N-1}\left(\frac{|S_n|}{n!} - \frac{1}{(n-1)!}\right)z^n + \sum_{n=1}^{\infty}\frac{z_n}{(n-1)!} < 1 + \sum_{n=1}^{N-1}\left(\frac{|S_n|}{n!} - \frac{1}{(n-1)!}\right)z^n + \sum_{n=1}^{\infty}\frac{(n+1)}{n!}z^n = \sum_{n=1}^{N-1}\left(\frac{|S_n|}{n!} - \frac{1}{(n-1)!}\right)z^n + \exp(z)[(1+z)], \tag{20}$$

and so from Eq. (17),

$$|\Sigma_1| < 1 + z + \left(\sum_{n=1}^{N-1}\left(\frac{|S_n|}{n!} - \frac{1}{(n-1)!}\right)z^n\right)\exp(-z), \quad s \gg 1. \tag{21}$$

In precisely the same manner, we may show (Lemma 2.1)

$$|\Sigma_3| < 1 + z + \left(\sum_{n=1}^{N-1}\left(\frac{|S'_n|}{n!} - \frac{1}{(n-1)!}\right)z^n\right)\exp(-z), \quad s \gg 1. \tag{22}$$

Moreover, Σ_2 and Σ_4 clearly have the same asymptotic behavior for $s \gg 1$ with Σ_1, and, respectively, Σ_3.[24] Hence from Eqs. (21) and (22), it follows that $|F(s)|$ and, consequently, also $|f(s)|$ for $s \gg 1$ behaves weaker than

$$2(1+z)(1+1/z_1), \quad N = \text{finite}. \tag{23}$$

We conclude now that Eq. (23) ensures the validity of Eq. (11) and thus the existence of the integral in Eq. (5). Evidently Eq. (6) is also satisfied. For the parameters a and b, we can choose, by virtue of Eqs. (3) and (8),

$$a = \frac{\alpha}{2|\beta|^{1/2}(i\cos\lambda/2 - \sin(\lambda/2))}, \qquad b = \frac{|\beta|^{1/2}}{3}(i\cos(\lambda/2) - \sin(\lambda/2)). \tag{24}$$

The value for the parameter b given in Eq. (24) is consistent with the one following from Eq. (13) for $k = 4$. The boundary condition that must be fulfilled reads

$$\lim_{|x|\to\infty} y(x) = 0$$

$$\text{when } \begin{cases} 0 \leqslant \arg(x) < \pi/3 - \lambda/6, & 0 \leqslant \lambda < \pi, \\ -\lambda/6 < \arg(x) < \pi/3 - \lambda/6, & -\pi < \lambda < 0. \end{cases} \tag{25}$$

From Eqs. (5) and (24), it is seen that Eq. (25) is indeed satisfied and so, owing to the fact that Eq. (12), which is actually the starting point for the proof of the theorem, is valid for every $|E| < \infty$, the last part of Theorem 1 follows. It should be noted that, by virtue of Eq. (10), c_n can easily be written as a determinant and that, as the structure of Eq. (10) shows, there do not exist E-values for which the series for $f(s)$ in Eq. (9) rigorously terminates.

Lemma 2.1: The sum S_n defined in Eq. (18) satisfies inequality (19).

Proof: It is sufficient to consider here $0 \leqslant \lambda < \pi/2$ since $\lambda = \pi/2$ has been investigated elsewhere.[25] We take the function

$$g(z) = \frac{\cos(\lambda z/6)}{\Gamma(n+1-z)[\Gamma(2z+1)]^{2/3}}, \quad z = x + iy, \tag{26}$$

where $\Gamma(Z)$ denotes the Gamma function, and by utilizing the technique for the conversion of a sum into a complex integral,[26] we get

$$S_n = \frac{1}{2i}\oint \Gamma(n+1)g(z)\csc(\pi z)dz. \tag{27}$$

The contour integral in Eq. (27) is taken counterclockwise around a closed rectangular contour encircling the points $0,1,\ldots,n$. Therefore it breaks up into four integrals, the first of which is

$$I_1 = \frac{\Gamma(n+1)}{2i}\int_{-L}^{L}\frac{[\exp(-i\lambda/24 - \lambda y/6) + \exp(i\lambda/24 + \lambda y/6)]\exp(i\pi/4)}{\Gamma(n+1+\frac14 - iy)[\Gamma(-\frac{2}{4}+1+2iy)]^{2/3}[\exp(-\pi y) - i\exp(\pi y)]}dy \tag{28}$$

with $z = -\frac14 + iy$. The relations[27]

$$|\Gamma(\tfrac12 + 2iy)| = \pi^{1/2}/(\cosh 2\pi y)^{1/2}, \quad |\Gamma(\bar Z)| = |\Gamma(Z)|, \quad Z = X + iY \tag{29}$$

permit us to write

$$|I_1| \leqslant \frac{\Gamma(n+1)}{(2\pi)^{1/3}}\int_0^L \frac{[\exp(\pi y + \lambda y/6) + \exp(\pi y - \lambda y/6)][\exp(2\pi y) + \exp(-2\pi y)]^{1/3}}{|\Gamma(n+1+\frac14+iy)||1 - i\exp(2\pi y)|}dy. \tag{30}$$

Bearing in mind the investigation of the integrals along the three other sides of the contour, we now choose

$$L = n + \tfrac14, \quad n \gg 1, \quad (L \gg 1), \tag{31}$$

and investigate Eq. (30) for the case of Eq. (31). In this respect, we can apply the general formula for the asymptotic behavior of $\Gamma(Z)$ as given by Morse and Feshbach[28]:

$$|\Gamma(X+1-iY)| = (2\pi)^{1/2}[(X^2+Y^2)^{(2X+1)/4}]\exp(-Y\phi + X), \quad \phi = \tan^{-1}(Y/X), \quad |Z| \gg 1, \tag{32}$$

and thus for $n \gg 1$ and $0 \leqslant y \leqslant L = n + \frac14$,

$$\frac{\Gamma(n+1)}{|\Gamma(n+1+\frac14+iy)|} = \frac{\exp(\phi y)}{n^{1/4}[1 + y^2/(n+\frac14)^2]^{(4n+3)/8}}, \qquad \phi = \tan^{-1}\frac{y}{n+\frac14}. \tag{33}$$

By virtue of Eq. (31) and (33) max $\phi = \pi/4$, and since, always,

$$[(y^2/(n+\tfrac{1}{4})^2)+1]^{(4n+3)/8} \geqslant 1$$

for $0 \leqslant y \leqslant L = n + \frac{1}{4}$, Eq. (30) goes into

$$|I_1| \leqslant \frac{1}{n^{1/4}(2\pi)^{1/3}} \int_0^L \frac{(\exp[(\pi+\pi/4+\lambda/6)y]+\exp[(\pi+\pi/4-\lambda/6)y])[\exp(2\pi y)+\exp(-2\pi y)]^{1/3}}{[1+\exp(4\pi y)]^{1/2}}\,dy$$

$$< \frac{1}{n^{1/4}(2\pi)^{1/3}} \int_0^\infty \frac{\cdots}{\cdots}\,dy. \tag{34}$$

As the last integral in Eq. (34) obviously exists, we denote it with I, and so Eq. (34) yields

$$|I_1| < n^{-1/4} I (2\pi)^{-1/3}, \quad L = n + \tfrac{1}{4}, \quad n \gg 1, \quad (L \gg 1). \tag{35}$$

The second integral I_2, along the contour, satisfies with $z = x - iL$

$$|I_2| \leqslant \Gamma(n+1) \int_{-1/4}^{n+1/4} \frac{\exp(\pi L + L\lambda/6)}{|\Gamma(n-x+1+iL)||\Gamma(2x+1+2iL)|^{2/3}|\exp[2\pi(ix+L)]-1|}\,dx, \tag{36}$$

where we have used $|\Gamma(\bar{Z})| = |\Gamma(Z)|$ and the obvious relation $\exp(\pi L + L\lambda/6) \geqslant \exp(\pi L - L\lambda/6)$. Introducing now Eq. (31) into Eq. (36), we can utilize Eq. (32) for the Gamma function because in all three cases $|Z| \gg 1$ holds. The result of some elementary calculations is

$$|I_2| \leqslant \frac{G}{(2\pi)^{1/3}} \exp\left(\frac{-1}{4}\right) \int_{-1/4}^{n+1/4} (\exp\{x/3 - \ln(n)[(x+1)/3] + L(\phi + 4\phi'/3 + \lambda/6 - \pi)\})dx, \tag{37}$$

$$\phi = \tan^{-1}\frac{L}{n-x}, \quad \phi' = \tan^{-1}\frac{L}{x}, \quad G = \frac{1}{1-\exp(-2\pi L)}, \quad \frac{1}{\exp(2\pi L)-1} > \frac{1}{|\exp[2\pi(ix+L)]-1|}. \tag{38}$$

The quantities

$$\left[\left(\frac{n-x}{L}\right)^2+1\right]^{(2n-2x+1)/4}, \quad \left[\left(\frac{x}{L}\right)^2+1\right]^{(4x+1)/6}, \quad 2^{(4x+1)/3} \tag{39}$$

TABLE I. S_n as a function of n: $\lambda = \pi/4$.

n	S_n	n	S_n	n	S_n	n	S_n
0	1.000 000	33	− 0.025 845	66	− 0.013 807	99	− 0.009 903
1	0.375 423	34	− 0.025 857	67	− 0.013 649	100	− 0.009 821
2	− 0.016 961	35	− 0.025 687	68	− 0.013 496	101	− 0.009 741
3	− 0.246 162	36	− 0.025 368	69	− 0.013 344	102	− 0.009 662
4	− 0.363 513	37	− 0.024 932	70	− 0.013 196	103	− 0.009 584
5	− 0.406 697	38	− 0.024 409	71	− 0.013 050	104	− 0.009 509
6	− 0.402 953	39	− 0.023 824	72	− 0.012 906	105	− 0.009 434
7	− 0.371 599	40	− 0.023 201	73	− 0.012 765	106	− 0.009 361
8	− 0.326 010	41	− 0.022 559	74	− 0.012 626	107	− 0.009 289
9	− 0.275 157	42	− 0.021 914	75	− 0.012 489	108	− 0.009 218
10	− 0.224 798	43	− 0.021 279	76	− 0.012 355	109	− 0.009 148
11	− 0.178 383	44	− 0.020 663	77	− 0.012 223	110	− 0.009 080
12	− 0.137 745	45	− 0.020 074	78	− 0.012 093	111	− 0.009 013
13	− 0.103 617	46	− 0.019 516	79	− 0.011 965	112	− 0.008 947
14	− 0.076 008	47	− 0.018 992	80	− 0.011 840	113	− 0.008 882
15	− 0.054 482	48	− 0.018 503	81	− 0.011 718	114	− 0.008 818
16	− 0.038 358	49	− 0.018 050	82	− 0.011 598	115	− 0.008 755
17	− 0.026 841	50	− 0.017 632	83	− 0.011 480	116	− 0.008 693
18	− 0.019 117	51	− 0.017 246	84	− 0.011 365	117	− 0.008 632
19	− 0.014 412	52	− 0.016 892	85	− 0.011 252	118	− 0.008 573
20	− 0.012 025	53	− 0.016 565	86	− 0.011 141	119	− 0.008 514
21	− 0.011 346	54	− 0.016 264	87	− 0.011 033	120	− 0.008 455
22	− 0.011 860	55	− 0.015 986	88	− 0.010 927	121	− 0.008 398
23	− 0.013 148	56	− 0.015 728	89	− 0.010 824	122	− 0.008 342
24	− 0.014 878	57	− 0.015 488	90	− 0.010 723	123	− 0.008 286
25	− 0.016 795	58	− 0.015 263	91	− 0.010 624	130	− 0.007 970
26	− 0.018 713	59	− 0.015 051	92	− 0.010 527	140	− 0.007 499
27	− 0.020 500	60	− 0.014 851	93	− 0.010 432	150	− 0.007 087
28	− 0.022 073	61	− 0.014 660	94	− 0.010 339	160	− 0.006 724
29	− 0.023 382	62	− 0.014 478	95	− 0.010 248	170	− 0.006 400
30	− 0.024 406	63	− 0.014 302	96	− 0.010 159	180	− 0.006 111
31	− 0.025 147	64	− 0.014 132	97	− 0.010 072	190	− 0.005 849
32	− 0.025 619	65	− 0.013 967	98	− 0.009 987	200	− 0.005 612

have been discarded in the denominator of the integral of Eq. (38) since they are $\geqslant 1$. Further, the maximum value of $\phi + 4\phi'/3 + \lambda/6 - \pi$ is $\cong -\pi/12 + \lambda/6$. On carrying out the integration in Eq. (37), we obtain

$$|I_2| < \frac{3G\exp[(2\lambda-\pi)/48]}{(2\pi)^{1/3}\exp[n(\pi/12-\lambda/6)]}\left[\frac{\exp(-1/12)}{n^{1/4}} - \frac{1}{n^{1/3}}\left(\frac{e}{n}\right)^{(4n+1)/12}\right], \tag{40}$$

$$L = n + \tfrac{1}{4}, \quad n \gg 1 \quad (L \gg 1).$$

For the third integral on the contour we can easily deduce, with $z = n + \frac{1}{4} + iy$,

$$|I_3| \leqslant 2\Gamma(n+1)\int_0^L \frac{\exp(\lambda y/6)}{|\Gamma(\frac{3}{4}+iy)|\,|\Gamma(2n+\frac{1}{2}+1+2iy)|^{2/3}|\exp(-\pi y)+i\exp(\pi y)|}\,dy. \tag{41}$$

Using Eqs. (31) and (32), Eq. (41) is transformed into

$$|I_3| \leqslant \left(\frac{e}{n}\right)^{n/3}\frac{(2\pi)^{1/6}}{n^{1/6}2^{(4n+5)/3}}\int_0^L \frac{\exp\{4y[\tan^{-1}y/(n+\frac{1}{4})]/3+\lambda y/6\}}{|\Gamma(\frac{3}{4}+iy)|\,|\exp(-\pi y)+i\exp(\pi y)|}\,dy, \tag{42}$$

where the term $[y^2/(n+\frac{1}{4})^2+1]^{(2n+1)/3} \geqslant 1$ appearing in the integrand's denominator has been neglected. Noting that $\pi/4 = \max(\tan^{-1}y/(n+\frac{1}{4}))$, the integral I' in Eq. (42) fulfills

$$I' < \int_0^\infty \frac{\exp[y(\pi/3+\lambda/6)]}{|\Gamma(\frac{3}{4}+iy)|\,|\exp(-\pi y)+i\exp(\pi y)|}\,dy = I'' < \infty, \tag{43}$$

since I'' exists by virtue of the behavior of $|\Gamma(\frac{3}{4}+iy)|$ for $y \to \infty$ according to Eq. (32). Hence

$$|I_3| < \left(\frac{e}{n}\right)^{n/3}\frac{(2\pi)^{1/6}}{n^{1/6}2^{(4n+5)/3}}I'', \quad L = n + \tfrac{1}{4}, \quad n \gg 1, \quad (L \gg 1). \tag{44}$$

The last integral I_4 along the contour for which $z = x + iL$ with x varying between $n + \frac{1}{4}$ and $-\frac{1}{4}$ is readily seen to fulfill Eq. (36) and, thus, also Eq. (40). On recalling now Eqs. (27), (35), (40), and (44), we deduce that $|S_n|$, from a certain sufficiently *large*

TABLE II. S_n as a function of n: $\lambda = \pi/2$.

n	S_n	n	S_n	n	S_n	n	S_n
0	1.000 000	33	− 0.032 012	66	− 0.013 614	99	− 0.009 694
1	0.391 483	34	− 0.031 781	67	− 0.013 502	100	− 0.009 616
2	− 0.008 831	35	− 0.031 180	68	− 0.013 382	101	− 0.009 540
3	− 0.253 777	36	− 0.030 295	69	− 0.013 256	102	− 0.009 465
4	− 0.385 976	37	− 0.029 201	70	− 0.013 125	103	− 0.009 392
5	− 0.439 150	38	− 0.027 969	71	− 0.012 988	104	− 0.009 320
6	− 0.439 445	39	− 0.026 661	72	− 0.012 847	105	− 0.009 249
7	− 0.406 677	40	− 0.025 328	73	− 0.012 704	106	− 0.009 179
8	− 0.355 478	41	− 0.024 013	74	− 0.012 559	107	− 0.009 111
9	− 0.296 311	42	− 0.022 748	75	− 0.012 412	108	− 0.009 043
10	− 0.236 356	43	− 0.021 560	76	− 0.012 266	109	− 0.008 976
11	− 0.180 255	44	− 0.020 464	77	− 0.012 120	110	− 0.008 911
12	− 0.130 739	45	− 0.019 471	78	− 0.011 976	111	− 0.008 846
13	− 0.089 139	46	− 0.018 586	79	− 0.011 833	112	− 0.008 782
14	− 0.055 798	47	− 0.017 810	80	− 0.011 694	113	− 0.008 719
15	− 0.030 393	48	− 0.017 138	81	− 0.011 557	114	− 0.008 657
16	− 0.012 181	49	− 0.016 566	82	− 0.011 424	115	− 0.008 596
17	− 0.000 193	50	− 0.016 084	83	− 0.011 294	116	− 0.008 535
18	0.006 637	51	− 0.015 684	84	− 0.011 168	117	− 0.008 476
19	0.009 370	52	− 0.015 355	85	− 0.011 046	118	− 0.008 417
20	0.009 003	53	− 0.015 087	86	− 0.010 928	119	− 0.008 359
21	0.006 428	54	− 0.014 870	87	− 0.010 814	120	− 0.008 302
22	0.002 411	55	− 0.014 695	88	− 0.010 703	121	− 0.008 246
23	− 0.002 413	56	− 0.014 552	89	− 0.010 597	122	− 0.008 190
24	− 0.007 538	57	− 0.014 433	90	− 0.010 494	123	− 0.008 136
25	− 0.012 578	58	− 0.014 332	91	− 0.010 394	130	− 0.007 824
26	− 0.017 254	59	− 0.014 243	92	− 0.010 297	140	− 0.007 361
27	− 0.021 379	60	− 0.014 160	93	− 0.010 203	150	− 0.006 958
28	− 0.024 845	61	− 0.014 078	94	− 0.010 113	160	− 0.006 602
29	− 0.027 602	62	− 0.013 996	95	− 0.010 024	170	− 0.006 286
30	− 0.029 651	63	− 0.013 909	96	− 0.009 939	180	− 0.006 001
31	− 0.031 026	64	− 0.013 818	97	− 0.009 855	190	− 0.005 745
32	− 0.031 788	65	− 0.013 720	98	− 0.009 773	200	− 0.005 513

but *finite n*, say n_1, onwards remains bounded. Therefore we can find a finite $N > n_1$ such that $|S_n| < n$, $n \geqslant N$, and the proof of Eq. (19) is complete. In the case of Σ_3, we have to investigate the sum S'_n that derives from S_n in Eq. (18) by replacing $\cos(\lambda k/6)$ with $\sin(\lambda k/6)$. This is done by substituting $\sin(\lambda z/6)$ for $\cos(\lambda z/6)$ in Eq. (26), and it is readily seen, since such a substitution amounts to inessential replacements of $+$ signs by $-$ signs, that S'_n satisfies Eq. (19) too, and therefore also Eq. (22). It is clear that the proof of Eq. (19) is based mainly on Eq. (32) which retains its validity even for small $|Z|$.[29] Consequently, Eq. (19) is probably satisfied for small n too. This is numerically confirmed (cf. Appendix).

III. DISCUSSION AND CONCLUSIONS

An analysis of Eq. (4)[3,11] reveals that $\exp(ax + bx^3)$ is the correct asymptotic behavior of $y(x)$, a and b being appropriately chosen. Hence we have

$$y(x) = \exp(ax + bx^3)h(x) \tag{45}$$

and

$$h''(x) + (2a + 6bx^2)h'(x) + (a^2 + E + 6bx)h(x) = 0, \tag{46}$$

a and b fulfilling Eq. (8) yet again.

A natural thing to do is to solve Eq. (46) in the context of the theory of differential equations.[20] Therefore

$$h(x) = \sum_{n=0}^{\infty} q_n x^n, \quad q_0 \neq 0, \tag{47}$$

the q's satisfying the four-term recursion relation

$$n(n+1)q_{n+1} + 2anq_n + (a^2 + E)q_{n-1} + 6b(n-1)q_{n-2} = 0,$$
$$q_{-1} = q_{-2} = 0, \quad n = 0,1,2,\ldots . \tag{48}$$

Clearly, up to this point, no restriction on x needs to be imposed apart from the one necessary to ensure $y(x) \to 0$ as $|x| \to \infty$, as is done, for instance, in Ref. 1 for complex x,β with the choice $b = -\beta^{1/2}/3$. However, in the context of Eqs. (45)–(48), one must verify that $h(x)$ for $|x| \to \infty$ remains sufficiently well behaved. In other words, one has to examine the asymptotic behavior of $h(x)$ as $|x| \to \infty$. Such an investigation has been carried out in Ref. 16 and it leads to some analytic properties of the solutions y_s and y_d (Sec. I). Clearly linear combinations

$$y = C_1 y_s + C_2 y_d, \quad y_2 = C_3 y_s + C_4 y_d, \quad C_1C_4 - C_2C_3 \neq 0 \tag{49}$$

generate linearly independent solutions y_1, y_2 which themselves form a basic set. Now in the present work, we are interested in analytically examining $\alpha x^2 + \beta x^4$ further. So, instead of attempting to analytically construct physical solutions expressible as in Eq. (49), in which case we would also obtain exact formulas for the corresponding eigenvalues, we have tried the ansatz (5), whence the restriction $\mathrm{Re}(x) > 0$ follows, in order to see whether any solution different from those in Eq. (49) follows. As now Eqs. (5)–(10), and (25) imply

TABLE III. S_n as a function of n: $\lambda = 3\pi/4$.

n	S_n	n	S_n	n	S_n	n	S_n
0	1.000 000	33	−0.048 433	66	−0.014 278	99	−0.009 312
1	0.417 944	34	−0.047 160	67	−0.014 235	100	−0.009 261
2	0.005 927	35	−0.045 053	68	−0.014 143	101	−0.009 209
3	−0.264 676	36	−0.042 328	69	−0.014 007	102	−0.009 157
4	−0.422 471	37	−0.039 186	70	−0.013 835	103	−0.009 103
5	−0.494 111	38	−0.035 808	71	−0.013 632	104	−0.009 048
6	−0.503 182	39	−0.032 355	72	−0.013 406	105	−0.008 992
7	−0.469 687	40	−0.028 957	73	−0.013 162	106	−0.008 935
8	−0.409 946	41	−0.025 722	74	−0.012 907	107	−0.008 876
9	−0.336 755	42	−0.022 732	75	−0.012 646	108	−0.008 817
10	−0.259 722	43	−0.020 044	76	−0.012 384	109	−0.008 756
11	−0.185 679	44	−0.017 695	77	−0.012 125	110	−0.008 694
12	−0.119 137	45	−0.015 703	78	−0.011 872	111	−0.008 631
13	−0.062 728	46	−0.014 069	79	−0.011 629	112	−0.008 568
14	−0.017 626	47	−0.012 783	80	−0.011 398	113	−0.008 505
15	0.016 085	48	−0.011 822	81	−0.011 180	114	−0.008 441
16	0.039 087	49	−0.011 158	82	−0.010 977	115	−0.008 377
17	0.052 564	50	−0.010 756	83	−0.010 790	116	−0.008 314
18	0.057 979	51	−0.010 578	84	−0.010 618	117	−0.008 251
19	0.056 916	52	−0.010 587	85	−0.010 461	118	−0.008 189
20	0.050 942	53	−0.010 744	86	−0.010 319	119	−0.008 127
21	0.041 530	54	−0.011 012	87	−0.010 190	120	−0.008 066
22	0.029 992	55	−0.011 358	88	−0.010 075	121	−0.008 007
23	0.017 451	56	−0.011 751	89	−0.009 972	122	−0.007 948
24	0.004 824	57	−0.012 165	90	−0.009 879	123	−0.007 890
25	−0.007 175	58	−0.012 576	91	−0.009 795	130	−0.007 570
26	−0.018 024	59	−0.012 965	92	−0.009 719	140	−0.007 124
27	−0.027 375	60	−0.013 318	93	−0.009 650	150	−0.006 745
28	−0.035 032	61	−0.013 624	94	−0.009 586	160	−0.006 405
29	−0.040 921	62	−0.013 875	95	−0.009 527	170	−0.006 097
30	−0.045 074	63	−0.014 067	96	−0.009 470	180	−0.005 821
31	−0.047 596	64	−0.014 197	97	−0.009 417	190	−0.005 573
32	−0.048 650	65	−0.014 267	98	−0.009 364	200	−0.005 349

the subdominant solution, $y(x)$, thus deduced, is essentially different from the ones in Eq. (49) because it can not be expressed as $y(x) = Cy_s$, $C = \text{const.}$, as $y(x)$ is not defined at $\text{Re}(x) = 0$. Here, we note that the existence of a subdominant solution to Eq. (4) can be ascertained either by utilizing the method of Ref. 15, or Sternberg's theory[22] which is readily applicable to differential equations of arbitrary order also possessing nonpolynomial coefficients.

Owing to the nonuniform convergence of the series $\Sigma c_n s^n$ and, thus, of the series $\Sigma \exp(-sx) c_n s^n$ for $s \in [0, \infty]$, integration term-by-term in Eq. (5) is not permitted. Indeed, such an integration would yield

$$h(x) = \sum_{n=0}^{\infty} \frac{n!}{x^{n+1}} c_n, \tag{50}$$

which is easily seen to diverge for any $|x| < \infty$ due to the fact that $c_n \sim 1/(n)!^{2/3}$ for $n \gg 1$. One can argue now that Eq. (46) may be solvable in some other representation, say in the form of a continued fraction, which eventually might be transformed into a converging series proceeding in powers of $1/x$. In such a case, the relation

$$h(x) = \int_0^{\infty} \exp(-sx) f(s) ds = \sum_{n=0}^{\infty} \frac{d_n}{x^{n+1}} \tag{51}$$

would hold. Equation (51) implies that there exists some transformation which can be applied to the integral in Eq. (51), thus rendering the integration possible, so that the series in $1/x$ can be rederived. Now it can be shown[25] that $h(x)$, as given by the integral in Eq. (51), fulfills

$$|h(x)| < M/\text{Re}(x), \quad \text{Re}(x) \gg 1, \quad 0 < M < \infty. \tag{52}$$

Equation (52) shows that $h(x) \to 0$ as $\text{Re}(x) \to \infty$ more rapidly than $1/\text{Re}(x)$ as implied by the alternative representation of $h(x)$ by the series in Eq. (51). Consequently that kind of representation can not exist, and this in turn shows that the solution of this paper is *not* obtainable by introducing $t = 1/x$ in Eq. (46) and attempting to solve the resulting equation by a series expansion in t.

In a recent work[25] we have considered the $\lambda = 0$ case, that is, $\beta < 0$ and $x > 0$ in Eqs. (2)–(4), and discussed its relation to the usual $\alpha x^2 + \beta x^4$ interaction where $-\infty < x < \infty$. In the present contribution, we have constructed a subdominant solution to Eq. (4) valid in the cut β-plane, the cut being made along the positive real axis from zero to ∞. Furthermore, due to the discussion following Eq. (52), the above solution may not be accessible to a perturbation expansion proceeding in powers of $1/x$ and convergent for $\text{Re}(x) > 0$. On the other hand, the appearance of $1/\beta$ in Eq. (24) suggests that $y(x)$ defined by Eqs. (5)–(10) may not be derivable by a perturbation procedure which is based on the αx^2 interaction and which, of course, in the limit $\beta \to 0$, yields a solution of the latter. It is perhaps interesting to observe that, although there is no obvious reason to expect it, the nonperturbative character of the subdominant solution persists also in the context of the cut-off theory considered in the present work.

TABLE IV. S_n as a function of n: $\lambda = \pi$.

n	S_n	n	S_n	n	S_n	n	S_n
0	1.000 000	33	− 0.089 738	66	− 0.020 169	99	− 0.009 084
1	0.454 460	34	− 0.085 193	67	− 0.020 024	100	− 0.009 186
2	0.029 078	35	− 0.078 615	68	− 0.019 657	101	− 0.009 263
3	− 0.276 129	36	− 0.070 550	69	− 0.019 101	102	− 0.009 316
4	− 0.471 356	37	− 0.061 523	70	− 0.018 393	103	− 0.009 344
5	− 0.572 608	38	− 0.052 018	71	− 0.017 566	104	− 0.009 347
6	− 0.598 471	39	− 0.042 465	72	− 0.016 657	105	− 0.009 327
7	− 0.567 843	40	− 0.033 230	73	− 0.015 696	106	− 0.009 285
8	− 0.498 421	41	− 0.024 612	74	− 0.014 716	107	− 0.009 223
9	− 0.405 778	42	− 0.016 841	75	− 0.013 741	108	− 0.009 143
10	− 0.302 885	43	− 0.010 082	76	− 0.012 797	109	− 0.009 048
11	− 0.199 967	44	− 0.004 436	77	− 0.011 902	110	− 0.008 940
12	− 0.104 575	45	0.000 051	78	− 0.011 073	111	− 0.008 822
13	− 0.021 822	46	0.003 384	79	− 0.010 322	112	− 0.008 696
14	0.045 297	47	0.005 608	80	− 0.009 657	113	− 0.008 564
15	0.095 544	48	0.006 802	81	− 0.009 085	114	− 0.008 430
16	0.129 064	49	0.007 071	82	− 0.008 606	115	− 0.008 295
17	0.147 034	50	0.006 537	83	− 0.008 221	116	− 0.008 161
18	0.151 336	51	0.005 333	84	− 0.007 926	117	− 0.008 030
19	0.144 280	52	0.003 596	85	− 0.007 716	118	− 0.007 903
20	0.128 366	53	0.001 463	86	− 0.007 585	119	− 0.007 782
21	0.106 101	54	− 0.000 936	87	− 0.007 525	120	− 0.007 668
22	0.079 850	55	− 0.003 480	88	− 0.007 527	121	− 0.007 561
23	0.051 744	56	− 0.006 057	89	− 0.007 580	122	− 0.007 462
24	0.023 607	57	− 0.008 574	90	− 0.007 677	123	− 0.007 371
25	− 0.030 076	58	− 0.010 951	91	− 0.007 807	130	− 0.006 993
26	− 0.027 162	59	− 0.013 121	92	− 0.007 962	140	− 0.006 755
27	− 0.047 848	60	− 0.015 037	93	− 0.008 132	150	− 0.006 527
28	− 0.064 639	61	− 0.016 664	94	− 0.008 309	160	− 0.006 168
29	− 0.077 315	62	− 0.017 983	95	− 0.008 486	170	− 0.005 818
30	− 0.085 891	63	− 0.018 984	96	− 0.008 657	180	− 0.005 551
31	− 0.090 571	64	− 0.019 673	97	− 0.008 817	190	− 0.005 336
32	− 0.091 702	65	− 0.020 060	98	− 0.008 960	200	− 0.005 131

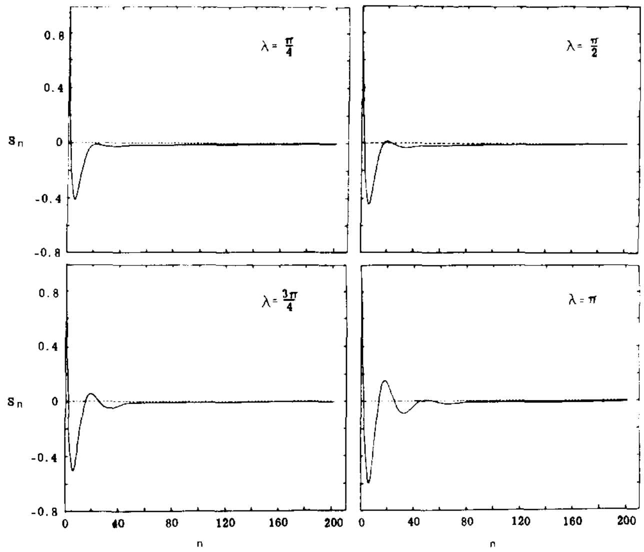


FIG. 1. S_n as a function of n.

Finally, we point out that it might be of some interest to further investigate the role of cut-off distances in problems involving a wider class of potentials, bearing also in mind possible applications to vibrational spectra, with the aim to see whether or not many features of the solutions of this paper apply to more general potentials.

ACKNOWLEDGMENTS

We are most indebted to J. Rosenberg for providing us with computer facilities. One of us (GPF) wishes to thank I. Murphy for a very helpful discussion on the applicability of the criterion in Eq. (14).

APPENDIX

We have investigated the sum S_n defined in Eq. (18) numerically. The results pertaining to four different λ-values are presented in Tables I–IV and Fig. 1. It is perhaps worth pointing out that, although not of relevance to this paper, very similar results are obtained for S_n in the case $\pi \leqslant \lambda < 2\pi$ and clearly for $\lambda' = \lambda + 12K'\pi, K' = 0,1,2,\ldots$; for $\lambda = 2\pi$, S_n diverges as $n \to \infty$. From Fig. 1 it is seen that $|S_n|$ remains always smaller than 1, $0 \leqslant \lambda \leqslant \pi$, and tends slowly to zero for $n \to \infty$. The slow rate at which $|S_n| \to 0$ as $n \to \infty$ is predicted by the theoretical treatment in the Lemma; in fact, the dominant behavior of $|S_n|$ for $n \gg 1$ is determined by $|I_1|$ in Eq. (35) since $|I_2|, |I_3|, |I_4|$, as shown by Eqs. (40) and (44), approach zero much more rapidly and, therefore, $|S_n| \to 0$ as $n \to \infty$ apparently like $n^{-1/4}$.

Further we note that, most probably, Eq. (19) cannot be verified by application of theorems from the general theory of series because the terms in S_n are *themselves* n-dependent and, thus, those theorems become inapplicable. The n-dependence of the terms of S_n is the reason why we had to consider an n-dependent contour of integration in the lemma.

[1]C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[2]G. E. Sobelman, Phys. Rev. D **19**, 3754 (1979).
[3]B. Simon, Ann. Phys. (N.Y.) **58**, 76 (1970).
[4]J. J. Loeffel, A. Martin, B. Simon, and A. S. Wightman, Phys. Lett. B **30**, 656 (1969).
[5]S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, J. Math. Phys. **14**, 1190 (1973).
[6]F. T. Hioe, D. Macmillan, and E. W. Montroll, Phys. Rep. C **43**, 305 (1978).
[7]M. Lakshmanan and J. Prabhakaran, Lett. Nuovo Cimento **7**, 689 (1973).
[8]P. Lu, S. S. Wald, and B. L. Young, Phys. Rev. D **7**, 1701 (1973).
[9]K. Datta and A. Rampal, Phys. Rev. D **23**, 2875 (1981).
[10]G. P. Flessas and K. P. Das, Phys. Lett. A **78**, 19 (1980).
[11]G. P. Flessas, Phys. Lett. A **81**, 17 (1981); J. Phys. A: Math. Gen. **14**, L209 (1981).
[12]E. Magyari, Phys. Lett. A **81**, 116 (1981).
[13]M. Znojil, J. Phys. A: Math. Gen. **15**, 2111 (1982).
[14]A. Khare, Phys. Lett. A **83**, 237 (1981).
[15]P. F. Hsieh and Y. Sibuya, J. Math. Anal. Appl. **16**, 84 (1966).
[16]G. P. Flessas, Phys. Lett. A **95**, 361 (1983).
[17]I. G. Halliday and P. Surenyi, Phys. Rev. D **21**, 1529 (1980).
[18]O. Haan and K. H. Mütte, Nucl. Phys. B **155**, 46 (1979).
[19]P. M. Mathews, M. Seetharaman, S. Raghavan, and V. T. A. Bhargava, Phys. Lett. A **83**, 112 (1981).
[20]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1956), pp. 530–534.
[21]Equation (11) implies that $f(s)$ for $s \gg 1$ can behave, at the most, as some power of, say s^N, $N =$ finite. In such a case, the relation $|s^M f(s) \exp(-sx)| \leqslant A, A > 0$, and $M > 1$ is satisfied for arbitrarily large s, which in turn ensures the absolute convergence of the integral in Eq. (5).
[22]A numerical investigation confirms Eq. (23) from which, by applying a powerful theorem [W. Sternberg, Mathematische Annalen **81**, 119 (1920)] on the asymptotic integration of differential equations, we can show that, in fact, $f(s) \to 0$ for $s \to \infty$ and b given by Eq. (24). In this paper, we have preferred to use the argument following Eq. (14), since it may be taken as representing a case which does not seem to be covered by a theorem (Ref. 20, p. 387) concerning the comparison of two power series of the same argument z for $|z| \to \infty$. Along the same lines, it can be seen that the second solution of Eq. (7) has to be to discard d since it makes the integral in Eq. (5) divergent at $s = \infty$.
[23]Reference 20, p. 396.
[24]This can be seen from Lemma 2.1 by taking, for Σ_2 and Σ_4, a contour encircling the points $1,\ldots,n$ with x varying between b and $n + b$ where $1/3 < b < 1$.
[25]G. P. Flessas, R. R. Whitehead, and A. Rigas, J. Phys. A: Math. Gen. **16**, 85 (1983). The case $\pi/2 < \lambda < \pi$ can be dealt with as in the lemma; see also the Appendix of the present work.
[26]Reference 20, p. 413.
[27]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1956), p. 256.
[28]Reference 20, p. 476.
[29]H. Jeffreys and B. B. Jeffreys, *Methods of Mathematical Physics* (Cambridge U. P., Cambridge, England, 1956), p. 467.

Infinitely many commuting symmetries and constants of motion in involution for explicitly time-dependent evolution equations

W. Oevel
Universität Paderborn, D 4790 Paderborn, Federal Republic of Germany
A. S. Fokas
Department of Mathematics, Clarkson College, Potsdam, New York 13676

(Received 10 August 1982; accepted for publication 1 April 1983)

The existing symmetry approach to exactly solvable time-evolution equations in one spatial dimension is extended to the case that the equation depends explicitly on the time and space coordinates. In this context (commuting) symmetries, constants of motion (in involution), strong and hereditary symmetries (squared-eigenfunction operators), and Hamiltonian structures are discussed. The cylindrical Korteweg–de Vries equation is used as an illustrative example.

PACS numbers: 03.65. – w

1. INTRODUCTION

One of the important aspects associated with a given evolution equation (i.e., the flux of a vector field on a manifold) is the investigation of its "algebraic properties." This includes the evaluation of the symmetry group of this equation, i.e., one looks for a Lie group acting on the manifold in such a way that solutions of the dynamical system are mapped to solutions. Instead of giving this group action explictly, one rather tries to calculate the corresponding Lie algebra or, more specifically, the infinitesimal generators of one-dimensional subgroups. These generators can be characterized easily in terms of the usual commutator on the Lie algebra of vector fields.

Further information about the given equation can be obtained by looking for conserved quantities, i.e., for realvalued functions on the manifold which are constant along the trajectories of the given dynamical system. If the manifold is symplectic and if the system admits a Hamiltonian formulation, then the well-known classical Noether theorem establishes a correspondence between symmetries and conservation laws. Such a system is called completely integrable if there are "enough" conserved quantities in involution, which means that the corresponding symmetries commute.

Let u be a smooth function on the real line vanishing rapidly at infinity, and let $K(u)$ be a smooth vector field on the space S of these functions. For completely integrable equations of the type $u_t = K(u)$, it has turned out that certain integro-differential operators are of prime importance for analyzing the algebraic properties. These operators have been given various names in the literature: squared eigenfunction operators,[1] recursion operators,[2] strong symmetries,[3] hereditary symmetries,[3] Kähler operators,[4] or regular operators.[5] Several investigators, in particular Magri,[4] Gel'fand and Dorfman,[5] and Fokas and Fuchssteiner[6] have extensively studied the structure of these operators and their connection to the Hamiltonian formulation. Other relevant works include Refs. 7–15. It turns out that sets of infinitely many commuting symmetries and constants of motion in involution are easily characterized within this framework.

The aim of this paper is to extend the above studies to equations of the form

$$u_t = K(u,x,t). \tag{1}$$

Notation

We will consider evolution equations of type (1), where u is an element of the space S of functions on the real line, vanishing rapidly for $|x|\to\infty$. Let K be some differentiable map on this space, depending explicitly on the space and time variables x and t.

We define the directional derivative of a function φ on S by

$$\varphi'(u)[v] := \frac{\partial}{\partial\epsilon}\varphi(u+\epsilon v)|_{\epsilon=0}. \tag{2}$$

Given two explicitly space-time-dependent functions $\gamma(u,x,t)$ and $\sigma(u,x,t)$ on S, decaying sufficiently fast for $|x|\to\infty$, we define a scalar product:

$$\langle\gamma,\sigma\rangle := \int_{-\infty}^{\infty}\gamma(u(x,t),x,t)\sigma(u(x,t),x,t)\,dx. \tag{3}$$

If $p(u,x,t)$ is a differentiable real-valued function, then its gradient w.r.t. u is defined by

$$\langle \operatorname{grad} p, f\rangle = \frac{\partial}{\partial\epsilon}p(u+\epsilon f,x,t)|_{\epsilon=0}, \tag{4}$$

where f is any element of S. We call p the potential of grad p. A function $\gamma(u,x,t)$ is a gradient, iff $\gamma'^* = \gamma'$, where * means transposition w.r.t. the scalar product $\langle\cdot,\cdot\rangle$. Its potential then is given by

$$p(u) = \int_0^1 \langle\gamma(\lambda u),u\rangle\,d\lambda. \tag{5}$$

2. SYMMETRIES, CONSERVED QUANTITIES, STRONG SYMMETRIES AND GENERALIZED NOETHER THEOREM

The notions of symmetries, conserved covariants, strong symmetries (or recursion operators[2]), and Noether operators for equations of the form $u_t = K(u)$ have been well established, see, for example, Ref. 6.

Rather than recalling their mathematical definitions, let us illustrate these notions using the Korteweg–de Vries (KdV) equation as an example:

$$u_t + u_{xxx} + uu_x = 0, \tag{6}$$

where the subscripts denote differentiation w.r.t. the corresponding variable. Equation (6) remains invariant under the

 0022-2488/84/040918-05$02.50

following four-parameter Lie-point group of transformations:

$$x' = e^{\delta}(x + \alpha - \gamma t), \quad t' = e^{3\delta}(t + \beta), \quad u' = e^{2\delta}(u + \gamma). \quad (7)$$

The above transformations (space and time translations, Galilean and scaling transformations) are uniquely characterized by the following *infinitesimal generators of symmetries*[16]:

$$\sigma_1 = u_x, \quad \sigma_2 = u_{xxx} + uu_x, \quad \Sigma_1 = 1 - tu_x,$$
$$\Sigma_2 = 2u + xu_x - 3t(u_{xxx} + uu_x). \quad (8)$$

We now establish a Hamiltonian formulation for the KdV. Let us recall that an antisymmetric operator $J(u)$ is called symplectic, if the 2-form $\omega(u)$ defined by

$$\omega(u)(\cdot,\cdot) = \langle J(u)\cdot,\cdot\rangle \quad (9)$$

is closed. This means that the bracket

$$\{a,b,c\} := \langle J'(u)[a]b,c\rangle \quad (10)$$

satisfies the Jacobi identity for arbitrary functions a, b, and c in S.

An antisymmetric operator $\Theta(u)$ is called implectic (inverse-symplectic), if it has the same properties as the inverse of a symplectic operator, i.e., if the bracket

$$\{\{a,b,c\}\} := \langle b,\Theta'(u)[\Theta(u)a]c\rangle \quad (11)$$

satisfies the Jacobi identity for all functions a, b, and c in S.

Such operators J and Θ obviously correspond to the symplectic forms of classical mechanics which give rise to the definition of Hamiltonian systems. Analogously, a vector field $K(u)$ on S is called *Hamiltonian*, if one can write either

$$J(u)K(u) = \operatorname{grad} p(u) \quad (12)$$

or

$$K(u) = \Theta(u) \operatorname{grad} \tilde{p}(u) \quad (13)$$

with J symplectic, Θ implectic, and p and $\tilde{p}$ some functions on S. The well-known Noether theorem of classical mechanics goes through for (12) and (13): If $q(u)$ is a conservation law for K, then $\Theta(u)$ grad $q(u)$ is the generator of a symmetry for K. Conversely, if $\sigma(u)$ generates a symmetry of K, then $\gamma(u) = J(u)K(u)$ is a *conserved covariant*. The notion of a conserved covariant is a mathematical generalization of the gradient of a conserved quantity. Namely, if the functional $p(u)$ is conserved w.r.t. a given evolution equation, then $\gamma = \operatorname{grad} p$ is a conserved covariant. Conversely, if γ is a conserved covariant *and* if γ is a gradient function, then its potential $p(u)$ is a conserved quantity. We note that there exist conserved covariants which are not gradients and hence do not give rise to conserved quantities of the usual type.

An operator $\Theta(u)$ which maps conserved covariants of K to symmetries of K will be called a *Noether operator* for K. Conversely, an operator $J(u)$ mapping symmetries to conserved covariants will be called an *inverse Noether operator*.

For the KdV we find

$$u_t = -u_{xxx} - uu_x = D \operatorname{grad} \int_{-\infty}^{\infty} \left[\frac{1}{2}(u_x)^2 - \frac{u^3}{6}\right] dx, \quad (14)$$

where $D = \partial/\partial x$ is the differential operator w.r.t. the space variable. Clearly D is implectic and its inverse D^{-1} is symplectic so that they are Noether and inverse Noether operators, respectively, for the KdV. Throughout this paper we understand that

$$(D^{-1}f)(x) = \int_{-\infty}^{x} f(\xi)\, d\xi \quad (15)$$

if the function f vanishes fast enough as x goes to $-\infty$ and

$$(D^{-1}f)(x) = \int_0^x f(\xi)\, d\xi \quad (16)$$

if f does not obey this boundary condition.

Applying D^{-1} to the symmetries (8), we find the following conserved covariants:

$$\gamma_1 = u, \quad \gamma_2 = u_{xx} + u^2/2, \quad \Gamma_1 = x - tu,$$
$$\Gamma_2 = D^{-1}u + xu - 3t(u_{xx} + u^2/2). \quad (17)$$

It is clear that γ_1, γ_2, and Γ_1 are gradient functions with the potentials

$$p_1 = \int_{-\infty}^{\infty} \frac{u^2}{2} dx, \quad p_2 = \int_{-\infty}^{\infty} \left[\frac{u^3}{6} - \frac{1}{2}(u_x)^2\right] dx,$$
$$\mathrm{P}_1 = \int_{-\infty}^{\infty} \left(xu - \frac{tu^2}{2}\right) dx. \quad (18)$$

However, $\Gamma_2' \neq (\Gamma_2')^*$ and hence the scaling symmetry does not give rise to a constant of the motion.

Since the KdV is an "exactly solvable" evolution equation, it possesses infinitely many symmetries. A *strong symmetry* is an operator $\phi(u)$ which maps a symmetry to a symmetry and hence, in principle, generates infinitely many symmetries from a given one.

For the KdV the Lenard operator[2]

$$\Phi(u) = D^2 + \tfrac{2}{3}u + \tfrac{1}{3}u_x D^{-1} \quad (19)$$

is such a strong symmetry.

Note that the equations

$$u_t = (\Phi(u))^n u_x, \quad n = 1,2,3,\cdots, \quad (20)$$

define the Lax hierarchy associated to the KdV.

Starting with Σ_2, we find a second hierarchy $(\Phi(u))^n \Sigma_2$ of symmetries for the KdV which are nonlocal and do not give rise to conservation laws.

All the above notions can be easily defined for equations of the form (1):

Definition 1: (i) A function $\sigma(u,x,t)$ is a *symmetry* of (1) iff

$$\frac{\partial \sigma}{\partial t} + \sigma'[K] - K'[\sigma] = 0. \quad (21)$$

(ii) A function $\gamma(u,x,t)$ is a *conserved covariant* of (1) iff

$$\frac{\partial \gamma}{\partial t} + \gamma'[K] + K^*[\gamma] = 0. \quad (22)$$

(iii) An operator valued function $\Theta(u,x,t)$ is called a *Noether operator* of (1) iff

$$\frac{\partial \Theta}{\partial t} + \Theta'[K] - \Theta K'^* - K'\Theta = 0. \quad (23)$$

(iv) An operator valued function $\Phi(u,x,t)$ is called a *strong symmetry* of (1) iff

$$\frac{\partial\phi}{\partial t} + \phi'[K] + \Phi K' - K'\Phi = 0. \tag{24}$$

Equations (21)–(24) must hold identically in the variables u, x, and t; $\partial/\partial t$ denotes the derivative w.r.t. the explicit time dependence.

Remark 1: (i) From the above definitions it follows by elementary calculus that Φ maps symmetries to symmetries, Φ^* maps conserved covariants to conserved covariants, and Θ maps conserved covariants to symmetries.

(ii) One may define an operator valued function $J(u,x,t)$ as an *inverse Noether operator* of (1) if

$$\frac{\partial J}{\partial t} J'[K] + JK' + K'^*J = 0. \tag{25}$$

Then J maps symmetries to conserved covariants.

(iii) If Θ and J are Noether and inverse Noether operators for (1), then $\Phi = \Theta J$ is a strong symmetry of (1). Furthermore, the hierarchy $\{\phi^n\Theta\}$ is a hierarchy of Noether operators for (1).

It is well known that if an equation is in a Hamiltonian form, then there exists a relationship between its symmetries and the constants of the motion. If Eq. (1) admits a Hamiltonian formulation, then a generalization of Noether's theorem is given by the following theorem:

Theorem 1: Assume that Eq. (1) can be written in the form

$$u_t = \Theta\mu, \tag{26}$$

where (i) $\Theta(u,x,t)$ is an implectic operator[17] and (ii) the operator $\Theta(u,x,t)$ and the function $\mu(u,x,t)$ satisfy

$$\frac{\partial\Theta}{\partial t} - \Theta(\mu' - \mu'^*)\Theta = 0. \tag{27}$$

There Θ is a Noether operator for Eq. (1).

Proof: From $K = \Theta\mu$ we derive $K' = \Theta'[\cdot]\mu + \Theta\mu'$. Using the Jacobi identity for implectic operators, we find

$$\langle b,(\Theta_t + \Theta'[K] - \Theta K'^* - K'\Theta)c\rangle$$
$$= \langle b,(\Theta_t - \Theta(\mu' - \mu'^*)\Theta)c\rangle \tag{28}$$

for arbitrary b and c in S. Hence (27) implies (23) and vice versa.

Remark 2: One may easily obtain a similar result for inverse Noether operators. Namely, if $JK = \tilde{\mu}$, where J is a symplectic operator and where J and the function $\tilde{\mu}$ satisfy

$$\frac{\partial J}{\partial t} + \tilde{\mu}' - \tilde{\mu}'^* = 0, \tag{29}$$

then (1) admits J as an inverse Noether operator.

Example: Consider the cylindrical KdV (c-KdV) equation[18]:

$$u_t + u_{xxx} + uu_x + u/2t = 0. \tag{30}$$

The inverse scattering transform (IST) method for (30) is given in Ref. 19, and the conservation laws are given in Ref. 20. Here we use this equation as an illustrative example and prove that it possesses a set of infinitely many commuting symmetries and a set of infinitely many constants of motion in involution.

Bountis[21] has shown that (30) admits a four-parameter group of Lie-point transformations which are characterized by the following generators of symmetries:

$$K_0 = -3u_x, \quad G_0 = 2\sqrt{t}\,(u_x - 1/2t),$$
$$K_1 = 2u + xu_x + 3tu_t, \tag{31}$$
$$G_1 = -2\sqrt{t^3}u_t - \sqrt{t}\,xu_x - 2\sqrt{t}\,u + x/2\sqrt{t}.$$

The strong symmetry of (30) is given by the operator (see Sec. 4)

$$\Phi(u,x,t) = t(D^2 + \tfrac{2}{3}u + \tfrac{1}{3}u_x D^{-1}) - \tfrac{1}{3}x - \tfrac{1}{6}D^{-1}. \tag{32}$$

The operator

$$\Theta_1 = (1/t)D \tag{33}$$

is implectic, and (30) can be written as

$$u_t = \Theta_1\mu \tag{34}$$

with

$$\mu = -t(u_{xx} + u^2/2) - \tfrac{1}{2}D^{-1}u. \tag{35}$$

Verifying that (27) is satisfied with Θ_1 and μ defined by (33) and (35), respectively, we see that Θ_1 is a Noether operator for the c-KdV. Hence $\Theta_2 = \phi\Theta_1$ is a second Noether operator and Θ_1^{-1} is an inverse Noether operator. One checks that Θ_2 is again implectic.

We find a first hierarchy of symmetries using K_0 as starting point

$$K_n = \Phi^n K_0, \tag{36}$$

to which corresponds a hierarchy of conserved covariants

$$\gamma_n = \Theta_1^{-1}K_n. \tag{37}$$

However, neither the K_n's commute nor are the above conserved covariants gradients. It will turn out that one needs to use the second hierarchy of symmetries

$$G_n = \Phi^n G_0 \tag{38}$$

to obtain commuting symmetries as well as constants of motion in involution. However, in order to prove these results, we need the notion of factorizable hereditary symmetries.[6]

3. FACTORIZABLE HEREDITARY OPERATORS AND INVOLUTIONARY PROPERTY

We recall that an operator valued function $\Phi(u,x,t)$ is called *hereditary*[3] iff it satisfies

$$\phi'[\phi a]b - \phi\phi'[a]b = \phi'[\phi b]a - \phi\phi'[b]a \tag{39}$$

for arbitrary elements a and b in S. This is equivalent[22] to

$$\phi^2[|f,g|] + [|\phi f,\phi g|] = \phi([|\phi f,g|] + [|f,\phi g|]) \tag{40}$$

for arbitrary functions $f(u)$ and $g(u)$ on S, where $[|\ ,\ |]$ denotes the commutator

$$[|f,g|] = f'[g] - g'[f]. \tag{41}$$

The properties of hereditary operators and their close relationship to symplectic/implectic operators have been throughly investigated in Refs. 3, 5, and 22. Their main use results from the following two fundamental properties:

(i) Hereditary operators generate hierarchies of *commuting* symmetries.

(ii) Hereditary operators which are factorizable in terms

of symplectic/implectic operators generate conserved covariants which are gradients and in involution.
These two properties can also be used in connection with evolution equations of type (1).

Theorem 2: Let $\phi(u,x,t)$ be a hereditary operator and $G_0(u,x,t)$ a function such that

$$\phi'[G_0] + \phi G_0' - G_0'\phi = 0. \tag{42}$$

Then: (i) All the flows $G_n = \phi^n G_0$ commute, i.e.,

$$[|G_n,G_m|] = 0 \quad \forall n,m\in\mathbb{N}\cup\{0\}. \tag{43}$$

(ii) If ϕ admits a symplectic/implectic factorization, i.e., if ϕ can be written as $\Phi = \Theta J$, where Θ and J form a *compatible* (see Ref. 6) pair of implectic and symplectic operators, then all γ_n's defined by

$$\gamma_n := JG_n, \quad n\in\mathbb{N}\cup\{0\}, \tag{44}$$

are gradients, if γ_0 is a gradient. Furthermore, all γ_n's are in involution w.r.t. the Poisson bracket

$$\{\gamma_n,\gamma_m\} := \langle\gamma_n,\Theta\gamma_m\rangle. \tag{45}$$

Proof: (i) An equivalent formulation for (42) is

$$\phi[|G_0,f|] = [|G_0,\phi f|], \tag{46}$$

$f(u)$ being an arbitrary function on S. We say that ϕ commutes with G_0, if (46) is satisfied. From (40) we see that if a hereditary ϕ commutes with G_0, then it also commutes with ϕG_0, $\phi^2 G_0$, etc. So we immediately have

$$[|G_n,G_m|] = [|\phi^n G_0,\phi^m G_0]| = \phi^{n+m}[|G_0,G_0|] = 0 \tag{47}$$

for each n,m.

(ii) We first consider an arbitrary function $G(u)$ on S. Let $\gamma = JG$ be a gradient. From $\gamma' = \gamma'^*$ and using the Jacobi identity for symplectic operators, we find

$$\langle (J'[G] + JG' + G'^*J)a,b\rangle = 0 \tag{48}$$

for all elements a and b in S. Hence JG is a gradient iff

$$J'[G] + JG' + G'^*J = 0. \tag{49}$$

If ϕ commutes with G, i.e., (42) holds, we calculate

$$(J\phi)'[G] + J\phi G' + G'^*J\phi = (J'[G] + JG' + G'^*J)\phi = 0. \tag{50}$$

So $J\phi G$ is another gradient, provided $J\phi$ is symplectic. What compatibility condition have J and ϕ to satisfy such that J, $J\phi$, $J\phi^2$,$\cdots$ are symplectic operators? Let us assume that ϕ factorizes in the form $\phi = \Theta J$ with Θ implectic and J symplectic and invertible. We say that Θ and J are *compatible*,[6] if $J^{-1} + \Theta$ is implectic, which is the same as saying that $J^{-1} - \lambda\Theta$ is implectic for all $\lambda\in\mathbb{R}$. Assuming that $J^{-1} - \lambda\Theta$ is invertible for all λ in a neighborhood of zero, we conclude that $(J^{-1} - \lambda\Theta)^{-1} = J(1-\lambda\Theta J)^{-1}$ is symplectic. As the symplectic operators form a vector space, we differentiate this operator w.r.t. λ and find that

$$\frac{d^n}{d\lambda^n} J(1-\lambda\Theta J)^{-1}|_{\lambda=0} = n!\, J\phi^n \tag{51}$$

is symplectic for any $n\in\mathbb{N}$. So if we generate a hierarchy $G_n = \phi^n G_0$ by means of a hereditary operator $\phi = \Theta J$, which factorizes into a compatible pair of implectic and symplectic operators, then (50) tells us by induction that $J\phi^n G_0$ is closed for any n, if JG_0 is closed.

The involutionary property of the set $\{\gamma_n = JG_n\}$ follows easily. Note that we have

$$\Phi^n\Theta = \Theta(\Phi^*)^n, \tag{52}$$

and

$$\gamma_n = J\phi^n G_0 = (\phi^*)^n\gamma_0, \tag{53}$$

so we find for $n > m$

$$\langle\gamma_n,\Theta\gamma_m\rangle = \langle(\phi^*)^{n-m}\gamma_m,\Theta\gamma_m\rangle = \langle\gamma_m,\Theta\gamma_n\rangle = -\langle\gamma_n,\Theta\gamma_m\rangle = 0. \tag{54}$$

Example: Let us return to the c-KdV and consider the hierarchy of symmetries generated by G_0:

$$G_n(u) := \phi^n(u)G_0(u), \quad G_0(u) = 2\sqrt{t}\,(u_x - 1/2t). \tag{55}$$

Using $J = \Theta_1^{-1} = tD^{-1}$, we obtain the following hierarchy of conserved covariants:

$$\begin{aligned}
\gamma_0 &= 2\sqrt{t^3}u - \sqrt{t}\,x\\
\gamma_1 &= 2\sqrt{t^5}(u_{xx} + u^2/2) - \sqrt{t^3}xu + \sqrt{t}\,x^2/4\\
\gamma_n &:= tD^{-1}G_n
\end{aligned} \tag{56}$$

Note that $\gamma_{-1} := 6\sqrt{t}$ is another conserved covariant belonging to this second hierarchy. It is easily verified that Φ and G_0 satisfy (42) and that γ_0 is a gradient function. One checks that Φ is hereditary (see also Sec. 4) and factorizable ($\phi = \Theta_2\Theta_1^{-1}$) into a compatible pair of implectic and symplectic operators and hence we conclude:

(i) All G_n's commute.

(ii) All γ_n's are gradients and in involution w.r.t. the Poisson bracket

$$\{\gamma_n,\gamma_m\}_1 = \langle\gamma_n,\Theta_2\gamma_m\rangle. \tag{57}$$

Remark 3: As $\Theta_2 = \Theta_1\phi^*$, we see that we can as well use the Poisson bracket

$$\{\gamma_n,\gamma_m\}_2 := \langle\gamma_n,\Theta_1\gamma_m\rangle = \{\gamma_n,\gamma_{m-1}\}_1. \tag{58}$$

4. THE CONNECTION TO INVERSE SCATTERING

From the above discussion it is evident that the operator $\Phi(u,x,t)$ plays a fundamental role for the investigation of the algebraic structure of an equation. On the other hand, the existence of an isospectral eigenvalue problem is essential for solving an equation via the IST. The connection between these two aspects are given in Refs. 23 and 24, where it is shown that knowledge of the eigenvalue problem yields algorithmically a strong symmetry for the whole hierarchy of equations solvable by this eigenvalue problem. Furthermore, this strong symmetry is also hereditary. The extension of these results to equations of type (1) is straightforward.

Proposition: Assume there is given an isospectral eigenvalue problem for the evolution equation (1), i.e., we have an operator $L(u,x,t)$ depending on u and the space—and time—coordinates, such that the spectrum of L remains invariant, when u evolves in time according to (1). Then the eigenvalues λ, which depend on u, are conserved quantities for (1). By G_λ we denote the gradient of the eigenvalues w.r.t. u, and we try to evaluate this gradient in terms of the eigenfunction Ψ

given by

$$L\Psi = \lambda \Psi. \tag{59}$$

If one can transform (59) into an eigenvalue equation satisfied by G_λ, say

$$\Delta G_\lambda = f(\lambda) G_\lambda, \tag{60}$$

where $\Delta(u,x,t)$ is an operator acting on G_λ and f is a function, then the operator Δ^* is a strong and hereditary symmetry for Eq. (1).

Example: The isospectral eigenvalue problem associated with the c-KdV equation is given by[20]

$$(tD^2 - x/12 + \tfrac{1}{6}tu)\Psi = \lambda \Psi. \tag{61}$$

Taking the directional derivative of (61) and forming the scalar product with Ψ, it follows that

$$G_\lambda := \operatorname{grad} \lambda = t\Psi^2. \tag{62}$$

If Ψ satisfies Eq. (61), one easily finds the equation satisfied by $t\Psi^2$, namely

$$(tD^2 + \tfrac{2}{3}tu - \tfrac{1}{3}tD^{-1}u_x - \tfrac{1}{3}x + \tfrac{1}{6}D^{-1})(t\Psi^2) = 4\lambda\,(t\Psi^2). \tag{63}$$

Hence one directly obtains the hereditary symmetry of the c-KdV given by (32).

ACKNOWLEDGMENTS

The authors thank T. Bountis and B. Fuchssteiner for many valuable discussions.

[1] M. J. Ablowitz, D. J. Kaup, A. C. Newell, and H. Segur, Stud. Appl. Math. **53**, 249 (1974).
[2] P. J. Olver, J. Math. Phys. **18**, 1212 (1977).
[3] B. Fuchssteiner, Nonlin. Anal. Theory, Methods & Applications **3**, 849 (1979).
[4] F. Magri, J. Math. Phys. **19**, 1156 (1978): "A geometrical approach to the nonlinear solvable equations," *Lecture Notes in Physics*, Vol. 120, edited by M. Boiti, F. Pimpinelli, and G. Soliani (Springer-Verlag, Berlin, 1980), p. 233.
[5] I. M. Gel'fand and I. Ya Dorfman, Funct. Anal. Appl. **13**, 248 (1979); **14** (1980).
[6] (a) A. S. Fokas and B. Fuchssteiner, Lett. Nuovo Cimento **28**, 299 (1980); (b) B. Fuchssteiner and A. S. Fokas, Physica **D4**, 47 (1981).
[7] I. M. Gel'fand, L. A. Dikii, Funct. Anal. Appl. **10**, 259 (1976).
[8] P. J. Olver, Math. Proc. Camb. Phil. Soc. **88**, 71 (1980).
[9] W. Symes, J. Math. Phys. **20**(4), 721 (1979).
[10] A. P. Fordy and J. Gibbons, J. Math. Phys. **22**, 1170 (1981); Commun. Math. Phys. **77**, 21 (1980).
[11] B. A. Kupershmidt and G. Wilson, Invent. Math. **62**, 403 (1981); Commun. Math. Phys. **81**, 189 (1981).
[12] Y. Kosmann-Schwarzbach, "Vector fields and Generalized Vector fields on fibered manifolds," Lect. Notes Math. **792**, 307 (1980); Lett. Math. Phys. **5**, 229 (1981).
[13] G. Z. Tu. Commun. Math. Phys. **77**, 289 (1980).
[14] B. G. Konopelchenko, Phys. Lett. A **75**, 447 (1980); A **79**, 39 (1980); B **95**, 83 (1980); B **100**, 254 (1981).
[15] F. Calogero and A. Degasperis, Nuovo Cimento B **32**, 201 (1976); B **39**, 1 (1977).
[16] A. S. Fokas, J. Math. Phys. **21**(6), 1318 (1980).
[17] We stick to definitions (10) and (11), even if the operators depend explictly on the time-variable.
[18] S. Maxon and J. Viecelli, Phys. Fluids **17**, 1614 (1974).
[19] F. Calogero and A. Degasperis, Lett. Nuovo Cimento **23**, 150 (1978).
[20] F. Calogero and A. Degasperis, Lett. Nuovo Cimento **23**, 155 (1978).
[21] T. Bountis, private communication.
[22] B. Fuchssteiner, Prog. Theoret. Phys. **65**(3), (1981).
[23] B. Fuchssteiner, "Application of Spectral-Gradient Methods to Nonlinear Soliton Equations," preprint, 1980.
[24] A. S. Fokas and R. L. Anderson, "On the use of isospectral eigenvalue problems for obtaining hereditary symmetries for Hamiltonian systems," J. Math. Phys. **23**(6), 1066 (1982).

Geometry of the dynamical systems with time-dependent constraints and time-dependent Hamiltonian: An approach towards quantization

Adnan Hamoui[a)] and André Lichnerowicz
Collège de France, Physique Mathématique, 75231 Paris Cedex 05, France

(Received 5 April 1983; accepted for publication 28 October 1983)

The quantization of a dynamical system with time-dependent constraints or time-dependent Hamiltonians presents particular difficulties not only from a technical, but also from a theoretical point of view. We give here an approach to the evolution of such a quantum system in terms of star products on symplectic or canonical manifolds. These star products are particular deformations of the usual product of functions, deformations that start with the Poisson bracket. The introduction of the star products corresponds to a generalization of the Weyl–Wigner quantization on a flat phase space. This approach leads to the introduction of "two times" denoted t and τ, playing different roles; the time t plays the role of a geometrical time corresponding to the control, while the other, τ, which parametrizes the evolution, plays the role of a dynamical time. The classical as well as the quantum results point out the role played by the variable $(t + \tau)$. This reflects the coherence of the formalism.

PACS numbers: 03.65.Ca, 02.40. + m, 03.20. + i

INTRODUCTION

(1) In the attempt to understand more clearly the relationship between classical and quantum mechanics, certain authors,[1,2] inspired by the Weyl–Wigner quantization, have viewed quantization as a deformation of the following two algebraic structures for functions (corresponding to classical observables) on phase space:

—the associative algebra defined by the ordinary product of functions.

—the Lie algebra defined by the Poisson bracket.

In these attempts, the mechanical systems considered were autonomous, and, consequently, their Hamiltonians were time-independent. This enabled them to make a direct use of the full richness of the symplectic geometry. Indeed, if F is the 2-form defining the symplectic structure and H is the Hamiltonian, the equations of motion of the system considered read: $i(X_H)F = -dH$.

Here we aim to extend those attempts to nonautonomous systems with time-dependent constraints and Hamiltonians which are explicitly time-dependent. Evidently, such systems are not closed insofar as there are external influences which alter the characteristics of the system. But here also we restrict ourselves to systems with a finite number of degrees of freedom. In this case, the previous analysis is not directly applicable as our manifold, on which we define the motions and the Hamiltonian, has odd dimensions and hence cannot be a symplectic manifold. The notion of *canonical manifold* corresponds to this situation.

This problem of quantization has received relatively little systematic attention,[3,4] although its study is necessary for an important number of physical problems, related for example to lasers or to the interactions with electromagnetic fields in general. Also certain natural problems lead to the study of time-dependent harmonic oscillators, both classical and quantum others result from time-dependent boundary conditions.[5]

In this work, we present a coherent approach to the evolution of quantum systems. Our approach is grounded on deformed products[1,6] (often called star products, Secs. 7 and 8) since with them the correspondence between classical and quantum mechanics appears to be more precise and clear, and the interpretation of this approach in terms of the conventional quantum mechanics is relatively direct.[7]

We note that if the phase space of the system admits a nonvanishing second Betti number, there exist for the system several quantizations corresponding to convenient equivalence classes of star products.[2] Similar circumstances occur for a canonical manifold.[8]

(2) Our approach is based on a doubling of the time variable. Thus one introduces "two times" denoted t and τ, playing different roles; the time t plays the role of *geometrical time* corresponding to the control of the system, while the other τ, which parametrizes the evolution plays the role of a *dynamical time*. The distinction between these two times is made in a natural way (see Sec. 6). Usual approximations can be considered as limiting cases of our approach corresponding respectively to big or small variations of t and τ. The classical as well as the quantum results point out the role played by the variable $(t + \tau)$. This shows, *a posteriori*, that the results can be interpreted in terms of one time, as desired.

After setting up the appropriate geometric framework, we first study classical mechanics, and later, in terms of its deformation, we consider quantum mechanics, the deformation parameter being $\nu = \hbar/2i$. The formalism developed in this work is susceptible to play a role in the analysis of various open systems as, for example, in statistical mechanics.

I. THE GEOMETRIC FRAMEWORK

(3) The state space as canonical manifold: (a) The *state space* of a dynamical system with time-dependent constraints and n degrees of freedom is given by a $(2n + 1)$-*dimensional* differentiable manifold W admitting an interesting geometric structure. The manifold W is supposed always connected and paracompact. The classical time can be de-

a) Permanent address: Department of Mathematics, Kuwait University, Kuwait.

 0022-2488/84/040923-09$02.50

scribed by means of a C^∞-map t: $W \to \mathbb{R}$ such that $dt \neq 0$. We have thus a distinguished function $t \in N = N(W)/ = C^\infty(W;\mathbb{R})$.

One of the authors has proved[8] that W admits then *a structure* (Λ,t) *of canonical manifold* that is a structure of regular Poisson manifold[8] of codimension 1, such that its symplectic foliation is given by the global coordinate t. We see that Λ is a 2-tensor of rank $2n$ satisfying in the sense of the Schouten brackets[9,10]

$$[\Lambda,\Lambda] = 0 \tag{3.1}$$

and

$$[\Lambda,t] = 0. \tag{3.2}$$

The canonical manifold (W,Λ,t) admits the foliation $t = \text{const}$ such that Λ defines each leaf Σ_t as a symplectic hypersurface. The group of the canonical transformations of the dynamical system is nothing but the group of the automorphisms of the canonical structure.[8]

This canonical structure should be completed by *a vector field* E such that

$$i(E)\,dt = 1, \quad \mathscr{L}(E)\Lambda = 0, \tag{3.3}$$

where $i(\cdot)$ is the inner product and $\mathscr{L}(\cdot)$ the Lie derivative. (W,Λ,t,E) *is the geometric representation of the state space of the considered dynamical system.*

(b) Consider a $2n$-dimensional symplectic manifold (M,F). Introduce the manifold $W = M \times \mathbb{R}$ and denote by t the projection $W \to \mathbb{R}$, by E the vector field $\partial/\partial t$. The structure 2-tensor of (M,F) defines on W a 2-tensor Λ of rank $2n$ such that (W,Λ,t) is a canonical manifold (called a product canonical manifold). We say that a dynamical system has *time-independent constraints* if its state space can be defined by a product canonical manifold (W,Λ,t,E).

(c) In the general case, there are on W atlases of charts $\{x^a\} = \{x^{\bar 0} = t; x^h\}$ $(a,b,\dots = \bar 0,1,\dots,2n;\ h = 1,\dots,2n)$ such that E admits only as nonvanishing component $E^{\bar 0} = 1$. In such a chart of domain U, we have according to (3.2):

$$\Lambda^{\bar 0 h} = 0.$$

More precisely, it follows from the study of the canonical manifolds[8] that there are atlases of charts $\{x^a\} = \{x^{\bar 0},x^h\} = \{x^{\bar 0},x^\alpha,x^{\bar\alpha}\}$ $(\alpha = 1,\dots,n;\ \bar\alpha = \alpha + n)$ such that E and Λ admit only as nonvanishing components:

$$E^{\bar 0} = 1, \quad \Lambda^{\alpha\bar\alpha} = -\Lambda^{\bar\alpha\alpha} = 1. \tag{3.4}$$

Such a chart is sayed to be a *canonical chart* for (W,Λ,t,E). We obtain the usual notations of classical mechanics if we set

$$x^\alpha = p_\alpha, \quad x^{\bar 0} = q^0 = t, \quad x^{\bar\alpha} = q^\alpha. \tag{3.5}$$

(d) Consider the $2(n+1)$-manifold $\widetilde W = W \times \mathbb{R}$ and let π: $\widetilde W \to W$ be the corresponding projection. Denote by $x^0 = p_0$ the canonical coordinate of $\mathbb{R}$ and by $\widetilde Z$ the vector field $\partial/\partial x^0 = \partial_0$ of W. The elements Λ and E of W define on $\widetilde W$ a 2-tensor and a vector denoted by the same notations. Introduce on $\widetilde W$ the 2-tensor:

$$\widetilde\Lambda = \widetilde Z \wedge E + \Lambda. \tag{3.6}$$

Take for W a canonical chart $\{x^\alpha = p_\alpha, x^{\bar 0} = q^0, x^{\bar{}} = q^\alpha\}$ of domain U. We obtain for $\widetilde W$ a chart $\{x^A\} = \{x^i = p_i, x^{\bar i} = q^i\} = \{x^0,x^\alpha,x^{\bar 0},x^{\bar\alpha}\}$ $(A,B,\dots = 0; \bar 0,1,\dots 2n;\ i = 0,\alpha)$ of domain $\widetilde U = U \times \mathbb{R}$ such that $\widetilde\Lambda$ admits the nonvanishing components:

$$\widetilde\Lambda^{0\bar 0} = -\widetilde\Lambda^{\bar 0 0} = 1, \quad \widetilde\Lambda^{\alpha\bar\alpha} = -\widetilde\Lambda^{\bar\alpha\alpha} = 1. \tag{3.7}$$

The tensor $\widetilde\Lambda$ has the rank $2(n+1)$ and satisfies

$$[\widetilde\Lambda,\widetilde\Lambda] = 0.$$

Therefore, $(\widetilde W,\widetilde\Lambda)$ is a $2(n+1)$-dimensional symplectic manifold such that $\pi_* \widetilde\Lambda = \Lambda$. For the corresponding 2-form $\widetilde F$, the introduced chart is a canonical chart and we have locally:

$$\widetilde F_{|\widetilde U} = dx^0 \wedge dx^{\bar 0} + \sum_\alpha dx^\alpha \wedge dx^{\bar\alpha} = dp_0 \wedge dq^0 + dp_\alpha \wedge dq^\alpha. \tag{3.8}$$

Thus p_i is canonically conjugate to q^i with respect to $\widetilde F$; in particular p_0 *is by definition, canonically conjugate to* $q^0 = t$.

(4) Integral curves and flows: Let M be a m–dimensional manifold and Z a vector field on M. An *integral curve* (or an orbit) of the vector field Z through $x \in M$ is a smooth curve satisfying

$$\frac{d\gamma(s)}{ds} = Z(\gamma(s)), \quad \gamma(0) = x,\ \gamma(s) = y \tag{4.1}$$

for all $s \in I$, where I is an open real interval centered at the origin. If $\{y^a\}$ $(a = 1,\dots,m)$ is a chart of M of domain U, we have on U

$$\frac{d}{ds}(y^a(\gamma(s))) = Z^a(\gamma(s)).$$

It is well known that the integral curves of Z define a flow f_s so that, for s sufficiently small, we have

$$y = f_s(x) = f(x;s) \tag{4.2}$$

with, if s, s' are sufficiently small,

$$f(f(x;s);s') = f(x;s+s'). \tag{4.3}$$

Therefore, the formula (4.2) is often written under the form

$$y = \exp(sZ)\cdot x. \tag{4.4}$$

Let u_0 be a function, element of $N(M)$. If we set

$$u_s = f_s^* u_0 = u_0 \circ f_s,$$

it follows from the properties of the Lie derivative that u_s satisfies at each point x:

$$\frac{d}{ds} u_s = \mathscr{L}(Z) u_s \tag{4.5}$$

and takes the initial value $u_0 \in N(M)$ at $s = 0$. Therefore, the evolution in s of each function u_s, defined on $M \times I$, and solution of (4.5), is strictly connected with the flow of the integral curves of Z.

II. CLASSICAL DYNAMICS

(5) Orbits of the dynamical system: Let (W,Λ,t,E) be the state space of the considered dynamical system and $(\widetilde W,\widetilde\Lambda)$ the symplectic manifold associated with the state space in the sense of subsection (3)(d). Here we consider essentially functions that are elements of $N(W)$. Such a function $u \in N(W)$ admits in $\widetilde W$ an inverse image $\pi^* u$ that we will also denote (abusing the notation) by u, as it is independent of p_0.

(a) Dynamics is determined on the state space by a function $H \in N(W)$, the classical (time-dependent) *Hamiltonian* of the system. This Hamiltonian determines on W a vector field:

$$Y_H = E + [\Lambda, H]. \tag{5.1}$$

We have proved by means of an invariant approach the following[8]:

Proposition: The motions of the dynamical system are described, with respect to the state space (W,Λ,t,E) and in function of the time t, by the integral curves of Y_H.

Denote by $\{\ ,\ \}$ the Poisson bracket of the state space and introduce a canonical chart $\{q^0 = t, q^\alpha, p_\alpha\}$ of domain U of this space. For a motion $c(t)$, the proposition can be translated by

$$\frac{dq^0}{dt}(c(t)) = 1 \tag{5.2}$$

and

$$\frac{dq^\alpha}{dt}(c(t)) = \frac{\partial H}{\partial p_\alpha}(c(t)) = \{H, q^\alpha\}(c(t)),$$

$$\frac{dp_\alpha}{dt}(c(t)) = -\frac{\partial H}{\partial q^\alpha}(c(t)) = \{H, p_\alpha\}(c(t)) \tag{5.3}$$

or

$$\frac{dq^\alpha}{dt} = \frac{\partial H(t)}{\partial p_\alpha}, \quad \frac{dp_\alpha}{dt} = -\frac{\partial H(t)}{\partial q^\alpha},$$

which are usual Hamilton's equations of motion.

We note that we impose that, along the considered integral curves of Y_H, the evolution parameter is precisely t.

(b) Introduce the Poisson bracket $\{\ ,\ \}^{\sim}$ of the symplectic manifold $(\widetilde{W},\widetilde{\Lambda})$. We remark that, according to the definition of $\widetilde{\Lambda}$, we have

$$\{p_0, u\}^{\sim} = \doteq \frac{\partial u}{\partial t} \quad [\text{for } u \in N(W)]. \tag{5.4}$$

The Hamiltonian vector field of $(\widetilde{W},\widetilde{\Lambda})$ corresponding to $(p_0 + H) \in N(\widetilde{W})$:

$$\widetilde{Y}_H = [\widetilde{\Lambda}, p_0 + H] \tag{5.5}$$

is such that $\pi_* \widetilde{Y}_H = Y_H$ and admits the component

$$\widetilde{Y}^0 = -\frac{\partial H}{\partial t}.$$

Therefore, the projection by π of the integral curves of $\widetilde{Y}_H$ are the integral curves of Y_H and Hamilton's equations can be completed by

$$\frac{dp_0}{dt} = -\frac{\partial H(t)}{\partial q^0}. \tag{5.6}$$

(c) It can be easily verified that if $u \in N(W)$, its total derivative with respect to t, along the orbits of Y_H, is given by

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \{H(t), u(t)\}. \tag{5.7}$$

Inversely, (5.7) determines the orbits, and thus it is equivalent to Hamilton's equations.

If $\tilde{u} \in N(\widetilde{W})$, its total derivative with respect to t, along the orbits of $\widetilde{Y}_H$, is given by

$$\frac{d\tilde{u}}{dt} = \{p_0 + H, \tilde{u}\}^{\sim}. \tag{5.8}$$

If $\tilde{u} = u$, (5.8) reduces to (5.7). Also, if we take $\tilde{u} = p_0 + H$, the relation (5.8) gives

$$p_0 + H = \text{const.}$$

Thus one can roughly say that, *along the orbits, p_0 is equal to the negative of the energy up to an additive constant.*

(6) The change of time variable: (a) Suppose that we parametrize the integral curves of Y_H by means of an arbitrary parameter τ. We obtain $\gamma(\tau)$ satisfying

$$d\gamma(t)/d\tau = Y_H(\gamma(\tau)) \quad [\gamma(0) = x,\ \gamma(t) = y], \tag{6.1}$$

and, in canonical coordinates, we see that

$$\frac{dt}{d\tau} = 1. \tag{6.2}$$

(6.1) is thus equivalent to Hamilton's equations. We have seen [subsection (4)] that the evolution in τ of a one-parameter family u_τ of elements of $N(W)$, satisfying at the points x of W

$$\frac{du_\tau}{d\tau} = \mathscr{L}(Y_H)u_\tau \tag{6.3}$$

and taking the value $u_0 \in N(W)$ at $\tau = 0$, is strictly connnected with the flow f_τ of the integral curves of Y_H. For a sufficiently small τ, we have

$$u_\tau = f_\tau^* u_0 = u_0 \circ f_\tau$$

with

$$f(f(x;\tau)\tau') = f(x;\tau + \tau').$$

It follows from (6.2) that

$$t(y) = t(x) + \tau. \tag{6.4}$$

Introduce a canonical chart $\{x^{\bar{0}} = t, x^\alpha, x^{\bar{\alpha}}\}$ of the state space of domain U. On this domain we set $\hat{x} = \{x^\alpha, x^{\bar{\alpha}}\}$. The flow f_τ can be described on U, with evident notations, in the following way,

$$\hat{y} = \hat{f}(\hat{x}, t(x); \tau)$$

with (6.4), and we have, for suitable τ,τ',

$$\hat{f}(\hat{f}(\hat{x}, t(x); \tau), t(x) + \tau; \tau') = \hat{f}(\hat{x}, t(x); \tau + \tau').$$

(b) We adopt systematically in the following part a different point of view to that of orbits, but similar to that introduced[1] in the case of a time-independent Hamiltonian on a phase space (M,F). In this vein, we consider only now *the one-parameter family u_τ of elements of $N(W)$ taking the value $u_0 \in N(W)$ for $\tau = 0$ and subject to satisfy the differential equation* (6.3), *that is,*

$$\frac{du_\tau}{d\tau} = \frac{\partial u_\tau}{\partial t} + \{H, u_\tau\} \tag{6.5}$$

or

$$\frac{du_\tau}{d\tau} = \{p_0 + H, u_\tau\}^{\sim}. \tag{6.6}$$

The introduced functions, elements of $N(W)$, depend upon $x \in W$, where the global coordinate t appears. For $u_0 = t$, we obtain as solution of (6.5) the function $t + \tau$, in agreement with (6.4).

We see how, in this approach, one is led to the introduction of two time variables having different roles (a role of geometric coordinate and a role of evolution parameter) noted respectively t and τ. In this context, Eq. (6.5) can be considered as the main intrinsic equation of classical dynamics.

III. QUANTUM MECHANICS

(7) Star products on a symplectic manifold: Let M be a differentiable manifold and $N = N(M)$ the space of smooth functions on M. We denote by $E(N;\nu)$ the space of formal power series in $\nu\in\mathbb{C}$ with coefficients in N; let

$$u_\nu = \sum_{r=0}^{\infty} \nu^r u_r \quad (u_r \in N).$$

(a) Consider a symplectic manifold (M,Λ). A star-product on (M,Λ) is a bilinear map $N\times N \to E(N;\nu)$ given by

$$u *_\nu v = uv + \sum_{r=1}^{\infty} \nu^r C_r(u,v) \quad (u,v\in N), \tag{7.1}$$

where the C_r are bilinear differential operators vanishing on the constants and satisfying the following conditions:

(1) $C_1(u,v) = \{u,v\} = P(u,v)$ (the Poisson bracket on M);

(2) the canonical extension of $u *_\nu v$ as a bilinear map from $E(N;\nu)\times E(N;\nu)$ into $E(N;\nu)$ is associative;

(3) C_r is symmetric if r is even and antisymmetric if r is odd. We have according to the assumptions

$$u *_\nu 1 = 1 *_\nu u = u, \quad u *_{-\nu} v = v *_\nu u. \tag{7.2}$$

The notion of star product[1,2,6] has been introduced from the consideration of the deformations of an associative structure (here the usual product of functions) and from the connection of these deformations with quantization.

Neroslavsky and Vlassov[11] have proved that, on each symplectic manifold (M,Λ) such that the third Betti number $b_3(M)$ of the manifold vanishes, there are star products. Lecomte and de Wilde[12] have shown that each symplectic manifold that is a cotangent bundle admits star products.

A star product defines by skew-symmetrization an alternate bilinear map $N\times N\to E(N;\lambda)$ (where $\lambda = \nu^2$) given by

$$\begin{aligned}[u,v]_* &= (2\nu)^{-1}(u *_\nu v - v *_\nu u) \\ &= P(u,v) + \sum_{r=1}^{\infty} \lambda^r C_{2r+1}(u,v).\end{aligned} \tag{7.3}$$

The canonical extension of $[u,v]_*$ as an alternate bilinear map from $E(N;\lambda)\times E(N;\lambda)$ into $E(N;\lambda)$ defines on $E(N;\lambda)$ a structure of Lie algebra. We say that the star product (7.1) generates by skew-symmetrization a deformation (7.2) of the Poisson Lie algebra.

(b) Consider the cotangent bundle of $\mathbb{R}^n$, let $\mathbb{R}^n\times\mathbb{R}^n = \mathbb{R}^{2n}$, with its canonical symplectic structure defined by the 2-tensor Λ. Introduce the bidifferential operators P^r of maximum order r for each argument, defined in the canonical chart $\{x^h\}$ of $\mathbb{R}^{2n}$ by

$$P^r(u,v) = \Lambda^{h_1k_1}\cdots\Lambda^{h_rk_r}\partial_{h_1\cdots h_r}u\,\partial_{k_1\cdots k_r}v. \tag{7.4}$$

We set $P^0(u,v) = uv$, and we have $P^1 = P$.

Given a formal function $f(z)$ with constant coefficients such that $f(0) = 1$, substitute P^r to z^r in the expansion of $f(\nu z)$; we obtain a bilinear map $(u,v)\in N\times N\to u *_\nu v = f(\nu P)(u,v) \in E(N;\nu)$. We wish to choose f so that we define thus a star product. The answer is given by the following[1]:

Proposition: For the flat symplectic manifold (R^{2n},Λ) there is a unique formal function of the Poisson bracket that generates a star product: it is the exponential function.

We have

$$u *_\nu v = \sum_{r=0}^{\infty} \frac{\nu^r}{r!} P^r(u,v) = \exp(\nu P)(u,v), \tag{7.5}$$

which generates the deformation of the Poisson Lie algebra $(\lambda = \nu^2)$

$$[u,v]_* = \sum_{r=0}^{\infty} \frac{\lambda^r}{(2r+1)!} P^{2r+1}(u,v) = \nu^{-1}\sinh(\nu P)(u,v). \tag{7.6}$$

It is remarkable that, for $\nu = \hbar/2i$, we deduce from (7.6) a bracket $(2/\hbar)\sin(\hbar P/2)$ given in 1949 by Moyal[13] in the context of the Weyl–Wigner quantization. The star product (7.5) is called the *Moyal star product*.

(c) Recall that two star products, denoted by $*_\nu$ and $*'_\nu$, are said to be *equivalent* if there exists a linear map from N into $E(N;\nu)$ given by

$$T_\nu = \mathrm{Id}_N + \sum_{r=1}^{\infty} \nu^r T_r, \tag{7.7}$$

where the T_r are differential operators, such that

$$T_\nu(u *'_\nu v) = (T_\nu u) *_\nu (T_\nu v) \quad (\forall u,v\in N). \tag{7.8}$$

Introduce now the following notation: we denote by Q^r a bidifferential operator of maximum order r in each argument satisfying the assumptions 1 and 3 of a and such that its principal symbol coincides with the principal symbol of P^r (which is independent of the chart). We take $Q^0(u,v) = uv$, $Q^1 = P$. Introduce the following:

Definition: A Vey star product on (M,Λ) is a star product of the form

$$u *_\nu v = \sum_{r=0}^{\infty} \frac{\nu^r}{r!} Q^r(u,v). \tag{7.9}$$

It is possible to prove[6]:

Proposition: On an arbitrary symplectic manifold, each star product is equivalent to a Vey star product.

(8) Tangential star products on a regular Poisson manifold: (a) Let (W,Λ) be a regular Poisson manifold of dimension m and codimension q. *A star product* is defined on (W,Λ) by

$$u *_\nu v = \sum_{r=0}^{\infty} \nu^r C_r(u,v) \quad [u,v\in N(W)], \tag{8.1}$$

where $C_0(u,v) = uv$ and where the bidifferential operators C_r satisfy the same assumptions as in subsection (7)(a) for $r\geqslant 1$.

Introduce on W an atlas of charts $\{x^a\} = \{x^s,x^h\}$ ($s = 1,\ldots, q$; $h = q+1,\ldots m$) *adapted to* Λ and so to the foliation. We say that (8.1) is *a tangential star product*[14] if, for each domain U of an adapted chart, the local expressions of the C_r ($r\geqslant 1$) contain only the tangential derivatives $\partial_{h_1\cdots h_l}$ of the arguments. A tangential star product of (W,Λ) induces a star product on each symplectic leaf of (W,Λ).

All the considerations of subsection (7) are valid for the tangential star products on a regular Poisson manifold. Such a tangential star product is said to be *a Vey tangential star product* if it induces on the leafs Vey star products. It is possible to prove[14] that each tangential star product of (W,Λ) is equivalent to a Vey tangential star product of (W,Λ).

One of us has proved under a general cohomology condition[14] the existence of tangential star products on (W,Λ) by means of an adaptation of the argument of Neroslavsky and Vlassov. On the other hand, consider a manifold M admitting a foliation $\mathscr{F}$ of codimension q. The cotangent bundle $T^*\mathscr{F}$ of the foliation admits a natural structure Λ of regular Poisson manifold of codimension q, the leafs of $(T^*,\mathscr{F},\Lambda)$ being the cotangent bundles of the leafs of $\mathscr{F}$. The argument of Lecomte and de Wilde shows that $(W=T^*\mathscr{F},\Lambda)$ admits always tangential star products.

(9) The formal equation of quantum dynamics: Come back to our dynamical system and let (W,Λ,t) be *the canonical manifold* defined by the state space. In the following part, *we introduce systematically on (W,Λ,t) a given tangential star product, denoted by $*_\nu$, invariant under E,* that will induce the quantization.

(a) Introduce the space $N^c=N^c(W)=C^\infty(W,\mathbb{C})$ and denote by $E(N^c;\nu)$ the space of formal power series in $\nu\in\mathbb{C}$ with coefficients in N^c. Consider a one-parameter family u_τ of elements of $E(N^c;\nu)$:

$$u_\tau=\sum_{r=0}^{\infty}\nu^r u_{(r)\tau}\quad (u_{(r)\tau}\in N^c) \tag{9.1}$$

satisfying the differential equation deduced from (6.5) by deformation:

$$\frac{du_\tau}{d\tau}=\frac{i}{\hbar}2\nu\left\{\frac{\partial u_\tau}{\partial t}+[H,u_\tau]_*\right\}, \tag{9.2}$$

where the distinguished value of ν suggested by the Moyal product $\nu=\hbar/2i$ appears.

According to (9.1), the differential equation (9.2) is equivalent to the system:

$$\frac{du_{(s)\tau}}{d\tau}=\frac{2i}{\hbar}\frac{\partial u_{(s-1)}}{\partial t}+\frac{2i}{\hbar}\sum_{r+2r'=s-1}C_{2r'+1}(H,u_{(r)\tau})$$
$$(r,r'\geqslant 0;\ s=0,1,\ldots). \tag{9.3}$$

It is easy to deduce from (9.3) by induction on s the following:

Proposition: There is a unique solution u_ν of (9.2) taking a given value $u_0\in E(N^c;\nu)$ at $\tau=0$.

In particular, for $u_0=t$, we obtain the solution $t+(i/\hbar)2\nu\tau$.

(b) Suppose that *our dynamical system admits time-independent constraints* [see subsection (3)(b)], the Hamiltonian being time-independent. We have a symplectic manifold (M,F) of dimension $2n$ and the state space is given by the corresponding product canonical manifold so that $W=M\times\mathbb{R}$ and $\widetilde{W}=M\times\mathbb{R}^2$. The space $\mathbb{R}^2$ defined by the canonical coordinates $\{x^0=p_0, x^0=q^0=t\}$ admits a canonical symplectic structure, and we denote by $*_\nu^{(1)}$ the corresponding Moyal star product. Suppose that (M,F) admits a Vey star product $*_\nu^{(2)}$. *We can deduce from $*_\nu^{(1)}$ and $*_\nu^{(2)}$ a Vey star product on the symplectic manifold $(\widetilde{W},\widetilde{\Lambda})$* in the following way: we set

$$u^{(1)}*_\nu^{(1)}v^{(1)}=\sum_{r=0}^{\infty}\nu^r C_r^{(1)}(u^{(1)},v^{(1)})\quad [u^{(1)},v^{(1)}\in N(\mathbb{R}^2)]$$

and

$$u^{(2)}*_\nu^{(2)}v^{(2)}=\sum_{s=0}^{\infty}\nu^s C_s^{(2)}(u^{(2)},v^{(2)})\quad [u^{(2)},v^{(2)}\in N(M)].$$

If

$$\widetilde{C}_r=\sum_{s+s'=r}C_s^{(1)}C_{s'}^{(2)}\quad (s,s'\geqslant 0) \tag{9.4}$$

the $\widetilde{C}_r$ are bidifferential operators on $N(\widetilde{W})$; we set

$$\tilde{u}\tilde{*}\tilde{v}=\sum_{r=0}^{\infty}\nu^r\widetilde{C}_r(\tilde{u},\tilde{v})\quad [\tilde{u},\tilde{v}\in N(\widetilde{W})]. \tag{9.5}$$

We can prove that (9.5) is *associative*. In fact, if $u^{(1)},v^{(1)},w^{(1)}\in N(\mathbb{R}^2)$, $u^2,v^2,w^{(2)}\in N(M)$, it follows from (9.4), (9.5) that we have, with evident notations,

$$(u^{(1)}\cdot u^{(2)})\tilde{*}_\nu(v^{(1)}\cdot v^{(2)})=(u^{(1)}*_\nu^{(1)}v^{(1)})\cdot(u^{(2)}*_\nu^{(2)}v^{(2)}),$$

and we deduce from the associativity of $*_\nu^{(1)}$ and $*_\nu^{(2)}$ that

$$((u^{(1)}\cdot u^{(2)})\tilde{*}_\nu(v^{(1)}\cdot v^{(2)}))\tilde{*}_\nu(w^{(1)}\cdot w^{(2)})$$
$$=(u^{(1)}\cdot u^{(2)})\tilde{*}_\nu((v^{(1)}\cdot v^{(2)})\tilde{*}_\nu(w^{(1)}\cdot w^{(2)})). \tag{9.6}$$

A degenerated function $\tilde{u}_d\in N(W)$ is, by definition, a finite sum of product of functions that are images of elements of $N(\mathbb{R}^2)$ and $N(M)$. It follows from (9.6) by linearity that we have for degenerated functions

$$(\tilde{u}_d\tilde{*}_\nu\tilde{v}_d)\tilde{*}_\nu\tilde{w}_d=\tilde{u}_\nu\tilde{*}_\nu(\tilde{v}_d\tilde{*}_\nu\tilde{w}_d). \tag{9.7}$$

Choose a point $\tilde{x}$ of $\widetilde{W}$ and a domain $\widetilde{U}$ of a chart which contains $\tilde{x}$. We restrict all the elements to the manifold $(\widetilde{U},\widetilde{\Lambda}_{|\widetilde{U}})$. The r-jet at $\tilde{x}$ of a function $\tilde{u}\in N(\widetilde{U})$ is a degenerated function. It follows from (9.7) that the associativity relation is satisfied at $\tilde{x}$ for each r and for arbitrary functions $\tilde{u},\tilde{v},\tilde{w}\in N(\widetilde{W})$. Therefore, (9.5) is *associative* and defines a star product. It is easy to see that $\tilde{*}_\nu$ is a Vey star product on the symplectic manifold $(\widetilde{W},\widetilde{\Lambda})$. If $u,v\in N=N(W)$, we set, by definition,

$$u*_\nu v=u\tilde{*}_\nu v.$$

We define thus on (W,Λ,t) the tangential star product $*_\nu$ *invariant under E*. We adopt this star product on the state space.

(c) For a dynamical system with time-independent constraints, Eq. (9.2) can be written, with the notations of b, under the form

$$\frac{du_\tau}{d\tau}=\frac{i}{\hbar}2\nu[p_0+H,u_\tau]_{\tilde{*}}$$
$$=\frac{i}{\hbar}\{(p_0+H)\tilde{*}_\nu u_\tau-u_\tau\tilde{*}_\nu(p_0+H)\}. \tag{9.8}$$

Introduce the $\tilde{*}_\nu$-powers $\tilde{u}^{(*)p}$ of a function $\tilde{u}\in N(\widetilde{W})(\tilde{u}^{(*)p}=\tilde{u}^{(*)p-1}\tilde{*}_\nu\tilde{u})$ and set

$$\mathrm{Exp}_{\tilde{*}}(\tilde{u}s)=\sum_{p=0}^{\infty}\frac{s^p}{p!}\tilde{u}^{(*)p}\quad (s\in C). \tag{9.9}$$

If $\tilde{u}=u$, the right member of (9.9) is independent of p_0, and we denote the left member by $\mathrm{Exp}_*(us)$. It is easy to verify that the solution of (9.2) or (9.8), which takes the value u_0 at $\tau=0$, can be written under the following formal way:

$$u_\tau=\mathrm{Exp}_{\tilde{*}}((i/\hbar)(p_0+H)\tau)\tilde{*}_\nu u_0\tilde{*}_\nu$$
$$\times\mathrm{Exp}_{\tilde{*}}(-(i/\hbar)(p_0+H)\tau). \tag{9.10}$$

In particular, we verify by means of a derivation with respect to p_0 that the right member is effectively independent of p_0.

(10) The spectrum of an observable: Consider now *the point of view of the mathematical analysis* and give to ν the value $\hbar/2i$. We denote by $*$ the star product for this value.

(a) We assume first that our dynamical system admits *time-independent constraints* and we use the notations on subsection (9)(b),(c).

Suppose that $u\in N = N(W)$ is such that we can define $\mathrm{Exp}_*((i/\hbar)us)$, for fixed t and s in a complex neighborhood of the origin as *a distribution on M*; t being always fixed, we obtain a spectral decomposition by means of a Fourier transform of $\mathrm{Exp}_*((i/\hbar)us)$ with respect to s in the sense of the distributions on M, exactly as in the purely time-independent case.

Suppose for simplicity that we consider the case where, t being fixed, $\mathrm{Exp}_*((i/\hbar)us)$ has a unique Fourier–Dirichlet expansion on M with respect to s (purely discontinuous spectrum):

$$\mathrm{Exp}_*((i/\hbar)us) = \sum_\lambda e^{(i/\hbar)\lambda s}\rho_\lambda$$

$$(\rho_\lambda \in N^c;\ \rho_\lambda(t) \not\equiv 0 \text{ for each } t), \tag{10.1}$$

where the λ depend upon t. This expansion is similar to the spectral expansion of an operator. The λ define for each *t the spectrum* of u. We obtain by differentiation of (10.1) with respect to s:

$$(\hbar/i)d\,\mathrm{Exp}_*((i/\hbar)us)/ds = \sum e^{(i/\hbar)\lambda s}\lambda\rho_\lambda$$

and, according to the definition of Exp_*:

$$(\hbar/i)d\,\mathrm{Exp}_*((i/\hbar)us)/ds = \sum e^{(i/\hbar)\lambda s}(u*\rho_\lambda)$$

$$= \sum e^{(i/\hbar)\lambda s}(\rho_\lambda * u).$$

It follows by uniqueness:

$$u*\rho_\lambda = \rho_\lambda * u = \lambda\rho_\lambda,$$

and we have $\Sigma\rho_\lambda = 1$.

(b) *Come back to a dynamical system with time-dependent constraints*. We have on the canonical manifold (W,Λ,t) the star product denoted by $*$.

Suppose that

$$u*\rho_\lambda = \rho_\lambda * u = \lambda\rho_\lambda, \quad \sum \rho_\lambda = 1, \tag{10.2}$$

where the functions ρ_λ are such that $\rho_\lambda(t) \not\equiv 0$ for each t and where the eigenvalues λ depend only upon t and are such that, for two different eigenvalues, $(\lambda' - \lambda)(t) \neq 0$ for each t. We say that the λ define for each *t the spectrum* of u; the ρ_λ define thus the (nonnormalized) *states* corresponding to the λ. We deduce from (1.2): $u = \Sigma\lambda\rho_\lambda$.

For two eigenvalues λ,λ', we deduce from (10.2)

$$\rho_\lambda * u * \rho_{\lambda'} = \lambda\rho_\lambda * \rho_{\lambda'} = \lambda'\rho_\lambda * \rho_{\lambda'}.$$

It follows that if $\lambda' \neq \lambda$, we have $\rho_\lambda * \rho_{\lambda'} = 0$. Moreover,

$$\rho_\lambda = \Bigl(\sum \rho_{\lambda'}\Bigr) * \rho_\lambda = \rho_\lambda * \rho_\lambda.$$

We see that ρ_λ is an *eigenprojector* corresponding to λ.

Show that *the spectrum and the ρ_λ's are characterized in a unique way by the relations* (10.2). In fact, consider another set $(\lambda',\rho'_{\lambda'})$ satisfying the same assumptions as the set (λ,ρ_λ), such that

$$u*\rho'_{\lambda'} = \rho'_{\lambda'} * u = \lambda'\rho'_{\lambda'}, \quad \sum \rho'_{\lambda'} = 1.$$

We have:

$$\rho'_{\lambda'} * u * \rho_\lambda = \lambda'\rho'_{\lambda'} * \rho_\lambda = \lambda\rho'_{\lambda'} * \rho_\lambda.$$

For each t such that $\lambda'(t) \neq \lambda(t)$, we have $(\rho'_{\lambda'} * \rho_\lambda)(t) = 0$, and so

$$\Bigl\{\Bigl(\sum \rho'_{\lambda'}\Bigr) * \rho_\lambda\Bigr\}(t) = \rho_\lambda(t) = 0.$$

We see that the two spectrums coincide necessarily. Moreover, the relations

$$\Bigl(\sum \rho'_{\lambda'}\Bigr) * \rho_\lambda = \rho_\lambda, \quad \rho'_\lambda * \Bigl(\sum \rho_{\lambda'}\Bigr) = \rho'_\lambda$$

imply $\rho'_{\lambda'} * \rho_\lambda = \rho_\lambda = \rho'_{\lambda'}$ and the eigenprojectors are equal.

There follows immediately from the characterization of the spectrum by (10.2):

Lemma: Substitute for the function $u\in N$ the function $(u + k)\in N$, where k depends only upon t. The spectrum of $(u + k)$ can be deduced from the spectrum of u by the change $\lambda \to \lambda + k$ and the corresponding states remain unchanged.

(c) A tangential star product on a regular Poisson manifold is said to be *nondegenerated* if, for a function $u\in N^c$, the relation $\bar{u} * u = 0$ (where $\bar{u}$ is the complex conjugate of u) implies $u = 0$. It is well known that the Moyal product on $\mathbb{R}^n$ and many other star products on a symplectic manifold or a Poisson manifold are nondegenerated.

Suppose that our star product $*$ on the canonical manifold (W,Λ,t) is *nondegenerated*. If $u,v\in N^c$, it follows from (7.2) that we have

$$\overline{u*v} = \bar{v} * \bar{u}.$$

We deduce then from (10.2), where u is real-valued, by complex conjugation

$$\bar{\rho}_\lambda * u = u * \bar{\rho}_\lambda = \bar{\lambda}\bar{\rho}_\lambda, \quad \sum \bar{\rho}_\lambda = 1. \tag{10.3}$$

If $\bar{\lambda} \neq \lambda$, it follows from (10.2) and (10.3) that we have $\bar{\rho}_\lambda * \rho_\lambda = 0$, which implies $\rho_\lambda = 0$, and we obtain a contradiction. Therefore, $\bar{\lambda} = \lambda$ and the spectrum of $u\in N$ is real. We deduce from the characterization of the set (λ,ρ_λ) by (10.2) that $\bar{\rho}_\lambda * \rho_\lambda = \rho_\lambda = \bar{\rho}_\lambda$ and we see that the ρ_λ's are real-valued. We have:

Proposition: For a nongenerated star product, the spectrum of all real-valued function is real and the corresponding eigenprojectors ρ_λ's are real-valued.

Let (Σ_t,F_t) be the symplectic leaf of (W,Λ,t) corresponding to the time t. Denote by η_t the symplectic volume element $F_t^n/n!$ on Σ_t. For each λ define $N_\lambda(t)$ by the formula

$$N_\lambda(t) = (2\pi\hbar)^{-n}\int_{\Sigma_t} \rho_\lambda(t)\eta_t.$$

If $N_\lambda(t)$ is finite, a normalized state of the dynamical system is defined by $\pi_\lambda = \rho_\lambda/N_\lambda$ and we know that, in the usual cases, $N_\lambda(t)$ is the multiplicity of the eigenvalue at the time t, in the sense of the conventional quantum mechanics.[1]

(d) In this context, *the main equation of quantum dynamics* [corresponding to (9.2)] *can be written under the form*

$$\frac{du_\tau}{d\tau} = \frac{\partial u_\tau}{\partial t} + [H,u_\tau]_* = \frac{\partial u_\tau}{dt} + \frac{i}{\hbar}(H * u_\tau - u_\tau * H). \tag{10.4}$$

(11) The evolution of the spectrum: Let u_τ be a solution of (10.4) taking the value u_0 at $\tau = 0$. If λ_τ is an eigenvalue corresponding to the eigenprojector ρ_τ, we have

$$u_\tau * \rho_\tau = \rho_\tau * u_\tau = \lambda_\tau \rho_\tau. \tag{11.1}$$

It follows from (11.1) by differentiation with respect to τ

$$\frac{du_\tau}{d\tau} * \rho_\tau + u_\tau * \frac{d\rho_\tau}{d\tau} = \frac{d\lambda_\tau}{d\tau}\rho_\tau + \lambda_\tau \frac{d\rho_\tau}{d\tau}. \tag{11.2}$$

We deduce from (11.2) by the left product of

$$\rho_\tau * \frac{du_\tau}{d\tau} * \rho_\tau + \lambda_\tau \rho_\tau * \frac{d\rho_\tau}{d\tau} = \frac{d\lambda_\tau}{d\tau}\rho_\tau + \lambda_\tau \rho_\tau * \frac{d\rho_\tau}{d\tau}.$$

We obtain

$$\frac{d\lambda_\tau}{d\tau}\rho_\tau = \rho_\tau * \frac{du_\tau}{d\tau} * \rho_\tau.$$

It follows similarly from (11.1) by differentiation with respect to t

$$\frac{\partial \lambda_\tau}{\partial t}\rho_\tau = \rho_\tau * \frac{\partial u_\tau}{\partial t} * \rho_\tau.$$

We obtain by difference

$$\left(\frac{d\lambda_\tau}{d\tau} - \frac{\partial \lambda_\tau}{\partial t}\right)\rho_\tau = \rho_\tau * \left(\frac{du_\tau}{d\tau} - \frac{\partial u_\tau}{\partial t}\right) * \rho_\tau. \tag{11.3}$$

But we deduce from Eq. (10.4) that

$$\rho_\tau * \left(\frac{du_\tau}{d\tau} - \frac{\partial u_\tau}{\partial t}\right) * \rho_\tau = \frac{i}{\hbar}(\rho_\tau * H * u_\tau * \rho_\tau - \rho_\tau * u_\tau * H * \rho_\tau);$$

let, according to (11.1),

$$\rho_\tau * \left(\frac{du_\tau}{d\tau} - \frac{\partial u_\tau}{\partial t}\right) * \rho_\tau = 0.$$

Therefore, $d\lambda_\tau/d\tau = \partial\lambda_\tau/\partial t$, and there is a function λ of $t \in \mathbb{R}$ such that

$$\lambda_\tau(t) = \lambda(t + \tau). \tag{11.4}$$

We have:

Theorem: The spectrum of u_τ is deduced from the spectrum of u_0 by $\lambda(t) \to \lambda(t + \tau)$.

(12) The evolution of the eigenprojectors: We search for the evolution of the eigenprojector ρ_τ of u_τ corresponding to $\lambda(t + \tau)$.

(a) We deduce from (11.2) and (11.4) the following relations:

$$\frac{du_\tau}{d\tau} * \rho_\tau + u_\tau * \frac{d\rho_\tau}{d\tau} = \lambda'(t + \tau)\rho_\tau + \lambda(t + \tau)\frac{d\rho_\tau}{d\tau}$$

and

$$\frac{\partial u_\tau}{\partial t} * \rho_\tau + u_\tau * \frac{\partial \rho_\tau}{\partial t} = \lambda'(t + \tau)\rho_\tau + \lambda(t + \tau)\frac{\partial \rho_\tau}{\partial t}.$$

According to Eq. (10.4) of quantum dynamics, we obtain by difference

$$u_\tau * \left(\frac{d\rho_\tau}{dt} - \frac{\partial \rho_\tau}{\partial t}\right) + [H,u_\tau]_* * \rho_\tau = \lambda(t + \tau)\left(\frac{d\rho_\tau}{d\tau} - \frac{\partial \rho_\tau}{\partial t}\right). \tag{12.1}$$

But we have

$$u_\tau * [H,\rho_\tau]_* + [H,u_\tau]_* * \rho_\tau = \lambda(t + \tau)[H,\rho_\tau]_*.$$

We are led to set:

$$\varphi_\lambda = \left(\frac{d\rho_\tau}{d\tau} - \frac{\partial \rho_\tau}{\partial t}\right) - [H,\rho_\tau]_*. \tag{12.2}$$

The relation (12.1) and the similar right relation deduced from (11.1) reduce to:

$$u_\tau * \varphi_\lambda = \varphi_\lambda * u_\tau = \lambda \varphi_\lambda, \tag{12.3}$$

where we have suppressed the argument $(t + \tau)$ of λ; φ_λ is thus an eigenfunction of u_τ corresponding to the considered eigenvalue. For a different eigenvalue $\hat{\lambda}$, we have for the corresponding eigenprojector $\rho_{\hat{\lambda}}$, according to the argument of subsection (10)(b):

$$\rho_{\hat{\lambda}} * \varphi_\lambda = 0. \tag{12.4}$$

Since $u_\tau = \Sigma \lambda \rho_\lambda$, we have

$$\frac{du_\tau}{d\tau} - \frac{\partial u_\tau}{\partial t} = \sum \lambda \left(\frac{d\rho_\lambda}{dt} - \frac{\partial \rho_\lambda}{\partial t}\right);$$

let, according to (12.2),

$$\frac{du_\tau}{d\tau} - \frac{\partial u_\tau}{\partial t} = [H,u_\tau]_* + \sum \lambda \varphi_\lambda.$$

It follows from (10.4) that

$$\sum \lambda \varphi_\lambda = 0,$$

and thus, according to (12.4),

$$\rho_\lambda * \sum \hat{\lambda} \varphi_{\hat{\lambda}} = \lambda \rho_\lambda * \varphi_\lambda = 0.$$

We obtain

$$u_\tau * \varphi_\lambda = \lambda \rho_\lambda * \varphi_\lambda = 0.$$

Therefore, we have, according to (12.3)

$$\lambda(t + \tau) \cdot \varphi_\lambda = 0. \tag{12.5}$$

(b) Substitute to the function $u_0 \in N$ the function $(u_0 + k) \in N$, where k depends only upon t. If u_τ is the solution of (10.4) taking the value u_0 at $\tau = 0$, we are led to substitute to u_τ the function $(u_\tau + k_\tau)$, where $k_\tau(t) = k(t + \tau)$. It follows from the lemma of subsection (10)(b) and from the theorem of subsection (11) that the spectrum of $(u_\tau + k_\tau)$ is deduced from the spectrum of u_0 by the change $\lambda(t) \to \lambda(t + \tau) + k(t + \tau)$. The eigenstates corresponding to u_τ and $(u_\tau + k_\tau)$ being the same, φ_λ given by (12.2) remains the same. We can choose k so that $(\lambda(t + \tau) + k(t + \tau)) \neq 0$ for a given $(t + \tau)$. It follows then from (12.5) that φ_λ vanishes necessarily. We have:

Theorem: Each eigenstate ρ_τ of u_τ satisfies the dynamical equation

$$\frac{d\rho_\tau}{d\tau} = \frac{\partial \rho_\tau}{\partial t} + [H,\rho_\tau]_*, \tag{12.6}$$

that is, (10.4).

(13) Invariants: Let $f \in N$ be a function satisfying the equation

$$\frac{\partial f}{\partial t} = [H,f]_* = 0. \tag{13.1}$$

The function $f_\tau = f$ is then a solution of (10.4) for each τ. Such a function is said to be *an invariant of considered quantum system*.

Let $\lambda(t)$ be an eigenvalue of the invariant f. It follows from the theorem of subsection (11) that we have for each τ

$$\lambda(t+\tau) = \lambda(t).$$

Therefore, all eigenvalue of f is necessarily constant. The corresponding eigenstate ρ satisfies, according to the theorem of subsection (12), the invariance equation

$$\frac{\partial \rho}{\partial t} + [H,\rho]_* = 0. \tag{13.2}$$

We have:

Corollary: The spectrum of each invariant of a quantum system is time-independent and each corresponding eigenstate satisfies Eq. (13.2).

Lewis[3] has disclosed interesting quantum invariants from the time-dependent harmonic oscillators. These invariants give a good information for the behavior of such oscillators.

Equation (13.2), that is,

$$\frac{\partial \rho}{\partial t} = \frac{i}{\hbar}(H*\rho - \rho*H) \tag{13.3}$$

appears as *the extension, in this framework, of the quantum Von Neumann equation*, for states described by invariant ρ's (see Ref. 7). The generalization of this equation is then Eq. (12.6).

(14) Interpretation for the dynamical systems with time-independent constraints: It is clear that *if u_τ, v_τ are solutions of (10.4), so is $u_\tau * v_\tau$.*

(a) We suppose now that *our dynamical system admits time-independent constraints* and a time-dependent Hamiltonian. We use the notations of subsection (9)(b),(c). We note by $\hat\eta$ the volume element of (M,F) defined by

$$\hat\eta = (2\pi\hbar)^{-n}(F^n/n!) \quad \text{(normalized Liouville measure)}.$$

We assume that our star product on (W,Λ,t) is such that if, for each t, the intersection of the supports of $u,v \in N$ restricted to the corresponding leaf is always compact, we have

$$\int_M [u,v]_* \hat\eta = 0. \tag{14.1}$$

More generally, we assume that (14.1) holds for suitable asymptotic conditions on v on each leaf. It is the case for a star product on (W,Λ,t) deduced from a Moyal product or from similar star products on (M,F). We have:

Lemma: If (14.1) holds and if $u_\tau, v_\tau, \cdots$ are solutions of Eq. (10.4), the quantities defined by

$$\langle u\rangle_{t,\tau} = \int_M u_\tau \hat\eta < \infty, \quad \langle u,v\rangle_{t,\tau} = \int_M (u_\tau * v_\tau)\hat\eta < \infty \cdots \tag{14.2}$$

depend only upon $(t+\tau)$.

In fact, we have for example

$$\frac{d\langle u\rangle_{t,\tau}}{d\tau} = \int_M \left(\frac{\partial u_\tau}{\partial t} + |H,u_\tau|_*\right)\hat\eta$$

and

$$\frac{\partial\langle u\rangle_{t,\tau}}{\partial t} = \int_M \frac{\partial u_\tau}{\partial t}\hat\eta.$$

It follows from (14.1) that

$$\frac{d\langle u\rangle_{t,\tau}}{d\tau} = \frac{\partial\langle u_{t,\tau}\rangle}{\partial t},$$

and the lemma is proved.

(b) The measured value $\bar u$ of the observable $u \in N$ at time t for a state ρ described by an element of N is in this dynamical framework[1]

$$\bar u(t) = \int_M (u*\rho)\hat\eta, \tag{14.3}$$

this formula being the extension of the classical Wigner formula (see Ref. 7) corresponding to the Moyal product. We note that, according to (14.1), we have

$$\int_M (u*\rho)\hat\eta = \int_M (\rho*u)\hat\eta$$

so that $\bar u(t)$ is real.

If u_τ and ρ_τ are solutions of (10.4) [see (12.6)], *the integral*

$$\bar u(t,\tau) = \int_M (u_\tau * \rho_\tau)\hat\eta$$

depends only upon $(t+\tau)$, according to the lemma. In particular, it is the case for an invariant state ρ satisfying (13.3): the measured value

$$\bar u(t,\tau) = \int_M (u_\tau * \rho)\hat\eta$$

depends only upon $(t+\tau)$.

[1] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. **111**, 61–151 (1978).

[2] A. Lichnerowicz, *Deformations and Quantization*, Lectures Notes in Mathematics 775 (Springer-Verlag, Berlin, 1979), pp. 105–121; *Quantum Theory, Groups, Fields and Particles*, edited by A. O. Barut (Reidel, Dordrecht, 1983), pp. 3–82.

[3] H. R. Lewis, J. Math. Phys. **9**, 1976–86 (1968); H. R. Lewis and W. B. Riesenfeld, J. Math. Phys. **10**, 1458–73 (1969).

[4] P. Camiz, A. Gerardi, C. Marchioro, E. Presutti, and E. Scacciatelli, J. Math. Phys. **12**, 2040-3 (1971).

[5] A. Munier, J. R. Burgan, M. Felix, and E. Fijalkow, J. Math. Phys. **22**, 1219–23 (1981).

[6] A. Lichnerowicz, Ann. Inst. Fourier Grenoble **32**, 157–209 (1982).

[7] S. de Groot, *La transformation de Weyl et la fonction de Wigner: une forme alternative de la mécanique quantique* (Presses Univ. de Montréal, Montreal, 1974).

[8] A. Lichnerowicz, J. Diff. Geom. **12**, 253–300 (1977); "Géométrie des transformations canoniques, un volume en l'honneur de J. Géhéniau," Bull. Soc. Math. Belgique **31**, 105–35 (1979).

[9]J. A. Schouten, *Convegno Internazionale Geom. Diff., Padova, 1953* (Cremonese, Rome, 1954).
[10]A. Nijenhuis, Indag. Math **17**, 390–403 (1955).
[11]O. M. Neroslavsky and A. T. Vlassov, C. R. Acad. Sci. Paris I **292**, 71–5 (1981).
[12]P. Lecomte and M. de Wilde, "Star-products on cotangent bundles," Lett. Math. Phys. (to be published).
[13]J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949); see also H. J. Groenewold, Physica **12**, 405–10 (1946).
[14]H. Basart and A. Lichnerowicz, C. R. Acad. Sci. Paris I **295**, 681–5 (1982).

Eigenvalues of anharmonic oscillators from a variational functional method

Gustavo A. Arteca, Francisco M. Fernández, and Eduardo A. Castro[a)]
INIFTA, Sección Química Teórica, Sucursal 4, Casilla de Correo 16, (1900) La Plata, Argentina

(Received 18 November 1982; accepted for publication 25 March 1983)

Analytic expressions for the eigenvalues of anharmonic oscillators are obtained from a simple variational method. Numerical results show striking accuracy in the whole range of the characteristic anharmonic oscillator parameters. The procedure also allows one to classify the set of states in three classes and yields a way of defining clearly the anharmonic and harmonic regimes.

PACS numbers: 03.65.Db, 03.65.Sq

I. INTRODUCTION

Recently, Rosen has shown[1] that eigenvalues associated with a set of oscillators can be obtained in an approximate way by minimization of a given functional. The extreme value of this functional satisfies both the virial theorem (VT) and the Hellmann–Feynman theorem (HFT). Rosen's results for the one-dimensional case are wholly coincident with those deduced by Orland[2] by means of a different method. Moreover, Fernández and Castro[3] have demonstrated that Rosen's method is applicable to more complex problems than those originally studied in Ref. 1.

The purpose of the present paper is to show that the variational functional method (VFM) is markedly useful for treating one-dimensional anharmonic oscillators. The paper is organized as follows. In Sec. II the variational functional (VF) is constructed by employing the pertinent Heisenberg inequality. The procedure permits us to know the amendments that may be introduced within Rosen's original method. In Secs. III and IV analytical expressions are derived for eigenvalues corresponding to the fourth- and sixth-order anharmonic oscillator, respectively. In spite of their extreme simplicity, the resulting formulas are strikingly accurate through the whole range of characteristic parameters. In Sec. V it is shown that one of the parameters introduced in the VF allows us to define in a very precise manner the harmonic and anharmonic regimes. Finally, in Sec. VI results are discussed and several possible applications of the VFM are analyzed.

II. THE METHOD

The study of $2k$-anharmonic oscillators

$$H = H^0 + \lambda x^{2k}, \quad H^0 = \tfrac{1}{2}(p^2 + x^2);$$
$$p = -i\frac{d}{dx}, \; k = 2, 3, \ldots \tag{1}$$

has received considerable attention lately, owing to their occurrence in several different contexts in quantum theory.[2] Here our purpose is to discuss this problem via the VFM, and the first step is to build an appropriate VF. A possible way is to apply the Heisenberg inequality

$$(\Delta p)^2(\Delta x)^2 = 2A \geqslant \tfrac{1}{4}, \tag{2}$$

where A is a constant depending on the quantum number n and the parameter λ. Taking into account that $\langle p \rangle = \langle x \rangle = 0$, Eq. (2) permits us to write the energy formula

$$E = \langle H \rangle = \epsilon(q) \equiv Aq^{-2} + \tfrac{1}{2}q^2 + F(q), \tag{3}$$

where

$$q^2 \equiv \langle x^2 \rangle \quad \text{and} \quad F(q) = \lambda \, \langle x^{2k} \rangle. \tag{4}$$

We are using abbreviated notation, and it is important to remember that $F(q)$ depends on λ and n. If the VF (3) fulfills the VT

$$\langle p^2 \rangle = \langle x^2 \rangle + 2k\lambda \, \langle x^{2k} \rangle, \tag{5}$$

then the equality

$$q^2 - 2Aq^{-2} = q\frac{\partial \epsilon}{\partial q} - q\frac{\partial F}{\partial q} = -2kF \tag{6}$$

must be satisfied. Equation (6) is obeyed at once if q and F are chosen as follows:

$$\frac{\partial \epsilon}{\partial q}(q = q^*) = 0, \quad E^* = \epsilon(q^*), \tag{7}$$

$$F(q) = \lambda B q^{2k}, \tag{8}$$

with B depending on λ and n. Moreover, if one assumes the forms (7) and (8), then the HFT

$$\frac{\partial E}{\partial \lambda} = \langle x^{2k} \rangle \tag{9}$$

allows us to relate the constants A and B through the relation

$$\frac{\partial A}{\partial \lambda} = -\lambda \frac{\partial B}{\partial \lambda} q^{*2k+2}. \tag{10}$$

The latter equation follows from (9) and the fact that E^* satisfies[3]

$$\frac{\partial E^*}{\partial \lambda} = \left(\frac{\partial \epsilon}{\partial q}\right)_\lambda (q = q^*)\frac{\partial q^*}{\partial \lambda} + \left(\frac{\partial \epsilon}{\partial \lambda}\right)_q (q = q^*)$$
$$= \left(\frac{\partial \epsilon}{\partial \lambda}\right)_q (q = q^*). \tag{11}$$

The most immediate way to satisfy Eq. (10) is to choose A and B to be independent of λ, and such a first approximation yields acceptable results.[4] However, our aim is to go a step further, and it is first necessary to delineate the required modifications to the Rosen[1] and Orland[2] methods.

The semiclassical approximation

$$\langle x^{2k} \rangle \simeq \langle x^2 \rangle^k,$$

[a)] To whom correspondence should be addressed.

 0022-2488/84/040932-07$02.50

invoked by both authors leads us immediately to $B = 1$, which means that a unique constant A remains in the VF. Orland[2] adjusted properly the value of this constant with the aim that E^* should reproduce the exact eigenvalue in the limit of large λ values. Afterward, two of us[3] showed that if A is obtained from the limit $\lambda \to 0$, then E^* presents a near correct dependence with n. This fact suggests that it is very convenient to adjust A from the correct result for $\lambda \to 0$ and B from the limit $\lambda \to \infty$.[4] As stated previously, our present purpose is to consider the variation of A and B with λ. However, this engenders the primary impasse of satisfying Eq. (10). The problem is simplified to a considerable extent when only B depends on λ. In order to achieve this simplification, it is necessary to redefine $\langle x^2 \rangle$ and $\langle x^{2k} \rangle$ in terms of A, B, and q in a manner consistent with the virial theorem (5) and the HFT (9). If we assume that A does not depend on λ, the functional built from Eqs. (3), (7), and (8), when being differentiated with regard to λ, allows us to obtain from the HFT [Eqs. (9) and (11)]

$$\frac{\partial E^*}{\partial \lambda} = q^{*^{2k}} \frac{\partial}{\partial \lambda}(\lambda B) = \langle x^{2k} \rangle. \tag{12}$$

Equation (12) makes up a new definition for $\langle x^{2k} \rangle$, instead of Eqs. (4) and (8). In a similar way, we can deduce an expression for $\langle x^2 \rangle$ within the context of this actual formulation where only B depends on λ. Considering that the VT (5) is satisfied by the VF

$$\langle H \rangle = \langle x^2 \rangle + (k+1)\lambda \langle x^{2k} \rangle = E^* = q^{*^{2k}} + (k+1)\lambda B q^{*^{2k}}, \tag{13a}$$

we can combine Eqs. (12) and (13a), and the result is

$$\langle x^2 \rangle = q^{*^2} - (k+1)\lambda^2 q^{*^{2k}} \frac{\partial B}{\partial \lambda}. \tag{13b}$$

This new equation must replace Eq. (4).

Since in this scheme A does not depend on λ, its value may be calculated from the limit $\lambda \to 0$:

$$E^*(\lambda = 0) = E(\lambda = 0),$$

so that

$$A = \tfrac{1}{2}(n + \tfrac{1}{2})^2. \tag{14}$$

The dependence of B with λ and n is discussed in the sections below.

III. 4-ANHARMONIC OSCILLATORS

According to our previous discussion, the eigenvalues E^* are given by (13a), q^* being the solution of the equation

$$(n + \tfrac{1}{2})^2 = q^{*^4} + 2k\lambda B q^{*^{2k+2}}. \tag{15}$$

The quantities E^*, B, and q^* are expressible as a λ power series for sufficiently small values of λ:

$$E^* = \sum_{i=0}^{\infty} E_i \lambda^i, \quad B = \sum_{i=0}^{\infty} B_i \lambda^i, \quad q^* = \sum_{i=0}^{\infty} q_i^* \lambda^i. \tag{16}$$

The correction terms B_i are deduced immediately when one substitutes (16) in (15). After some algebraic manipulations, we get

$$B_0 = E_1 / E_0^k, \tag{17a}$$

$$B_1 = \{2E_2 + k^2 E_1^2 / E_0\} / (2E_0^k). \tag{17b}$$

In order to obtain good results from the VF procedures in the small regime, it is necessary that the coefficients E_i coincide with the energy perturbative corrections. Then Eq. (17) determines the B_0 and B_1 values, since E_0, E_1, and E_2 may be calculated from perturbation theory.

A very interesting fact of this analysis is the coincidence of B_0 and B_1 with the first two terms of the expansion of $\langle x^{2k} \rangle / \langle x^2 \rangle^k$ in the λ power series. This fact reveals, in a transparent manner, that (i) the term $\lambda(\partial B / \partial \lambda)$ introduces high-order corrections, (ii) the new definition of $\langle x^{2k} \rangle$ and $\langle x^2 \rangle^k$ [Eqs. (12) and (13b)] differs to a very small extent from the first one [Eqs. (4) and (8)] in the small λ regime, i.e., B approaches $\langle x^{2k} \rangle / \langle x^2 \rangle^k$ up to the first-order in λ. When $k = 2$, Eqs. (17) gives

$$B_0 = \tfrac{3}{2}(1 + (2n+1)^{-2}), \tag{18a}$$

$$B_1 = \tfrac{1}{8}(2n+1)^{-3}((2n+1)^4 - 31(2n+1)^2 + 18). \tag{18b}$$

Considering only the first term in B, the VF takes the form

$$\epsilon(q) = (n + \tfrac{1}{2})^2 / 2q^2 + q^2/2 + \lambda \langle x^{2k} \rangle_0 q^{2k} / \langle x^2 \rangle_0^k, \tag{19}$$

where $\langle \ \rangle_0$ represents the average value calculated with an eigenstate of H^0. The change of variable $a^2 = \langle x^2 \rangle_0 / q^2$ transforms Eq. (19) in

$$\epsilon = \tfrac{1}{2}(n + \tfrac{1}{2})(a^2 + a^{-2}) + \lambda \langle x^{2k} \rangle_0 a^{2k}. \tag{20}$$

This last expression coincides with the H expectation value calculated with the H^0 eigenfunctions scaled with the parameter a. As a consequence, we can consider the scaling variational method[5] as a particular case of the VFM when $B = B_0$. It is convenient to substitute (16) by series expansions in $\lambda^{-2/(k+1)}$ in order to study the large λ value regime:

$$E^* = \lambda^{1/(k+1)} \sum_{i=0}^{\infty} e_i \lambda^{-2i/(k+1)}, \tag{21a}$$

$$B = \sum_{i=0}^{\infty} b_i \lambda^{-2i/(k+1)}. \tag{21b}$$

The replacement of (21) in (15) gives the following results for the first two coefficients:

$$b_0 = (e_0/(k+1))^{k+1} (2k(n + \tfrac{1}{2})^{-2})^k, \tag{22a}$$

$$b_1 = (2ke_0(k+1)^{-1}(n + \tfrac{1}{2})^{-2})^k \times (e_1 - (k+1)(n + \tfrac{1}{2})^2 / (4ke_0)). \tag{22b}$$

Our next problem is circumscribed to determine the e_i coefficients. This is the most troublesome aspect of the method because the eigenfunctions of the $2k$-oscillators (i.e., $H = \frac{1}{2}p^2 + \lambda x^{2k}$) are not known.

We can use the WKB method for large enough n values:

$$e_0(\text{WKB}) = c_1 (n + \tfrac{1}{2})^{2k/(k+1)}, \tag{23a}$$

$$c_1 = 2^{(k-2)/(k+1)} \left\{ (k+1)\pi \Gamma\left(\frac{1}{k}\right) \Gamma\left(\frac{1}{2k}\right)^{-2} \right\}^{2k/(k+1)}, \tag{23b}$$

$$e_1 = c_2 (n + \tfrac{1}{2})^{2/(k+1)}, \tag{23c}$$

$$c_2 = 2^{-(k-2)/(k+1)}\left\{(k+1)\pi\Gamma\frac{1}{k}\Gamma\left(\frac{1}{2k}\right)^{-2}\right\}^{2/(k+1))}$$
$$\times\Gamma\left(\frac{1}{k}\right)\Gamma\left(\frac{3}{2k}\right)^2\Gamma\left(\frac{1}{2k}\right)^{-2}\Gamma\left(\frac{3}{k}\right)^{-1}. \quad (23d)$$

It is necessary to introduce some modifications in the e_i coefficients to obtain good results for the low-lying eigenvalues. For example,[6,7]

$$e_0 = c_1\{(n+\tfrac{1}{2}) + d_n/(n+\tfrac{1}{2})\}^{2k/(k+1)} \quad (24)$$

allows one to obtain acceptable results, provided d_n is known. For the 4-anharmonic oscillator, we obtain

$$b_0 = (16c_1^3/27)(1 + 4d_n/(2n+1)^2)^4, \quad (25a)$$
$$b_1 = (2b_0)^{2/3}(4^{1/3}c_2 - 1/(2b_0^{1/3})). \quad (25b)$$

In the case of small quantum numbers, coefficients c_1 and c_2 can be deduced by means of a $\lambda^{-2/(k+1)}$ series expansion[6,7] but in the large n regime we can directly apply Eqs. (23b), (23d), and $d_n = 0$. It is convenient to analyze the behavior of $B = B(n, \lambda)$ as a function in the variables n and λ before continuing with its asymptotic development. The exact B value for each λ and n may be obtained from (15) whenever E^* is replaced by the corresponding eigenvalue. The curves shown in Figs. 1–3 were plotted using the results presented in Ref. 8.

The different eigenstates associated with the $2k$-oscillators can be classified as three different types according to the shape of the curves $B_n(\lambda)$ vs $\log\lambda$.

Type I: $B_n(\lambda)$ is a monotonic decreasing function ($B_1 < 0$, $b_1 > 0$).

Type II: $B_n(\lambda)$ presents a minimum ($B_1 < 0$, $b_1 < 0$). These states may be considered as a transition class between types I and III, and they are grouped in two subtypes corresponding to $B_0 > b_0$ (type IIA) and $B_0 < b_0$ (type IIB).

Type III: $B_n(\lambda)$ is a monotonic increasing function ($B_1 > 0$, $b_1 < 0$).

The plots $B_n(\lambda)$ vs $\log\lambda$ present an inflection point in all cases. Their significance will be discussed at length in Sec. V. Our results show that the states change from type I toward type III as n increases. For the 4-anharmonic oscillator, $n = 0$, $n = 1$ correspond to type I (Fig. 1), $n = 2$ to type IIB (Fig. 2), and $n > 2$ to type III (Fig. 3).

In what follows, we will try to find a simple form for the $B_n(\lambda)$ which makes allowance for the characteristics just discussed.

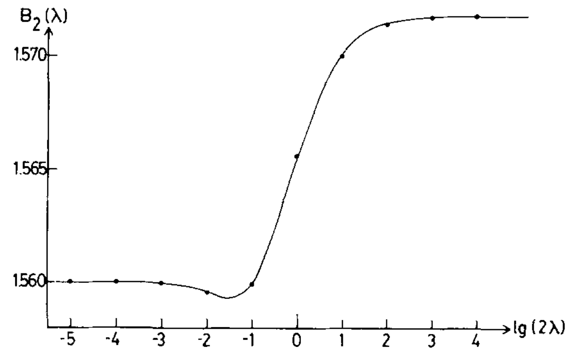


FIG. 2. Function $B_n(\lambda)$ for the $n = 2$ state of the 4-anharmonic oscillator. Full circles correspond to exact results.

The following equation is a very satisfactory expression:

$$B_n(\lambda) = b_0 + (B_0 - b_0)\{C_n(z)z + 1\}^{-2/3}, \quad z = \lambda(n + \tfrac{1}{2}). \quad (26)$$

This last formula does not give a $\lambda^{-2/3}$ power expansion, but it properly yields the first two terms and it is adequate for our actual purposes. Expanding Eq. (26) in a z series expansion and comparing the result with that formerly obtained, we deduce an expression for $C_n(0)$:

$$C_n(0) = \tfrac{3}{8}(b_0 - B_0)^{-1}(2n+1)^{-4}$$
$$\times((2n+1)^4 - 31(2n+1)^2 + 18). \quad (27)$$

A similar treatment of the $z^{-2/3}$ expansion gives

$$C_n(\infty) = (2b_0)^{-1}$$
$$\times((B_0 - b_0)(4^{1/3}c_2 - 1/(2b_0^{1/3}))^{-1})^{3/2}. \quad (28)$$

$C_n(z)$ can be approximated as

$$C_n(z) = C_n(0) + (C_n(\infty) - C_n(0))f(z),$$
$$0 \leqslant f(z) \leqslant 1, \quad \frac{\partial f}{\partial z} > 0 \quad (29)$$

between both limits.

The calculations show that $C_n(0)$ is greater than zero for every state except $n = 2$. This difference is consistent with the fact that the state $n = 2$ belongs to type II. Actually, the shape of $f(z)$ is not very important because even though $C(z)$ is chosen as a constant, the VF correctly reproduces the be-

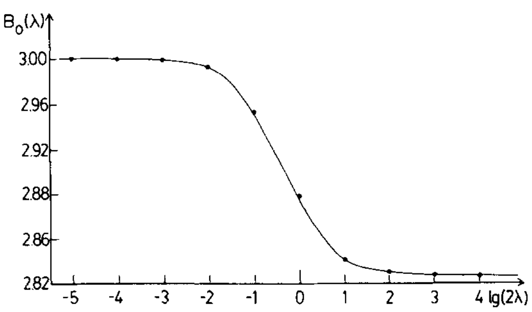


FIG. 1. Function $B_n(\lambda)$ for the ground state ($n = 0$) of the quartic anharmonic oscillator. The proposed analytical expression for $B_n(\lambda)$ (full line) is compared with the exact results (circles).

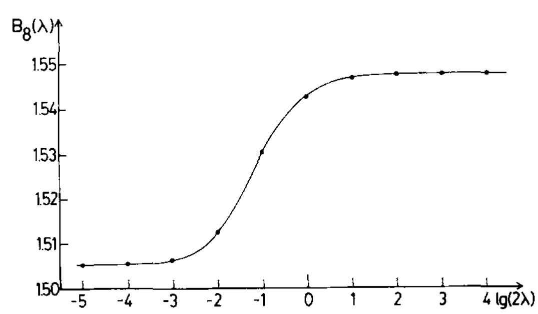


FIG. 3. Function $B_n(\lambda)$ for the $n = 8$ state of the 4-anharmonic oscillator. Full circles correspond to exact results.

havior of the eigenvalues. The only care that must be taken into account is that the introduction of $f(z)$ does not change the form of the B power series expansions in both regimes. A very convenient form for $f(z)$ is

$$f(z) = (z/(1+z))^{2/3}. \tag{30}$$

Nevertheless, for the $n = 2$ state of the 4-anharmonic oscillator (type IIB state), the best results are obtained when

$$f(z) = az/(1+az), \quad a \simeq 35. \tag{31}$$

This function adjusts the minimum of $B_n(\lambda)$ better than Eq. (30). As a general rule, those states belonging to types I and III of any $2k$-anharmonic oscillator can be approximated by way of the general function

$$B_n(\lambda) = b_0 + (B_0 - b_0)(zC_n(z) + 1)^{-2/(k+1)},$$
$$z = \lambda(n+\tfrac{1}{2})^{k-1}, \tag{32}$$

and substituting in Eq. (29)

$$f(z) = (z/(z+1))^{2/(k+1)}. \tag{33}$$

In the next section we will show that, in general, the B curves for type II states may be obtained as linear combinations of B curves for the adjacent states belonging to types I and III. This property illustrates the previous statement: States of type II make up a transition between states of type I and III. It is not necessary to apply the present procedure for the $n = 2$ state because function (31) already yields very good results (see Fig. 2).

Figures 1, 2, and 3 display the curves $B_0(\lambda)$, $B_2(\lambda)$, and $B_8(\lambda)$, respectively, which were obtained from Eq. (26). It can be seen that these curves fit the exact results (broken lines) very well. The foregoing discussion has made certain that in order to use the method for an anharmonic oscillator, it is necessary to previously find those states belonging to type II. In its turn, this may easily be done by studying the coefficients (18) and (22). For the last case, it is sufficient to employ the WKB results.

We have employed e_0(WKB) and e_1(WKB) for those states with $n \geqslant 10$, and the values e_0, e_1, and d_n from Refs. 6 and 7 when $n < 10$ in our calculations performed for the 4-anharmonic oscillator. In Table I we present the eigenvalues, obtained to great accuracy by Banerjee *et al.*,[8] together with our results and those given in Ref. 9. These last results were computed by means of an analytical expression which, the same as ours, employs adjustable parameters. Our results are better than those presented in Ref. 9 for almost the whole interval of λ and n values. The largest error in our eigenvalues is localized in the intermediate zone between the harmonic and anharmonic regime. In this zone, our formula makes certain just two decimal places, but it is far superior to that given in Ref. 9, where the numerical results are manifestly poorer than ours.

IV. 6-ANHARMONIC OSCILLATORS

In this section, we consider the case $k = 3$ with the purpose of verifying the suitability of the VFM to treat any $2k$-anharmonic oscillator. Since the procedure is closely related to that offered previously, here we restrict ourselves to a very succinct presentation. First, it may be seen at once that in

TABLE I. Eigenvalues of the 4-anharmonic oscillator.

n \ 2λ	10^{-4}	1	10^{4}
0	1.000 074 987[a]	1.392 35	22.861 609
	1.000 074 900[b]	1.392 03	22.861 608
	1.000 074 988[c]	1.397 8	23.09
1	3.000 374 90	4.648 81	81.903 317
	3.000 374 90	4.648 6	81.903 317
	3.000 374 91	4.670	82.57
2	5.000 974 62	8.655 0	160.685 91
	5.000 974 61	8.654 4	160.685 90
	5.000 974 6	8.663	160.81
5	11.004 571 36	23.297 4	457.664 57
	11.004 571 35	23.297 1	457.664 59
	11.004 571 40	23.299 8	457.677
10	21.016 550 25	53.449 1	1 082.888 52
	21.016 550 26	53.448 4	1 082.888 59
	21.016 550	53.450	1 082.891
100	202.494 080	1 035.544 18	21 997.240 27
	202.494 076	1 035.539	21 997.163 1
	202.494 10	1 035.544	21 997.240 49
1000	2134.242 5	21 932.783 71	471 075.928 38
	2134.247	21 932.781	471 075.916
	2134.252	21 932.783 86	471 075.928 46

[a] Exact results from Ref. 8.
[b] Present calculation: $2E^*$.
[c] Results from Ref. 9.

this case, the minimum condition (15) admits an exact solution

$$q^4 = ((\tfrac{1}{4} + 12\lambda AB)^{1/2} - \tfrac{1}{2})/(6\lambda B). \tag{34}$$

Then we obtain the explicit energy formula

$$E^* = \frac{((\tfrac{1}{4} + 12\lambda AB)^{1/2} - \tfrac{1}{2})^{1/2}(1 + 2((\tfrac{1}{4} + 12\lambda AB)^{1/2} - \tfrac{1}{2})/3)}{(6\lambda B)^{1/2}}, \tag{35}$$

which simplifies the calculation procedure to a great extent. The analysis of coefficients (17) and (22) for $k = 3$ makes certain that the states admit a classification according to the three types formerly introduced: $n < 2$, type I; $n = 2$, type IIA; $n = 3$, type IIB; and $n > 3$, type III.

From the power series expansion for E,

$$E \simeq (n+\tfrac{1}{2}) + \lambda(\tfrac{25}{4}(n+\tfrac{1}{2}) + \tfrac{5}{2}(n+\tfrac{1}{2})^3)$$
$$- \tfrac{1}{2}\lambda^2(\tfrac{19\,277}{128}(n+\tfrac{1}{2}) + \tfrac{4145}{16}(n+\tfrac{1}{2})^3 + \tfrac{393}{8}(n+\tfrac{1}{2})^5), \tag{36}$$

we deduce B_0 and B_1.

$$B_0 = \tfrac{5}{2}(1 + 5(2n+1)^{-2}), \tag{37a}$$

$$B_1 = (n+\tfrac{1}{2})^{-3}(\tfrac{57}{16}(n+\tfrac{1}{2})^5 - \tfrac{237}{8}(n+\tfrac{1}{2})^3 - \tfrac{8027}{256}(n+\tfrac{1}{2})). \tag{37b}$$

Besides, the expansion of E in the $\lambda^{-1/2}$ power series permits us to obtain b_0 and b_1 from Eq. (22):

$$b_0 = \tfrac{1}{6}(\tfrac{3}{2}c_1)^4(1 + 4d_n(2n+1)^{-2})^6,$$
$$\tfrac{1}{6}(\tfrac{3}{2}c_1)^4 = 2.775\,713\cdots, \tag{38a}$$
$$b_1 = (6b_0)^{3/4}(c_2 - \tfrac{1}{2}(6b_0)^{-1/4}). \tag{38b}$$

The value of the c_2 coefficient for every state may be obtained from Ref. 7. For $n > 5$ we use Eq. (22) and the WKB results (23). In a similar fashion as in the preceding example, let us write

write

$$B_n(\lambda) = b_0 + (B_0 - b_0)(zC_n(z) + 1)^{-1/2},$$
$$z = \lambda(n + \tfrac{1}{2})^2,\ n \neq 2, 3. \tag{39}$$

This expression is valid for all the states belonging to types I and III. Type II states will be discussed later.

A straightforward calculation gives

$$C_n(0) = 2(\tfrac{57}{16} - \tfrac{237}{4}(n + \tfrac{1}{2})^{-2} - \tfrac{8027}{256}(n + \tfrac{1}{2})^{-4}) \times (b_0 - B_0)^{-1}, \tag{40a}$$

$$C_n(\infty) = \tfrac{2}{3} b_0^{-1}((B_0 - b_0)(2c_2(6b_0)^{1/4} - 1)^{-1})^2. \tag{40b}$$

Furthermore, according to Eqs. (29) and (33), we know that

$$C_n(z) = C_n(0) + f(z)(C_n(\infty) - C_n(0)),\quad n \neq 2, 3, \tag{41a}$$

where

$$f(z) = (z/(1 + z))^{1/2}. \tag{41b}$$

For type II states, we write B as follows:

$$B_n(\lambda) = g_{n1}B_1(\infty) + g_{n2}B'(1, \lambda) + g_{n3}B_4(\infty) + g_{n4}B'(4, \lambda),\quad n = 2, 3, \tag{42a}$$

where

$$B'(4, \lambda) = (B_4(0) - B_4(\infty))(C'_4(z_4)z_4 + 1)^{-1/2}, \tag{42b}$$

$$z_i = \lambda(i + \tfrac{1}{2})^2, \tag{42c}$$

$$C'_4(z) = C_4(0) + (gC_4(\infty) - C_4(0))f(z), \tag{42d}$$

$$B'_1(1, \lambda) = (B_1(0) - B(\infty))(C_1(z_1)z_1 + 1)^{-1/2}, \tag{42e}$$

$$C_1(z_1) = C_1(0) + (C_1(\infty) - C_1(0))f(z_1). \tag{42f}$$

The coefficients g_{nj} and g are adjusted in such a way that $B_n(\lambda)$ correctly reproduces the first two terms in the B expansion as a z and $z^{-1/2}$ power series. After some algebraic manipulations we obtain

$$g = 0.548\,943\,017\,5,\quad g_{21} = 1,$$
$$g_{22} = 1.519\,682\,804,\quad g_{23} = -0.125\,092\,432,$$
$$g_{24} = 5.028\,711\,824. \tag{43}$$

The function $B_2(\lambda)$ obtained in this manner presents a minimum excessively deep, but this fact does not influence, to a large extent, the results (Fig. 4). This property shows clearly that the fitting of B is not decisive in the obtainment of acceptable results. In fact, in previous works it was possible to reproduce perfectly the behavior of the eigenvalues by introducing just one[1-3] or two[4] adjustable parameters within the VF (i.e., A and B were considered independent on λ). In very general terms, we can state that the main dependence of the VF on λ and n is assured by the VT and the HFT. The consideration of a variable B only introduces high-order corrections. The eigenvalues of the 6-anharmonic oscillator, calculated via the VFM, are displayed in Table II together with the exact results[8] and those reported by Mathews *et al.*[9] The conclusion, derivable from these results, is similar to those presented previously in Sec. III, with reference to Table I. However, we deem it appropriate to point out that our results are less accurate than those obtained previously. This is due to the fact that the larger k is, the less is the range of validity of the polynomials (16) and (21). This property notably influences the treatment of those states belonging to type II, especially when these polynomials are not valid around the minimum.

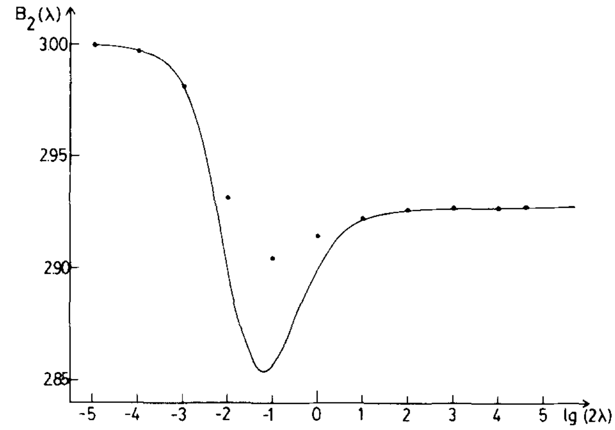


FIG. 4. Type IIA state for the 6-anharmonic oscillator. Approximate analytic expression for $B_2(\lambda)$ (full line) is compared with exact results (dots).

V. DEFINITION OF HARMONIC AND ANHARMONIC REGIMES FROM THE B FUNCTION

Undoubtedly, an appropriate definition of the harmonic and anharmonic regimes for the $2k$-anharmonic oscillator models possesses by itself great theoretical interest. Owing to this fact, the problem has received considerable attention from various authors.[6-8,10-12] Hioe *et al.*[6,7,10] defined the harmonic regime as that range of λ-values for which the energy levels differ from the harmonic ones by less than 10%, and the $2k$ regime as that range of λ-values for which energy levels differ from the $2k$ ones by less than 10%. Besides, they have shown that it is more convenient to employ z as a pa-

TABLE II. Eigenvalues of the 6-anharmonic oscillator.

n \ 2λ	10^{-4}	1	10^4
0	1.000 187 228[a]	1.435 624 62	11.478 798 04
	1.000 187 228[b]	1.433 4	11.476 4
	1.000 187 34[c]	1.481	12.18
1	3.001 308 84	5.033 395 94	43.457 784 74
	3.001 308 84	5.032 2	43.457 80
	3.001 310	5.15	44.98
2	5.004 664 71	9.966 622	90.821 278 91
	5.004 664 65	9.957 3	90.821 245
	5.004 668 0	10.04	91.56
10	21.277 841 68	78.958 068 58	771.456 964 19
	21.277 829	78.955 8	771.451 6
	21.278 1	78.963	771.474
100	283.323 081 57	2 287.793 566	22 821.749 929
	283.368	2 287.795 4	22 821.777
	283.340	2 287.793 74	22 821.750 0
1000	7346.838 298 7	71 700.060 494	716 823.285 169
	7346.36	71 696.3	716 789.6
	7346.850	71 700.060 51	716 823.285 171

[a] Exact results from Ref. 8.
[b] Present calculations: $2E^*$. The parameters used in $n = 1000$ were obtained from the limit $n \to \infty$.
[c] Results from Ref. 9.

rameter instead of λ. Here we will refer to the $2k$ regime simply as anharmonic regime. On the other hand, Kesarwani and Varshni[11,12] proposed the function

$$0<\alpha=\lambda^{2/3}/(1+\lambda^{2/3})<1 \tag{44}$$

as a measure of the anharmonicity degree of the 4-anharmonic oscillator, setting up $\lambda = 1$ as an intermediate regime. But as n increases, the same happens with the quotient $\langle x^{2k}\rangle/\langle x^2\rangle$, and the harmonic regime is shifted towards smaller λ values. This behavior is a plain confirmation that the suitable variable is z and not λ. Obviously, the function (44) will not be acceptable, unless it is writen in terms of z.

Our aim in this section is to propose a new definition of the different regimes, basing such a definition on the properties of the function $B_n(\lambda)$. This quantity is approximately a function dependent on z and takes finite values in both limits ($\lambda\to 0$ and $\lambda\to\infty$). In a broad sense, $B_n(\lambda)$ behaves as (44) if λ is replaced by z. The comparison of the plots[7] E vs $\ln\lambda$ with B vs $\ln\lambda$ (Figs. 1–4) shows clearly that the latter define the three regions (harmonic, intermediate, and anharmonic regimes) with greater neatness than the former. The B curves possess an inflection point in $z^* = \lambda^*(n+\frac{1}{2})^{k-1}$ given by

$$z^*\frac{\partial B}{\partial z}(z^*) = -z^{*2}\frac{\partial^2 B}{\partial z^2}(z^*). \tag{45}$$

Substituting (32) in (45) we obtain, in an approximate fashion, the relationship

$$z^* = \text{const}. \tag{46}$$

If we use this equation to define the transition among different regimes, we can appreciate that the harmonic regime is shifted towards small λ values as n and k increase. In Fig. 5 we display the limits of both regimes according to the definition of Hioe *et al.*[7] and our inflection points for the 4-anharmonic oscillator. Our results approximately satisfy the relation $z^* = 0.864$. This result is in line with the conclusions of Banerjee *et al.*[8] about the changes of regime, which were deduced from scaling considerations.

In closing this section, we consider it appropriate to point out that it is not mandatory to resort to Eq. (45) to

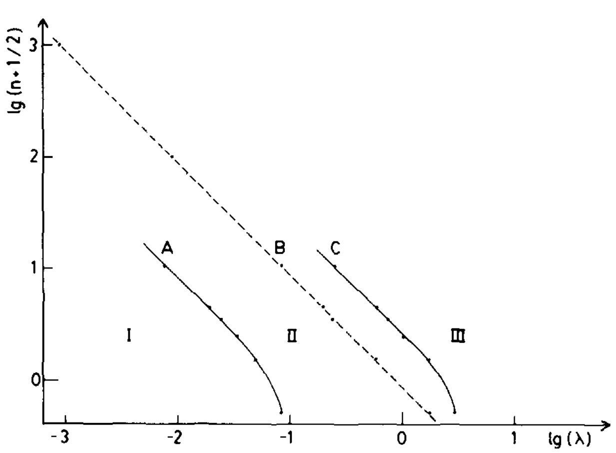


FIG. 5. Quadratic regime (I), boundary layer (II) and quartic regime (III) defined as in Ref. 7 by means of the curves A and C. Present inflection point calculations lay on curve B.

obtain λ^*, because the relationship

$$B_n(\lambda^*) = \tfrac{1}{2}(B_0 + b_0) \tag{47}$$

is fulfilled approximately. For type II like states with a very deep minimum, limits of the regimes may be very doubtful, but in the remaining states, they are neatly defined and can be seen without any difficulty.

VI. FURTHER COMMENTS AND CONCLUSIONS

The results presented in the preceding sections show clearly that the VFM is very useful to build analytical expressions associated with the eigenvalues of $2k$-anharmonic oscillators. The VFM by itself is not sufficient to obtain such eigenvalues, but we have seen that the method takes profitable advantage and efficiently uses the information regarding those systems. As a general rule, we can state that the VFM is a powerful tool to treat those quantum-mechanical systems that possess two extreme regimes. Whenever the information about these limits is properly introduced in the VF, it enables us to relate them in a correct way.

In the examples discussed in the present work, we have adjusted the parameters of the VF by means of the perturbation theory (small λ regime) and the WKB approximation (large λ regime). The eigenvalues obtained in this manner yield an acceptable result, even in the intermediate zone about which we have no information.

Time and again, theoretical physics must study quantum-mechanical models about which the eigenvalues and eigenfunctions are known for two extreme values of a certain parameter, but ignores the functional dependence of such eigenvalues on that parameter, which in turn prevents the relating of both results. The VFM would be of great help in these cases. A very interesting and illuminating example is a Rydberg atom in the presence of a magnetic field.[13] We have obtained very good results and they will be published elsewhere in a forthcoming paper. The fitting of the functions $B_n(\lambda)$ deserves, by itself, a brief comment. The functions proposed previously in Secs. III and IV are not essential to apply the method, and probably, there are simple enough relationships that could afford better results than the present ones. The point, significantly important, rests upon the fact that B is determined from power series expansions in both regimes. Besides, it is clear that it is much more convenient to approximate B and then to obtain E than to approach E directly. There are two principal reasons to support this assertion.

Firstly, B remains bounded in the whole range of (λ, n) values, which obviously does not occur with E. It allows us to adjust the B plots without any extra difficulty by means of two points $[N/N]$ Padé approximants.[14]

Second, it is worthwhile to note that the error introduced in the approximation of B is modulated and markedly reduced by the extremum condition imposed on the VF (that is to say, by the VT and HFT). For example, the function proposed in Sec. IV for $B_2(\lambda)$ gives too deep a minimum. However, results obtained for E are satisfactory and rather good.

Finally, we deem it appropriate to point out that the VFM may be applied without further difficulties to more complex problems than those just presented here. Recently, we have analyzed,[4] with particular success, several central field problems, bounded systems, and the hydrogen atom in a magnetic field. In all cases, we have used a constant value for B, so that the treatment presented in this paper may be considered as an improvement of previous methods.

[1]G. Rosen, Phys. Rev. A **20**, 1287 (1979).
[2]H. Orland, Phys. Rev. Lett. **42**, 285 (1979).
[3]F. M. Fernández and E. A. Castro, Phys. Rev. A **27**, 2735 (1983).
[4]F. M. Fernández, G. A. Arteca, and E. A. Castro (submitted).
[5]F. M. Fernández and E. A. Castro, Phys. Rev. A **27**, 663 (1983).
[6]F. T. Hioe and E. W. Montroll, J. Math. Phys. **16**, 1945 (1975).
[7]F. T. Hioe, D. MacMillen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).
[8]K. Banerjee, S. P. Bhatnagar, V. Choudhry, and S. S. Kanwal, Proc. R. Soc. London, Ser. A **360**, 575 (1978).
[9]P. M. Mathews, M. Seetharaman, S. Raghavan, and V. T. A. Bhargava, Pramāna **17**, 121 (1981).
[10]F. T. Hioe, D. MacMillen, and E. W. Montroll, Phys. Rep. **43**, 307 (1978).
[11]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **22**, 1983 (1981).
[12]R. N. Kesarwani and Y. P. Varshni, J. Math. Phys. **23**, 803 (1982).
[13]R. H. Garstang, Rep. Prog. Phys. **40**, 105 (1977).
[14]G. A. Baker, *Essentials of Padé Approximants* (Academic, New York, 1975).

Harmonic functions and matrix elements for hyperspherical quantum field models

R. E. Cutkosky
Physics Department, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213

(Received 4 October 1983; accepted for publication 18 November 1983)

Scalar and vector harmonic functions on the surface of a four-dimensional sphere are discussed. These functions are of use for numerical calculations in a hyperspherical formulation of quantum gauge field models. Formulas are derived for derivatives and integrals involving these functions.

PACS numbers: 03.65.Fd, 02.20. + b, 02.30.Bi, 11.15. − q

I. INTRODUCTION

In a recent paper,[1] a new Hamiltonian method was introduced for numerical calculations involving quantum gauge field models. The spatial domain was taken to be the surface of a four-dimensional hypersphere; this hyperspherical formalism has several convenient features, in particular, numerical calculations have manifest rotational invariance in all degrees of approximation. Also, group-theoretical methods involving the O(4) symmetry group can be used to classify states and evaluate matrix elements. A Lagrangian formulation on the surface of a five-dimensional hypersphere, given by Adler,[2] has many similar features. Although the properties of O(4) are widely known, and although many useful results exist in the literature,[3–5] most of the specific formulas needed for numerical calculations have not been available. In the present work, scalar and vector harmonics are constructed explicitly, and derivative formulas and integrals involving products of three of these functions are considered. Such integrals can be written as a product of two 3-j symbols and a reduced matrix element. Expressions for these reduced matrix elements are derived.

1.1. Notation

Let $\hat{r} = (x, y, z, w)$ denote a point on the hypersphere, which is taken to have a unit radius. If the hypersphere is parametrized in the usual way by polar angles, the scalar harmonic functions can be expressed in terms of the familiar Gegenbauer polynomials. For the present calculations, however, it is more convenient to use the following parametrization[3,4,6]:

$$\xi = x + iy = \cos\theta e^{i\alpha},$$
$$\zeta = z + iw = \sin\theta e^{i\beta}. \tag{1}$$

The volume element, in terms of these variables, is

$$d\Omega = d\alpha\, d\beta \sin\theta \cos\theta\, d\theta. \tag{2}$$

Let $M_{ij} = -M_{ji}$ denote the O(4) rotation operators, and let

$$J_i = \epsilon_{ijk} M_{jk} = L_{1i} + L_{2i},$$
$$K_i = M_{i4} - M_{4i} = L_{1i} - L_{2i}, \tag{3}$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are the generators of an SU (2)×SU (2) algebra. The lowering operators are

$$L_{1-} = \xi^* \frac{\partial}{\partial \zeta} - \zeta^* \frac{\partial}{\partial \xi},$$
$$L_{2-} = \xi^* \frac{\partial}{\partial \zeta^*} - \zeta \frac{\partial}{\partial \xi}. \tag{4}$$

The L_{i0} are diagonalized, and their eigenvalues m_i can be obtained from the expressions

$$J_0 = -i\frac{\partial}{\partial \alpha},\ K_0 = -i\frac{\partial}{\partial \beta}. \tag{5}$$

It is useful to define a parity operation P as follows:

$$(x,y,z,w) \to (-x, -y, -z, w),$$
$$\xi \to -\xi,\quad \zeta \to -\zeta^*, \tag{6}$$

which interchanges $\mathbf{L}_1$ and $\mathbf{L}_2$.

A four-dimensional notation will be used generally, that is, vectors will be four-vectors, except that the angular momentum operators defined in Eq. (3) will continue to be three-dimensional. Vectors which refer to physical quantities must be tangent to the hypersphere, and have only three independent components. Thus, the operator ∇ takes the form

$$\nabla_a = \partial_a - \hat{r}_a(\hat{r}_b\, \partial_b). \tag{7}$$

The operator ∇^2 is equal to the angular part of the ordinary four-dimensional Laplacian. The vector product is given by

$$(\mathbf{B}\times\mathbf{C})_a = \epsilon_{abcd} \hat{r}_d B_b C_c. \tag{8}$$

The curl of a vector $\mathbf{C}$ is obtained by substituting ∇ for $\mathbf{B}$ in (8), and can also be written as

$$(\text{curl}\ \mathbf{C})_a = -i\epsilon_{abcd} M_{db} C_c. \tag{9}$$

For calculations involving vectors, it is sometimes useful to introduce an auxiliary unit vector $\hat{\mathbf{r}}'$, which is contracted with the vector. Primed versions of the variables defined in Eqs. (1)–(3) will also be used; the tensor M'_{ij}, which acts on $\hat{r}'$, can be interpreted as a spin operator. The component substitution in Eq. (9) can be identified with M'. The permutation symbol converts this into the dual tensor $\tilde{M}'$, giving

$$\text{curl} = -M_{ab}\tilde{M}'_{ab} = 2\mathbf{L}_2\cdot\mathbf{L}'_2 - 2\mathbf{L}_1\cdot\mathbf{L}'_1. \tag{10}$$

1.2. Scalar and vector harmonic functions

Scalar harmonic functions correspond to the traceless, symmetric tensors of rank s formed from $\hat{r}$, and to the repre-

 0022-2488/84/040939-04$02.50

sentation $(s/2,s/2)$ of SU (2)×SU (2). The normalized function with the maximal values of m_i is $S_s^{s/2,s/2} = C_s(-\xi)^s$, where, using (2), $C_s^2 = (s+1)/2\pi^2$. The dimensionality of this representation is $(s+1)^2$, and the value of the Laplacian is $-s(s+2)$. By application of the lowering operators (4), the general scalar function is found to be

$$S_s^{s/2-a,s/2-b} = C_s[a!(s-a)!b!(s-b)!]^{1/2} \times \sum_k \frac{(\xi^*)^k \zeta^{b-k} (\zeta^*)^{a-k} (-\xi)^{s+k-a-b}}{k!(s+k-a-b)!(a-k)!(b-k)!}, \tag{11}$$

where the sum extends over all values giving nonnegative exponents. If either m or n is at the end of its range, there is only one term. These functions satisfy the following symmetries:

$$S_s^{\mu\nu *} = (-1)^{\mu+\nu} S_s^{-\mu,-\nu}, \qquad PS_s^{\mu\nu} = (-1)^s S_s^{\nu\mu}. \tag{12}$$

The scalar functions (11) have been obtained before in this coordinate system,[4,6] and are related to the representation matrices for SU (2).

The vector harmonic functions correspond to the tensors which are symmetric in v indices and have been antisymmetrized in one more, and which are linear in $\hat{r}'$ and of degree v in $\hat{r}$. There are $2v(v+2)$ such functions. The vector components are the coefficients of the components of $\hat{r}'$. The same notation will be used here to refer either to the function of $\hat{r}'$ or to the vector. As pointed out in Ref. 1, $\hat{r}\cdot\mathbf{V}$ and $\nabla\cdot\mathbf{V}$ have antisymmetry, but depend only on $\hat{r}$ and must therefore vanish. To obtain these functions, components of the $(v/2,v/2)$ representation (involving $\hat{r}$) are combined with components of the $(1/2,1/2)$ representation (involving $\hat{r}'$) by using the vector-addition coefficients for addition of $v/2$ and $1/2$ to give a total $L = (v \pm 1)/2$. This gives, for the components of the representation $((v+1)/2, (v-1)/2)$ which have the highest values of m_1,

$$V_{v+}^{(v+1)/2,(v-1)/2-b} = D_v \left[\frac{(v-1)!}{b!(v-1-b)!}\right]^{1/2} \times \zeta^b (-\xi)^{v-b-1} [(-\xi)\zeta' - \zeta(-\xi')], \tag{13}$$

where $D_v^2 = v/(4\pi^2)$. The functions in which m_2 is maximal are

$$\begin{aligned} V_{v+}^{(v+1)/2-a,(v-1)/2} &= \frac{D_v}{v(v+1)}\left[\frac{(v+1)!}{a!(v+1-a)!}\right]^{1/2} \\ &\times \{(v+1-a)(\zeta^*)^{a-1}(-\xi)^{v-a-1}[v\zeta'\zeta^*(-\xi) \\ &- (v-a)(-\xi')\zeta\zeta^* - a\zeta'^*\zeta(-\xi)] \\ &+ a(\zeta^*)^{a-2}(-\xi)^{v-a}[v\xi'^*\zeta^*(-\xi) \\ &- (v-a+1)(-\xi')\xi^*\zeta^* - (a-1)\zeta'^*\xi^*(-\xi)]\}. \end{aligned} \tag{14}$$

The general vector functions are more complicated than (11), but are not needed here. They satisfy the Hermiticity condition

$$\mathbf{V}_{v\pm}^{\mu\nu *} = (-1)^{\mu+\nu+1} \mathbf{V}_{v\pm}^{-\mu,-\nu}, \tag{15}$$

which can be used to determine the functions with minimal values of μ or ν from Eqs. (13), (14). The other components of this O(4) tensor representation, the functions $\mathbf{V}_{v-}$ in the representation $((v-1)/2, (v+1)/2)$, are obtained by application of the parity operation of Eq. (6):

$$P\mathbf{V}_{v+}^{\mu\nu} = (-1)^{v+1} \mathbf{V}_{v-}^{\nu\mu}. \tag{16}$$

The curl of the vector function is given by Eq. (10):

$$\operatorname{curl} \mathbf{V}_{v\pm}^{\mu\nu} = \mp (v+1) \mathbf{V}_{v\pm}^{\mu\nu}. \tag{17}$$

1.3. Free massive scalar and vector fields

The usual relativistic spectra are altered by use of the hyperspherical domain. To see what kinds of effects occur, consider some free fields of mass μ. Expansion of a scalar field in terms of the scalar functions S is straightforward. The following energies are obtained:

$$E = (\mu^2 + k^2(s))^{1/2}, \tag{18}$$

where $k^2(s) = s(s+2)$. The quantity k can be interpreted as the momentum. It takes on discrete values, and the direction is also quantized, corresponding to the $(s+1)^2$ independent functions S_s.

The massive spin-1 field has a number of complications, which help to illustrate some additional features of the hyperspherical formalism. The flat-space Hamiltonian is[7,8]

$$H = \frac{1}{2}\int dv \left[\mathbf{\Pi}^2 + \frac{(\nabla\cdot\mathbf{\Pi})^2}{\mu^2} + (\nabla\times\mathbf{A})^2 + \mu^2\mathbf{A}^2\right]. \tag{19}$$

The transverse and longitudinal fields must be separated,

$$\mathbf{\Pi} = \mathbf{\Pi}_t - (\nabla^2)^{-1}\nabla\Psi, \quad \mathbf{A} = \mathbf{A}_t + \nabla\Phi, \tag{20}$$

where the subscript t means the divergence vanishes. On the hypersphere, these fields are expanded in the orthonormal harmonic functions

$$\mathbf{A}_t = \Sigma q_{vn}\mathbf{V}_v^n, \quad \Phi = \Sigma Q_{sm} S_s^m, \qquad \mathbf{\Pi}_t = \Sigma p_{vn}\mathbf{V}_v^n, \quad \Psi = \Sigma P_{sm} S_s^m. \tag{21}$$

A condensed notation has been used, as in Ref. 1. Then, using the expression for the curl given by (17) as well as the value of the Laplacian,

$$H = \tfrac{1}{2}\Sigma_{\kappa n}\{p_{\kappa n}^2 + [(\kappa+1)^2 + \mu^2]q_{\kappa n}^2\} + \tfrac{1}{2}\Sigma_{\kappa m}\{P_{\kappa m}^2[k^{-2} + \mu^{-2}] + k^2\mu^2 Q_{\kappa m}^2\}, \tag{22}$$

where κ denotes the generic O(4) quantum number, and $k^2 = \kappa(\kappa+2)$. The energies of the transverse and longitudinal modes are slightly different

$$E = [\mu^2 + k^2 + h^2]^{1/2}, \tag{23}$$

where h is the helicity and has the values $0, \pm 1$. The value $\kappa = 0$ is allowed only for $h = 0$.

In these examples, the rotational invariance of the hyperspherical formalism has preserved the degeneracy associated with different momentum directions. There is, however, a helicity-dependent contribution to the effective mass arising from the curvature of the space.

2. EVALUATION OF INTEGRALS

A scalar integral containing a product of three harmonic functions can be written as a product of two 3-j symbols and a reduced matrix element. Similar integrals, involving the group O(3) and the surface of a three-dimensional sphere, are encountered in some atomic and nuclear calculations. In the O(3) problem, the reduced matrix element is calculated by exploiting a relation between the harmonic functions and the rotation matrix elements. The calculations described here use a different trick, which is based on the direct-product nature of the group and the fact that the functions have relatively simple expressions if either of the m_i values is at the end of its range. For one of the three functions, both m_i values are chosen to be maximal; for each of the other two, a different one of the two m_i values is chosen to be minimal. Except in one case, the more complicated expression (14) can also be avoided. The reduced matrix elements are then obtained by dividing by the 3-j symbols for these special cases. In fact, the combinations of quantum numbers which simplify the matrix elements are the same ones for which the summation formula for the 3-j symbols reduces to a single term[9]

$$\begin{pmatrix} \alpha & \beta & \gamma \\ -\alpha & \beta & \alpha-\beta \end{pmatrix} = (-1)^{\gamma-\alpha-\beta}\left[\frac{(2\alpha)!(2\beta)!}{(\alpha+\beta-\gamma)!(\alpha+\beta+\gamma+1)!}\right]^{1/2}. \quad (24)$$

It is therefore possible to express each of the reduced matrix elements by a relatively simple algebraic formula.

Four types of integrals are considered. The simplest integral, with three scalar functions $S_1S_2S_3$, does not occur in a pure gauge model but would be required in a model with scalar fields. Integrals involving the functions $S'\mathbf{V}\cdot\nabla S$ arise in the calculation of the longitudinal propagator and the Faddeev–Popov determinant, in the Coulomb gauge.[1] The integral of $\mathbf{V}'\cdot S\,\mathbf{V}$ arises in calculation of the color-charge density. The integral of $\mathbf{V}_1\cdot(\mathbf{V}_2\times\mathbf{V}_3)$ arises from the cubic self-coupling term in the magnetic energy.

In the magnetic energy, there is also a quartic self-coupling term, which involves the combination $(\mathbf{V}_1\times\mathbf{V}_2)\cdot(\mathbf{V}_3\times\mathbf{V}_4)$. It is shown in Sec. 2.5 that the integral of this quantity is given by an expansion, in which each term can be related to integrals of the third and fourth types described above.

2.1. SSS

The integral to be evaluated is

$$I_1 = \int d\Omega\, S_x^{ap} S_y^{bq} S_z^{cr} = R_1(x,y,z)\begin{pmatrix} \frac{x}{2} & \frac{y}{2} & \frac{z}{2} \\ a & b & c \end{pmatrix}\begin{pmatrix} \frac{x}{2} & \frac{y}{2} & \frac{z}{2} \\ p & q & r \end{pmatrix}, \quad (25)$$

where R_1 is the reduced matrix element. For a nonzero integral, the m_i values satisfy $a+b+c=p+q+r=0$, and $x/2, y/2, z/2$ satisfy the triangular condition. The integrand is a polynomial of degree 2σ, where $\sigma=(x+y+z)/2$, and the integral vanishes unless σ is an integer. The 3-j symbols also vanish unless σ is an integer. If the m_i are chosen to be $b=q=y/2$, $a=-x/2$, and $r=-z/2$, with c and p determined from the constraint, each of the factors in I_1 has just one term. The following integral is encountered:

$$\int d\Omega\, |\xi^2|^y |\zeta^2|^{\sigma-y} = 2\pi^2 y!(\sigma-y)!/(\sigma+1)!. \quad (26)$$

Combining coefficients from Eq. (11), and using Eq. (24) for the 3-j symbols, we obtain

$$R_1(x,y,z) = \frac{(-1)^\sigma}{\pi}\left[\frac{(x+1)(y+1)(z+1)}{2}\right]^{1/2}. \quad (27)$$

2.2. SVS

The next integral is

$$I_2 = \int d\Omega\, S_x^{ap}\mathbf{V}_{y+}^{bq}\cdot\nabla S_z^{cr} = R_2(x,y,z)\begin{pmatrix} \frac{x}{2} & \frac{y+1}{2} & \frac{z}{2} \\ a & b & c \end{pmatrix}\begin{pmatrix} \frac{x}{2} & \frac{y-1}{2} & \frac{z}{2} \\ p & q & r \end{pmatrix}. \quad (28)$$

The triangle constraint in this case gives $|x-z|<y<x+z$. The integrand in (28) is a polynomial of degree $2(\sigma-1)$, where $\sigma=(x+y+z+1)/2$ must again be an integer. The m_i values are chosen to be $b=(y+1)/2$, $q=(y-1)/2$, with a and r as before. For $\mathbf{V}\cdot\nabla$, we make the following replacements in Eq. (13): $\zeta'\to\partial/\partial\zeta^*$ and $\xi'\to\partial/\partial\xi^*$. There is again just one term in the final integral, giving the result

$$R_2(x,y,z) = [(-1)^\sigma/2\pi] \times\left[\frac{(x+1)(z+1)(\sigma-x)(\sigma-y)(\sigma-z)(\sigma+1)}{(y+1)}\right]^{1/2}. \quad (29)$$

Note that the triangle constraints are equivalent to the requirement that each of the factors $(\sigma-x)$, $(\sigma-y)$, and $(\sigma-z)$ in (29) be positive.

For the integral with V_{y-} instead of V_{y+}, we use the parity opertion of Eq. (6) in (28). It is found that the reduced matrix element is the same as given by (29).

2.3. VSV

There are two integrals of the third type, one in which the two V's have the same handedness, and one in which they have opposite handedness:

$$I_{3\pm} = \int d\Omega\, \mathbf{V}_{x\pm}^{ap}\cdot S_y^{bq}\mathbf{V}_{z+}^{cr} = R_{3\pm}(x,y,z)\begin{pmatrix} \frac{x\pm 1}{2} & \frac{y}{2} & \frac{z+1}{2} \\ a & b & c \end{pmatrix} \times\begin{pmatrix} \frac{x\mp 1}{2} & \frac{y}{2} & \frac{z-1}{2} \\ p & q & r \end{pmatrix}. \quad (30)$$

To evaluate the dot product in (30), let $\{A^*, A, B^*, B\}$ denote, respectively, the coefficients of $\{\xi', \xi'^*, \zeta', \zeta'^*\}$ in V. This gives

$$\mathbf{V}_1\cdot\mathbf{V}_2 = 2(A_1A_2^* + A_1^*A_2 + B_1B_2^* + B_1^*B_2). \quad (31)$$

The integrands are polynomials of degree 2σ, where $\sigma=(x+y+z)/2$ as in (25), and must again be an integer. For I_{3+}, the triangular condition is $|x-z|\leqslant y\leqslant x+z-2$, while for I_{3-}, it is $|x-z|+2\leqslant y\leqslant x+z$. The calculations, carried out as in the previous two cases, are simplified by the choices $a=(x\pm 1)/2$, $p=(x\mp 1)/2$, $c=-(z+1)/2$, $q=-y/2$. The results are

$$R_{3+}(x,y,z)=\frac{(-1)^{\sigma+1}}{\pi}\times\left[\frac{(y+1)(\sigma+2)(\sigma+1)(\sigma+1-y)(\sigma-y)}{2(x+1)(z+1)}\right]^{1/2},$$
$$R_{3-}(x,y,z)=\frac{(-1)^{\sigma}}{\pi}\times\left[\frac{(y+1)(\sigma+1-x)(\sigma-x)(\sigma+1-z)(\sigma-z)}{2(x+1)(z+1)}\right]^{1/2}. \quad (32)$$

2.4. VVV

There are again two integrals to be considered, one in which all three V's have the same handedness, and a second in which one V has a different handedness:

$$I_{4\pm}=\int d\Omega\,\epsilon_{\alpha\beta\gamma\delta}V^{ap}_{x+,\alpha}V^{bq}_{y\pm\beta}V^{cr}_{z+,\gamma}\hat{r}_\delta$$
$$=R_{4\pm}(x,y,z)\begin{pmatrix}\frac{x+1}{2} & \frac{y\pm1}{2} & \frac{z+1}{2}\\ a & b & c\end{pmatrix}\times\begin{pmatrix}\frac{x-1}{2} & \frac{y\mp1}{2} & \frac{z-1}{2}\\ p & q & r\end{pmatrix}. \quad (33)$$

Other handedness combinations can be obtained from (33) by using the parity operation and the permutation properties of the triple scalar product. The integrands are polynomials of degree 2σ, where $\sigma=(x+y+z+1)/2$ is an integer. The triangular constraints require $|x-z|<y$, and $y<x+z$ in I_{4+}, $y<x+z-2$ in I_{4-}. The integrals will be evaluated for the case $a=(x+1)/2$, $p=(x-1)/2$, $c=-(z+1)/2$, and $q=-(y\mp1)/2$.

The integrand in (33) is a determinant D, which can be rewritten as follows by using the notation of Eq. (31):

$$D=2\begin{vmatrix}\xi & A_1 & A_2 & A_3\\ \xi^* & A_1^* & A_2^* & A_3^*\\ \zeta & B_1 & B_2 & B_3\\ \zeta^* & B_1^* & B_2^* & B_3^*\end{vmatrix}. \quad (34)$$

The indices $\{1,2,3\}$ refer, respectively, to the three factors V_x, V_y, and V_z in (33). From Eq. (13) it is found that V_x and V_z can be expressed as $V_x=v_1(\xi\zeta'-\zeta\xi')$ and $V_z=v_3(\zeta^*\xi'^*-\xi^*\zeta'^*)$, so the determinant (34) can be simplified to

$$D=2v_1v_3(|\xi|^2+|\zeta|^2)(A_2\xi^*-A_2^*\xi+B_2\zeta^*-B_2^*\zeta). \quad (35)$$

Then, either (13) or (14) is used for V_y. The final results are

$$R_{4+}(x,y,z)=\frac{(-1)^{\sigma+1}}{\pi}\times\left[\frac{(\sigma+2)(\sigma+1)\sigma(\sigma-x)(\sigma-y)(\sigma-z)}{(x+1)(y+1)(z+1)}\right]^{1/2},$$
$$R_{4-}(x,y,z)=\frac{(-1)^{\sigma+1}}{\pi}\times\left[\frac{(\sigma+1)(\sigma+1-y)(\sigma-y)(\sigma-1-y)(\sigma-x)(\sigma-z)}{(x+1)(y+1)(z+1)}\right]^{1/2}. \quad (36)$$

The triangle constraints imply (along with other conditions) that in R_{3+}, $(\sigma-y)$ is positive, and that in R_{3-}, $(\sigma-x)$ and $(\sigma-z)$ are positive.

2.5. VVVV

The foregoing results can be applied to the integral

$$I_5=\int d\Omega\,(\mathbf{V}_\alpha\times\mathbf{V}_\beta)\cdot(\mathbf{V}_\gamma\times\mathbf{V}_\delta), \quad (37)$$

where α is a shorthand notation for the entire set of quantum numbers needed for identification of $\mathbf{V}_\alpha$. The cross product $\mathbf{V}_\alpha\times\mathbf{V}_\beta$ is a vector, although it is not transverse. It can be expanded as in Eqs. (20) and (21),

$$\mathbf{V}_\alpha\times\mathbf{V}_\beta=\sum_{vn}g^{\alpha\beta}_{rn}\mathbf{V}^n_v+\sum_{sm}f^{\alpha\beta}_{sm}\nabla S^m_s. \quad (38)$$

The expansion coefficients for the transverse term are given directly by Eq. (33),

$$g^{\alpha\beta}_{vn}=\int d\Omega\,\mathbf{V}^n_v\cdot(\mathbf{V}_\alpha\times\mathbf{V}_\beta). \quad (39)$$

The coefficients in the longitudinal term are obtained as follows:

$$-k(s)^2f^{\alpha\beta}_{sm}=\int d\Omega\,S^m_s\nabla\cdot(\mathbf{V}_\alpha\times\mathbf{V}_\beta)$$
$$=\int d\Omega\,S^m_s[\mathbf{V}_\beta\cdot(\nabla\times\mathbf{V}_\alpha)-\mathbf{V}_\alpha\cdot(\nabla\times\mathbf{V}_\beta)]$$
$$=(C^\alpha-C^\beta)h^{\alpha\beta}_{sm}, \quad (40)$$

where the C's are given by Eq. (10) and where

$$h^{\alpha\beta}_{sm}=\int d\Omega\,S^m_s\mathbf{V}_\alpha\cdot\mathbf{V}_\beta \quad (41)$$

is given by Eq. (30). Substitution of these expansions into (37) gives the result

$$\int d\Omega\,(\mathbf{V}_\alpha\times\mathbf{V}_\beta)\cdot(\mathbf{V}_\gamma\times\mathbf{V}_\delta)=\sum_{vn}g^{\alpha\beta}_{vn}g^{\gamma\delta}_{vn}+\sum_{sm}(C^\alpha-C^\beta)(C^\gamma-C^\delta)\frac{h^{\alpha\beta}_{sm}h^{\gamma\delta}_{sm}}{k(s)^2}. \quad (42)$$

Note that the expansions in (42) have a finite number of terms, and that each term contains a product of four 3-j symbols.

ACKNOWLEDGMENTS

I thank J. L. Balduz and L. C. Biedenharn for discussions.

This work was supported by the U.S. Department of Energy under contract DE-AC02-76ER03066.

[1] R. E. Cutkosky, Phys. Rev. Lett. **51**, 538 (1983).
[2] S. L. Adler, Phys. Rev. D **6**, 3445 (1972); Phys. Rev. D **8**, 2400 (1973).
[3] L. C. Biedenharn, J. Math. Phys. **2**, 433 (1161).
[4] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
[5] A. O. Barut and R. Wilson, J. Math. Phys. **17**, 900 (1976).
[6] M. A. B. Beg and H. Ruegg, J. Math. Phys. **6**, 677 (1965).
[7] G. Wentzel, *Quantum Theory of Fields* (Interscience, New York, 1949), p. 73.
[8] T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood Academic, New York, 1981), p. 55.
[9] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton University, Princeton, N.J., 1957), p. 48.

Solving the Schrödinger equation with use of 1/N perturbation theory

Leonard D. Mlodinow[a)]
California Institute of Technology, Pasadena, California 91125 and Institut des Hautes Etudes Scientifiques, 91440 Bures-sur-Yvette, France
Michael P. Shatz[b)]
California Institute of Technology, Pasadena, California 91125

(Received 12 April 1983; accepted for publication 24 June 1983)

The large N expansion provides a powerful new tool for solving the Schrödinger equation. In this paper, we present simple recursion formulas which facilitate the calculation. We do some numerical calculations which illustrate the speed and accuracy of the technique

PACS numbers: 03.65.Ge, 03.65.Sq, 02.60. + y

I. INTRODUCTION

The large N expansion provides an approximation to the low-lying spectrum and the corresponding wave functions of a quantum-mechanical system. The approximation is based on a solution of the classical equations of motion, about which a perturbation series is developed. In solid-state physics, such expansions are well known[1–4] as the "semiclassical theory of spin systems," where the expansion is formally an expansion in $1/s$, s being the spin. The generalization of those ideas to include coordinate space problems has come only recently,[4–10] having been inspired by interest in similar expansions being developed for quantum field theory.[5–7]

Large N expansions are not expansions in a coupling constant (hence they are often called "nonperturbative"); rather, each coefficient in the expansion is a function of the coupling constants. Thus even the leading orders of the expansion may exhibit qualitative features of the problem that are not readily evident in coupling-constant perturbation theory. Another advantage of the large N method is that it does not depend on the Hamiltonian's being the sum of two terms, one solvable and the other relatively small. Furthermore, the calculation can be done using a simple algebraic recursion method. It is best for low-lying states, hence complimentary to the WKB method.

Numerical work provides concrete evidence of the promise of large N expansions. For instance, in Ref. 11, the method was applied to the axially symmetric problem of a hydrogen atom in a uniform magnetic field; with the use of only a few terms, good results were obtained for the ground-state energy (for each subspace of fixed azimuthal angular momentum) even for strong fields where the perturbation series in e^2 diverges badly.[12–14] A recursion relation derived in Ref. 11 greatly facilitated the calculation. In this paper, we show how that recursion method can be extended to allow treatment of the excited states and the calculation of the wave function.

For the sake of simplicity, we will explicitly treat the problem of a single particle in a spherically symmetric potential. The generalization to more complicated problems is discussed at the end of Sec. III. In Sec. II, we introduce and discuss the technique. The calculation is reduced to quadratures. In Sec. III, we derive algebraic recursion relations (presented in detail in the Appendix) that are useful for rapidly computing the large N series for the spectrum and nodes. In Sec. IV, we examine the utility of the method in the context of a few simple problems. We obtain energies and nodes, typically correct to at least several significant figures. The calculations were performed on a VAX-11/780, essentially instantaneously. Alternatively, results accurate to within a few percent can quickly be obtained by hand.

II. LARGE N EXPANSION

We suppose that we wish to solve a given spherically symmetric Schrödinger equation with potential $V(r)$. Although we may have a specific dimensionality and angular momentum in mind, we consider the more general problem where those quantities are arbitrary.

After separating variables, we obtain the radial equation (we choose units such that $\hbar = m = 1$)

$$\left\{-\frac{1}{2}\left(\frac{d^2}{dr^2}+\frac{N-1}{r}\frac{d}{dr}\right)+\frac{l(l+N-2)}{2r^2}+\hat{V}(r)\right\}\phi(r) = E\phi(r), \tag{2.1}$$

where $\hat{V}(r)$ is an N-dimensional potential to be defined shortly. Equation (2.1) is brought to the form of a Schrödinger equation on the half-line by setting

$$\psi(r) = r^{(N-1)/2}\phi(r). \tag{2.2}$$

Then ψ satisfies

$$-\frac{1}{2}\frac{d^2\psi}{dr^2}+k^2\left[\frac{(1-1/k)(1-3/k)}{8r^2}+V(r)\right]\psi = E\psi, \tag{2.3}$$

where

$$k \equiv N+2l, \tag{2.4}$$

and we have chosen

$$\hat{V}(r) \equiv k^2 V(r). \tag{2.5}$$

This may differ from the true potential for the number of spatial dimensions (and angular momentum) of interest, but only by a constant factor which can be absorbed in the definition of the coupling constants and reinstated at the end of the calculation. We may now develop an expansion in inverse powers of k.

[a)] Bantrell Fellow. Research supported in part by the U.S. National Science Foundation under Grant 81-19979.
[b)] Work supported in part by the U.S. Department of Energy under Contract No. DE-AC03-81-ER40050.

Due to definition (2.5), in the large k limit, the eigenfunctions ψ in Eq. (2.3) peak about the minimum r_0 of the effective potential $\frac{1}{8}r^{-2} + V(r)$, and the entire eigenvalue spectrum concentrates around the value[13]

$$E^{(-2)} \equiv 1/8r_0^2 + V(r_0). \tag{2.6}$$

$k^2E^{(-2)}$ is the energy of a classical particle with angular momentum $L_{cl} = \frac{1}{2}k = l + \frac{1}{2}N$ executing circular motion in the potential $\hat{V}(r)$, and forms our "zeroth"-order approximation to the quantum eigenvalue spectrum. For finite k, the greater the radial quantum number (= the number of finite radial nodes), the greater the deviation of the wave function from r_0, and of the quantum energy from $k^2E^{(-2)}$. Hence our approximation is best for the low-lying states, in contrast to the WKB method.[15]

For the ground state, the perturbation series in powers of k^{-1} is computed as follows. Since the wave function is nodeless, we may define

$$\psi(r) \equiv e^{\mathscr{U}(x)}, \tag{2.7}$$

where

$$x \equiv r - r_0. \tag{2.8}$$

It is also convenient to define

$$V_{\rm eff}(x) \equiv \tfrac{1}{8}r(x)^{-2} + V(r(x)) - E^{(-2)}, \tag{2.9}$$

which differs from the above-mentioned effective potential only by a constant term, chosen so that the minimum of $V_{\rm eff}(x)$ is zero. Using Eq. (2.3), we now find

$$-\tfrac{1}{2}(\mathscr{U}''(x) + \mathscr{U}'(x)\mathscr{U}'(x)) + k^2V_{\rm eff}(x) + (-\tfrac{1}{2}k + \tfrac{3}{8})r(x)^{-2} = \mathscr{E},$$

$$\mathscr{E} = E - E^{(-2)}k^2, \quad \mathscr{U}'(x) \equiv \frac{d\mathscr{U}(x)}{dx}. \tag{2.10}$$

The recursive solution is obtained by expanding the energy and wave function,

$$\mathscr{E} = \sum_{n=-1}^{\infty} E^{(n)}k^{-n}, \quad \mathscr{U}'(x) \equiv \sum_{n=-1}^{\infty} u^{(n)}(x)k^{-n}, \tag{2.11}$$

and collecting terms of like order in k. Thus we obtain

$$u^{(-1)}(x) = -\sqrt{2V_{\rm eff}(x)}, \tag{2.12a}$$

$$\sqrt{2V_{\rm eff}(x)}\,u^{(0)}(x) = E^{(-1)} + \tfrac{1}{2}r(x)^{-2} + \tfrac{1}{2}u^{(-1)\prime}(x), \tag{2.12b}$$

$$\sqrt{2V_{\rm eff}(x)}\,u^{(1)}(x) = E^{(0)} - \tfrac{3}{8}r(x)^{-2} + \tfrac{1}{2}(u^{(0)\prime}(x) + u^{(0)}(x)u^{(0)}(x)), \tag{2.12c}$$

and

$$\sqrt{2V_{\rm eff}(x)}\,u^{(n+1)}(x) = E^{(n)} + \frac{1}{2}\left(u^{(n)\prime}(x) + \sum_{m=0}^{n} u^{(m)}(x)u^{(n-m)}(x)\right); \quad n>0. \tag{2.12d}$$

We must, at this point, discuss the branch of the square root to be taken in Eqs. (2.12). Since we are interested in the normalizable wave function, the positive branch must be taken (i.e., the square root is defined to be a positive number) in the region $x \geqslant 0$ (corresponding to $r \geqslant r_0$). At $x = 0$ ($r = r_0$), the square root vanishes. In the region $-r_0 \leqslant x \leqslant 0$ ($0 \leqslant r \leqslant r_0$), we take the negative branch (i.e., the square root is defined to be a negative number), since in the large k limit ψ peaks about $x = 0$ ($r = r_0$).

Equation (2.12a) gives the leading approximation to the wave function. Because $V_{\rm eff}(0) = 0$ and $u^{(0)}(0) < \infty$, we find from (2.12b) that

$$E^{(-1)} = -\tfrac{1}{2}r_0^{-2} - \tfrac{1}{2}u^{(-1)\prime}(0). \tag{2.13a}$$

This may be substituted into Eq. (2.12b), which then defines the first-order wave function correction $u^{(0)}(x)$. Similarly, the higher-order energy corrections are given by

$$E^{(0)} = \tfrac{3}{8}r_0^{-2} - \tfrac{1}{2}(u^{(0)\prime}(0) + u^{(0)}(0)u^{(0)}(0)), \tag{2.13b}$$

$$E^{(n)} = -\frac{1}{2}\left(u^{(n)\prime}(0) + \sum_{m=0}^{n} u^{(m)}(0)u^{(n-m)}(0)\right); \quad n>0. \tag{2.13c}$$

$E^{(n)}$ is calculated after $u^{(n)}(x)$, but before $u^{(n+1)}(x)$. The (unnormalized) wave function is then given by [recall that $r = r(x)$]

$$\psi(r) = \exp\left[\int^x \sum_{n=-1}^{\infty} u^{(n)}(y)k^{-n}\,dy\right]. \tag{2.14}$$

If one is not interested in the long distance behavior of the wave function, the method described in the next section is considerably more efficient than that here. It is derived by expanding the functions $u^{(n)}(x)$ themselves in a Taylor series.

Before we discuss the excited states, it is useful to treat the simple example of the Coulomb problem $V = -e^2/r$. According to our prescription, what we shall really do is to approximate the solutions to the Schrödinger equation (2.1), where

$$\hat{V}(r) \equiv -k^2\hat{e}^2/r. \tag{2.15}$$

The solution to the problem of interest is obtained through the identification $e^2 = k^2\hat{e}^2$. Furthermore, instead of the coordinate r, we introduce the dimensionless coordinate

$$\rho = 4\hat{e}^2 r \tag{2.16}$$

in Eq. (2.3). The number "four" in Eq. (2.16) was chosen because in these units, $r_0 = 1$. The energy is now measured in units of $16\hat{e}^4$. Thus we obtain

$$E^{(-2)} = -\tfrac{1}{8} \tag{2.17}$$

and

$$V_{\rm eff}(x) = x^2/8(1+x)^2, \quad x \in [-1, \infty]. \tag{2.18}$$

From Eq. (2.12a) we find that

$$u^{(-1)}(x) = -x/2(1+x), \tag{2.19}$$

and from (2.13a),

$$E^{(-1)} = -\tfrac{1}{4}. \tag{2.20}$$

Next, Eq. (2.12b) yields

$$u^{(0)}(x) = -\frac{2+x}{2(1+x)}, \tag{2.21}$$

and so on. We have computed

$$E_{\rm ground} = -\tfrac{1}{8}k^2\{1 + 2k^{-1} + O(k^{-2})\} = -2\hat{e}^4k^2\{1 + 2k^{-1} + O(k^{-2})\}, \tag{2.22}$$

which upon identification of e^2 with $\hat{e}^2k^2$, is identified as the beginning of the expansion of the exact ground-state energy

$$E_{\text{ground}} = -2e^4/(k-1)^2. \tag{2.23}$$

For the wave function, we have found (C and C' are irrelevant constants)

$$\begin{aligned}\psi(r) &= C\exp\{[\tfrac{1}{2}\ln(1+x) - \tfrac{1}{2}(1+x)]k \\ &\quad + [-\tfrac{1}{2}\ln(1+x) - \tfrac{1}{2}(1+x)] + O(k^{-1})\} \\ &= C' r^{(k-1)/2}\exp\{-2k\hat{e}^2[r + rk^{-1} + O(k^{-2})]\},\end{aligned} \tag{2.24}$$

the expansion of the exact (unnormalized) ground-state wave function

$$\psi(r) = r^{(k-1)/2}\exp\{-2e^2/(k-1)\}\, r. \tag{2.25}$$

We now consider the excited states. In order to account for the n nodes of the nth excited state, we alter the ansatz of Eq. (2.7):

$$\psi_n(r) \equiv \left[\prod_{i=1}^{n}(r - r_i)\right] e^{\mathscr{U}_n(x)}. \tag{2.26}$$

Because as k increases the wave function increasingly concentrates about the point r_0, we must have

$$r_i = r_0 + a_i^{(1)}k^{-1} + a_i^{(2)}k^{-2} + \cdots, \tag{2.27}$$

so that Eq. (2.26) becomes

$$\begin{aligned}\psi_n(r) &= \left[\prod_i (x - A_i)\right] e^{\mathscr{U}_n(x)}, \\ A_i &= a_i^{(1)}k^{-1} + a_i^{(2)}k^{-2} + \cdots.\end{aligned} \tag{2.28}$$

By substituting Eq. (2.28) in Eq. (2.3), we obtain a revised version of Eq. (2.10) for the excited states, which we again solve order by order in inverse powers of k. [The expansions of $\mathscr{E}_n$ and $\mathscr{U}_n(x)$ take the same form as before.]

To illustrate the procedure, we consider the first excited state

$$\psi_1(r) = (x - A_1)e^{\mathscr{U}_1(x)}. \tag{2.29}$$

The function $\mathscr{U}_1(x)$ satisfies

$$\begin{aligned}&-\tfrac{1}{2}[\mathscr{U}_1''(x) + \mathscr{U}_1'(x)\mathscr{U}_1'(x)](x-a) - \mathscr{U}_1'(x) \\ &+ [k^2 V_{\text{eff}}(x) + (-\tfrac{1}{2}k + \tfrac{3}{8})r(x)^{-2}](x-a) = (x-a)\mathscr{E}_1.\end{aligned} \tag{2.30}$$

Solving this to order k^2, we again find (the square root is defined as before) the result of Eq. (2.12a):

$$u_1^{(1)}(x) = -\sqrt{2V_{\text{eff}}(x)}. \tag{2.31}$$

This is the leading behavior of the exponent for all states. Examining the order k terms, we find

$$\begin{aligned}&x\sqrt{2V_{\text{eff}}(x)}\,u_1^{(0)}(x) \\ &= [E_1^{(-1)} + \tfrac{1}{2}r(x) + \tfrac{1}{2}u_1^{(-1)\prime}(x)]x + u^{(-1)}(x).\end{aligned} \tag{2.32}$$

The energy coefficient $E^{(-1)}$ is determined by differentiating (2.32) and evaluating the result at $x = 0$:

$$E_1^{(-1)} = -\tfrac{1}{2}r_0^{-2} - \tfrac{3}{2}u^{(-1)\prime}(0). \tag{2.33}$$

The result for the nth excited state is

$$E_n^{(-1)} = -\tfrac{1}{2}r_0^{-2} - (n + \tfrac{1}{2})u^{(-1)\prime}(0), \tag{2.34}$$

reflecting the fact that, to this order, the spectrum is harmonic[4,8,14] with frequency proportional to $(V_{\text{eff}}''(0))^{1/2}$ [recall that $V_{\text{eff}}(0) = V_{\text{eff}}'(0) = 0$]. The next step is to examine the order k^0 terms in Eq. (2.30). The node coefficient $a_1^{(1)}$ is found by evaluating the equation at $x = 0$, whereas the energy coefficient $E_1^{(0)}$ is again obtained by differentiating the equation once and then evaluating it at $x = 0$. $u^{(1)}(x)$ is then given by an algebraic equation similar to Eq. (2.32), and one may then proceed to order k^{-1}, and so on.

III. ALGEBRAIC RECURSION METHOD

If one is not interested in the long distance behavior of the wave function, then it is useful to expand the functions $u^{(n)}(x)$ in a Taylor series in x. This results in purely algebraic recursion relations for the quantities of interest. Toward that end, it is convenient to rescale coordinates in the radial equation (2.1):

$$r \equiv \sqrt{k}\,\rho. \tag{3.1}$$

This necessitates a change in the definition of $\hat{V}$:

$$\hat{V}(r) \equiv kV(r/\sqrt{k}). \tag{3.2}$$

As before, this amounts to a mere rescaling of coupling constants which can easily be undone at the end of the calculation. For instance, if we are interested in the potential

$$V(r) = -(\alpha/r)e^{-\beta r} + \lambda r, \tag{3.3}$$

then we consider

$$\hat{V}(r) = (-k^{3/2}\hat{\alpha}/r)e^{-k^{1/2}\hat{\beta}r} + k^{1/2}\hat{\lambda}r \tag{3.4}$$

and identify $k^{3/2}\hat{\alpha} = \alpha$, $k^{-1/2}\hat{\beta} = \beta$, and $k^{1/2}\hat{\lambda} = \lambda$ when we are done. Equation (2.3) must now be replaced by

$$-\frac{1}{2}\frac{d^2\psi}{d\rho^2} + k^2\left[\frac{(1 - 1/k)(1 - 3/k)}{8\rho^2} + V(\rho)\right]\psi = Ek\psi. \tag{3.5}$$

We define

$$x = \sqrt{k}(\rho - \rho_0), \tag{3.6}$$

ρ_0 being the point where the bracketed term in Eq. (3.5) is minimized (in the large k limit). Without loss of generality, we shall set $\rho_0 = 1$. The ensuing calculation parallels that of Sec. II. Note, however, that because of the rescalings (3.1) and (3.2), the intermediate equations will exhibit a different k dependence.

For the ground state, we define [$\rho = \rho(x)$]:

$$\psi(\rho) = e^{\mathscr{U}(x)}, \tag{3.7}$$

which leads to

$$\begin{aligned}&-\tfrac{1}{2}(\mathscr{U}''(x) + \mathscr{U}'(x)\mathscr{U}'(x)) + kV_{\text{eff}}(x) \\ &\quad + (-\tfrac{1}{2} + \tfrac{3}{8}k^{-1})\rho(x)^{-2} = \mathscr{E}, \\ &\mathscr{E} \equiv E - E^{(-2)}k,\end{aligned} \tag{3.8}$$

where $E^{(-2)}$ is given by (2.6) with r_0 replaced by ρ_0, and V_{eff} by (2.9), with r replaced by ρ. We now expand $\mathscr{E}$ as in Eq. (A4), and also expand $\mathscr{U}'$ in the double power series:

$$\begin{aligned}\mathscr{U}'(x) &= D_1^0 x + [C_0^0 + C_1^0 x^2]y + [D_1^1 x + D_2^1 x^3]y^2 \\ &\quad + [C_0^1 + C_1^1 x^2 + C_2^1 x^4]y^3 + \cdots,\quad y \equiv k^{-1/2}.\end{aligned} \tag{3.9}$$

The expansion of the potential has the form

$$W(x) \equiv kV_{\text{eff}}(x) + [-\tfrac{1}{2} + \tfrac{3}{8}k^{-1}]\rho(x)^{-2}$$
$$= [W_0^0 + W_1^0 x^2] + [W_0^1 x + W_1^1 x^3]y$$
$$+ \sum_{n=2}^{\infty} [W_0^n x^{n-2} + W_1^n x^n + W_2^n x^{n+2}]y^n. \quad (3.10)$$

It is because of the rescaling of coordinates that $\mathscr{U}$ and W are polynomials in x, to each order in y. Equation (3.8) can be solved order by order in x and y, beginning with the lowest order in y, but with the highest power of x within a given order of y. The energy appears only in the even orders in y. For instance, to order y^0 we find

$$D_1^0 = -(2W_1^0)^{1/2}, \quad E^{(-1)} = -\tfrac{1}{2}D_1^0 + W_0^0, \quad (3.11)$$

whereas to order y, we obtain

$$C_1^0 = W_1^1/D_1^0, \quad C_0^0 = -(1/D_1^0)[C_1^0 - W_0^1], \quad (3.12)$$

and so on. In order to treat the excited states, we include factors describing the nodes, as before. For example, we write the wave function for the first excited state as $[\rho = \rho(x)]$:

$$\psi_1(\rho) = (xy - \alpha)e^{\mathscr{U}_1(x)},$$
$$\alpha = \alpha(y) = \alpha_1 k^{-1} + \alpha_2 k^{-2} + \cdots. \quad (3.13)$$

Finally, although we have considered problems with spherical symmetry, more complicated problems can also be considered. For the ground state, application of the method to problems lacking spherical symmetry is straightforward. The function $\mathscr{U}(x)$ in Eq. (3.7) is merely replaced by a new function of as many coordinates as are required.

For example, for a problem with axial symmetry, $\mathscr{U}(x)$ becomes $\mathscr{U}(x,z)$ and Eq. (3.8) becomes (subscripts of x and z are used to denote the derivative)

$$-\tfrac{1}{2}(\mathscr{U}_{xx} + \mathscr{U}_x\mathscr{U}_x + \mathscr{U}_{zz} + \mathscr{U}_z\mathscr{U}_z) + W = \epsilon, \quad (3.14)$$

where W now has the form

$$W(x,z) = [W_{00}^0 + W_{20}^0 x^2 + W_{02}^0 z^2]$$
$$+ [W_{30}^1 x^3 + W_{21}^1 x^2 z + W_{12}^1 xz^2$$
$$+ W_{03}^1 z^2]y + \cdots. \quad (3.15)$$

[It may be necessary to rotate coordinates to eliminate a possible $W_{11}^0 xz$ term in Eq. (3.15). If so, that should be done. Such a rotation affects the numerical value of the coefficients W_{ij}^n, but leaves the form of Eq. (3.15) invariant, except that it eliminates the undesired term.] A calculation of the ground-state energies for an axially symmetric problem using this method can be found in Ref. 11.

The excited states require a little further explanation. Let us consider the first excited state, and for concreteness again consider a potential with axial symmetry. To leading order in y, the Schrödinger equation (3.5) becomes

$$-\frac{1}{2}\frac{d^2\psi}{dx^2} - \frac{1}{2}\frac{d^2\psi}{dz^2}$$
$$+ k(W_{20}^0 x^2 + W_{02}^0 z^2 + W_{00}^0) + O(k^{-1/2}) = \epsilon k. \quad (3.16)$$

This equation is separable; in particular, it is the equation for uncoupled harmonic oscillators: just two copies of what we obtain in the spherically symmetric case. To leading order, we determine the position of the node as before, considering separately the case of excitation in the x and z directions. In higher orders there is mixing between the directions. That is, the expansion for the node, which to leading order corresponds to an excitation in the x direction, also depends on z. To be precise, Eq. (3.13) is replaced by

$$\psi_1^{(x)}(x,z) = \left\{xy - \sum_{n=1}^{\infty}\sum_{m=0}^{n} y^{2n}(f_{nm}z^{2m} + yg_{nm}z^{2m+1})\right\}e^{\mathscr{U}(x,z)}, \quad (3.17)$$

and an analogous expression for the other state, with x and z changing roles. In Eq. (3.17), y is given by[11]

$$y = (N - 1 + 2|m|)^{-1/2}. \quad (3.18)$$

IV. DISCUSSION

We used the formulas given in the Appendix to calculate the energies and nodes of the three lowest-lying states (in each l sector) for several potentials. The results are displayed in Tables I–III. In Table I, we exhibit the energies of the low-lying states for various spherically symmetric potentials. We compare our results with the numerical calculations of Refs. 16 and 17. Tables II and III display the first ten partial sums and the first ten Shanks extrapolated partial sums, for the energies and nodes of the three lowest-lying states of the exactly solvable problem of the linear potential with $N = 3$, $l = 0$.

In computing Table I, we calculated the first 14 partial sums of the series for the energy. For the ground state and first excited state, we then determined the two successive partial sums with the smallest difference and kept as many significant figures as coincided. The results for the second excited state ($n = 2$) were obtained in the same way, except that we used the Shanks extrapolated partial sums.[11,18]

We obtain values for the nodes to about four decimal places accuracy. That is an indication that the behavior of the wave function near the classical orbit is good. The behavior of the wave function near the origin and at large distances is not apparent from its Taylor expansion, calculated in Sec. III; therefore it should be investigated using the methods of Sec. II.

The difference between the methods of Secs. II and III is that in Sec. III the coordinate is rescaled by a power of the coupling constant ($k^{-1/2}$). This has the effect of reordering the $1/k$ expansion of the wave function, but does not effect the expansion of the energy, which is coordinate-independent. This rescaling is done because it makes the calculation of the energy much easier. It is apparent from Eqs. (2.13) that only certain derivatives of the wave function at the origin ($x = 0$) contribute to a given coefficient in the expansion for the energy. When the method of Sec. III is employed, the only information about the wave function that is calculated at a given order is that relevant to the calculation of the energy to that order. Calculation of the rest is postponed. Although this method therefore makes calculation of the

TABLE I. Energies of the low-lying states for various spherically symmetric potentials ($N = 3$). The numbers in parentheses were calculated in Refs. 16 and 17. $C_1 = 2^{1.7}$, $C_2 = 2^{0.8}$, and $\tau_0 = 2^{-1/2}$.

l	n	$V(r)$: $\ln(\tau/\tau_0)$	$C_1 \tau^{-0.2}$	$C_2 \tau^{-0.8}$
$l=0$	0	1.044 33	− 2.680 3	− 1.2187
		(1.044 3)	(− 2.686)	(− 1.218)
	1	1.847 44	− 2.253	− 0.462
		(1.847 4)	(− 2.253)	(− 0.462)
	2	2.289	− 2.044	− 0.26
		(2.289 7)	(− 2.044)	(− 0.265)
$l=1$	0	1.641 141 33	− 2.344 946 1	− 0.500 439
		(1.643)	(− 2.345)	(− 0.500)
	1	2.150 947	− 2.100 738	− 0.280 64
		(2.151)	(− 2.101)	(− 0.281)
	2	2.490 942	− 1.950 72	− 0.187 3
		(2.491)	(− 1.951)	(− 0.187)
$l=2$	0	2.013 308 64	− 2.156 260 884	− 0.294 695 9
		(2.015)	(− 2.156)	(− 0.295)
	1	2.387 432 8	− 1.990 055 5	− 0.194 911
		(2.388)	(− 1.990)	(− 0.195)
	2	2.662 492 0	− 1.875 032	− 0.141 979
		(2.663)	(− 1.875)	(− 0.142)
$l=3$	0	2.284 141 353 37	− 2.029 064 889 5	− 0.201 913 656 6
		(2.286)	(− 2.029)	(− 0.202)
	1	2.579 783 312 93	− 1.904 866 7	− 0.146 342 0
		(2.581)	(− 1.905)	(− 0.146)
	2	2.810 445 386	− 1.812 502 0	− 0.112 787
		(2.811)	(− 1.812)	(− 0.113)

TABLE II. Spectrum of the linear potential $V = \lambda r$. We treat the three lowest states in the case $N = 3$, $l = 0$, $\lambda = 2^{7/2}$. For each state the left column displays the first ten partial sums, and the one to its right, the first ten Shanks extrapolated partial sums. For the second excited state, we also display a second Shanks extrapolation. The exact energies, E_n, calculated from Ref. 18, are also given.

$n=0$, $E_0 = 9.352\,430$		$n=1$, $E_1 = 16.351\,80$		$n=2$, $E_2 = 22.082\,24$		
9.905 78	9.346 29	9.905 78	16.2888	9.905 78	21.4760	22.0712
9.315 95	9.355 04	16.941 4	16.3558	24.566 9	22.1774	22.0843
9.347 93	9.353 01	16.222 1	16.3514	20.650 4	22.0523	22.0817
9.353 75	9.349 64	16.386 4	16.3519	23.153 1	22.0953	22.0824
9.352 91	9.352 29	16.341 9	16.3517	21.187 9	22.0754	22.0822
9.352 24	9.352 22	16.354 7	16.3518	22.873 7	22.0863	22.0823
9.352 29	9.352 50	16.350 8	16.3518	21.357 2	22.0796	22.0822
9.352 48	9.352 49	16.352 1	16.3518	22.761 5	22.0841	22.0822
9.352 50	9.352 34	16.351 7	16.3518	21.435 9	22.0809	22.0822
9.352 40	9.352 39	16.351 8	16.3518	22.704 1	22.0832	22.0822

TABLE III. Nodes of the wave functions of the first two excited states of the linear potential ($N = 3$, $l = 0$, $\lambda = 2^{7/2}$). α, β, and γ are defined in Eq. (A2). The exact answer was calculated from Ref. 18. The columns on the left display the first ten partial sums; those on the right, the Shanks extrapolated partial sums. The second line of the table displays the exact answers.

α 0.020 183 5		β − 0.013 933 6		γ − 0.268 043	
0.023 171	0.020 029	− 0.021 857	− 0.014 362	− 0.305 77	− 0.259 80
0.019 537	0.020 192	− 0.010 452	− 0.014 199	− 0.251 67	− 0.269 51
0.020 336	0.020 186	− 0.016 033	− 0.013 841	− 0.278 28	− 0.267 61
0.020 151	0.020 183	− 0.012 423	− 0.013 972	− 0.260 46	− 0.268 20
0.020 189	0.020 182	− 0.015 135	− 0.013 916	− 0.274 15	− 0.267 97
0.020 180	0.020 184	− 0.012 920	− 0.013 943	− 0.262 89	− 0.268 08
0.020 186	0.020 184	− 0.014 820	− 0.013 928	− 0.272 52	− 0.268 02
0.020 183	0.020 181	− 0.013 141	− 0.013 937	− 0.264 07	− 0.268 06
0.020 182	0.020 183	− 0.014 654	− 0.013 931	− 0.271 61	− 0.268 03
0.020 183	0.020 180	− 0.013 271	− 0.013 935	− 0.264 79	− 0.268 05

energy easier, it has a detrimental effect on the approximation of the wave function, especially far away from the classical minimum ($x = 0$).

We also obtained the eigenvalues of the three lowest-lying states to about four decimal places accuracy when $l = 0$. For $l = 3$, we obtained 6–12 place accuracy. The series, although asymptotic,[19] is, in general, divergent. The sequence of partial sums does not stabilize so well for the more highly excited states. For these states the Shanks transformed sequence is more stable, which allows one to obtain more accurate results. Perhaps other summation techniques, which have been successfully applied to divergent coupling-constant perturbation series,[20] are also useful here.

The computations were done on a VAX 11/780, using double precision arithmetic. For a given state, the entire calculation (energy, nodes, wave function expansion) took less CPU time than the system clock could measure. We warn the reader, however, that if one is not careful, significant roundoff error can occur as early as the eighth term. Roundoff error became significant in our calculation at about the 15th term. Presumably, it could be further forestalled by working on a larger machine. Another way to avoid roundoff error is to employ a symbolic manipulation program, such as MACSYMA. Comparison of our results with those obtained using MACSYMA confirmed that Tables I–III are not affected by roundoff error. We also used MACSYMA to compute Tables IV–V, which exhibit analytic expressions for the first several coefficients of the $1/N$ expansion for the energy and nodes of the low-lying states of power potentials. The coefficients are given as functions of the power.

TABLE IVa. Analytic expressions for the coefficients in the large N expansion of the ground state of the potential $V = e^2 r^{2\nu}$, $e^2 = k^{1-\nu}/(8\nu)$. f is the positive root of $4f^2 - 2 = 2\nu$. The expansion is defined in Eq. (A4).

n	$E_0^{(n)}$
-2	$\frac{1}{8} + \frac{1}{4(4f^2-2)}$
-1	$(f-1)/2$
0	$(-4f^6 + 23f^4 - 36f^3 - f^2 + 36f - 18)/(36f^2)$
1	$(16f^{11} - 40f^9 + 81f^7 - 48f^6 - 142f^5 + 276f^4 - 131f^3 - 156f^2 + 216f - 72)/(216f^4)$
2	$(-20\,288f^{16} + 25\,392f^{14} + 43\,200f^{13} - 21\,036f^{12} - 64\,800f^{11} + 18\,121f^{10} + 110\,700f^9 - 96\,609f^8 - 164\,700f^7 + 405\,687f^6 - 348\,300f^5 + 45\,133f^4 + 229\,500f^3 - 307\,800f^2 + 194\,400f - 48\,600)/(291\,600f^8)$

ACKNOWLEDGMENTS

LDM is grateful to Dr. Barbara Czermak and Dr. Dieter Maison for many stimulating discussions. We are also indebted to the Laboratory for Computer Science at the

TABLE IVb. Analytic expressions for the coefficients in the large N expansion for the energy of the first excited state for power potentials. For the notation, see Table IVa.

n	$E_1^{(n)}$
-2	$\frac{1}{8} + \frac{1}{4(4f^2-2)}$
-1	$(3f-1)/2$
0	$(-52f^6 + 227f^4 - 108f^3 - 157f^2 + 108f - 18)/(36f^2)$
1	$(176f^{11} - 632f^9 + 1227f^7 - 208f^6 - 1586f^5 + 908f^4 + 599f^3 - 676f^2 + 216f - 24)/(72f^4)$
2	$(-897\,728f^{16} + 2\,572\,752f^{14} + 1\,425\,600f^{13} - 5\,267\,316f^{12} - 3\,693\,600f^{11} + 8\,713\,351f^{10} + 4\,819\,500f^9 - 12\,116\,379f^8 - 2\,907\,900f^7 + 15\,177\,297f^6 - 6\,828\,300f^5 - 5\,298\,377f^4 + 6\,601\,500f^3 - 2\,835\,000f^2 + 583\,200f - 48\,600)/(291\,600f^6)$

TABLE IVc. Analytic expressions for the coefficients in the large N expansion for the energy of the second excited state for power potentials. For the notation see Table IVa.

n	$E_2^{(n)}$
-2	$\frac{1}{8} + \frac{1}{4(4f^2-2)}$
-1	$(5f-1)/2$
0	$(-148f^6 + 635f^4 - 180f^3 - 469f^2 + 180f - 18)/(36f^2)$
1	$(2480f^{11} - 9080f^9 + 17\,595f^7 - 1776f^6 - 22\,370f^5 + 7620f^4 + 10\,295f^3 - 5772f^2 + 1080f - 72)/(216f^4)$
2	$(-6\,362\,048f^{16} + 19\,711\,632f^{14} + 6\,696\,000f^{13} - 42\,669\,156f^{12} - 17\,820\,000f^{11} + 71\,965\,291f^{10} + 22\,990\,500f^9 - 94\,852\,839f^8 - 12\,892\,500f^7 + 105\,866\,877f^6 - 30\,658\,500f^5 - 45\,721\,757f^4 + 30\,712\,500f^3 - 7\,889\,400f^2 + 972\,000f - 48\,600)/(291\,600f^6)$

TABLE Va. Analytic expressions for the coefficients in the large N expansion of the node for the first excited state of power potentials. α is defined in Eq. (A2). For the notation see Table IVa.

n	α_n
1	$(-4f^3+7f-3)/(3f^2)$
2	$(16f^8+880f^6-720f^5-1581f^4+1620f^3$ $+272f^2-630f+135)/(270f^4)$
3	$(21\,824f^{13}-87\,760f^{11}+8064f^{10}-148\,180f^9$ $+445\,536f^8+218\,125f^7-908\,712f^6+258\,796f^5$ $+427\,014f^4-234\,455f^3-34\,272f^2+39\,690f-5670)$ $/(34\,020f^6)$

TABLE Vb. Analytic expressions for the coefficients in the large N expansion of the nodes for the second excited state of power potentials. β and γ are defined in Eq. (A2). For the notation see Table IVa.

n	β_n	γ_n
1	$-1/(2f)$	$(28f^3-49f+12)/(6f^2)$
2	$(592f^6-1976f^4+528f^3$ $+1789f^2-1032f+144)/(144f^4)$	$(-128f^8-14\,544f^6+6720f^5$ $+26\,328f^4-15\,120f^3-7831f^2$ $+5880f-720)/(720f^4)$
3	$(-896f^{11}-858\,880f^9+421\,632f^8$ $+3\,057\,720f^7-2\,159\,424f^6$ $-2\,967\,460f^5+2\,947\,248f^4$ $+211\,021f^3-925\,956f^2$ $+291\,600f-25\,920)$ $/(25\,920f^6)$	$(-4\,647\,424f^{13}+16\,066\,688f^{11}$ $-387\,072f^{10}+26\,367\,968f^9$ $-43\,884\,288f^8-64\,789\,016f^7$ $+90\,611\,131f^6+13\,777\,996f^5$ $-46\,487\,952f^4+9\,025\,153f^3$ $+5\,920\,236f^2-2\,222\,640f$ $+181\,440)/(544\,320f^6)$

Massachusetts Institute of Technology for allowing us the use of MACSYMA to perform algebraic manipulations.

APPENDIX

We present the recursion relations discussed in Sec. III of the paper. We wish to solve the differential equation

$$(-\tfrac{1}{2}\partial_x^2 + W - \mathscr{E})\psi = 0. \tag{A1}$$

For the first three states we try the ansätze

$$\begin{aligned}
\psi_1 &= e^{\mathscr{U}(x)},\\
\psi_2 &= (xy-\alpha(y))e^{\mathscr{U}(x)},\\
\psi_3 &= (x^2y^2 + xy\,\gamma(y) + \beta(y))e^{\mathscr{U}(x)}.
\end{aligned} \tag{A2}$$

(Although the $\mathscr{U}$'s are not the same for the three states, we do not write an index to distinguish them; no confusion should result.) and

This leads to the following differential equations for the ground state, and first and second excited states, respectively:

$$(\mathscr{U}'^2 + \mathscr{U}'' - 2W + 2\mathscr{E}) = 0, \tag{A3}$$

$$(xy-\alpha)(\mathscr{U}'^2 + \mathscr{U}'' - 2W + 2\mathscr{E}) + 2y\mathscr{U}' = 0,$$

$$(x^2y^2 + xy\,\gamma(y) + \beta(y))(\mathscr{U}'^2 + \mathscr{U}'' - 2W + 2\mathscr{E}) + 2y^2 + 2(2x^2y + y\,\gamma(y))\mathscr{U}' = 0.$$

We expand W, $\mathscr{E}$, γ, β, α, and $\mathscr{U}$ in power series as follows:

$$\begin{aligned}
\mathscr{E} &= \sum_{n=0}^{\infty} E^{(n-1)}k^{-n},\\
\alpha &= \sum_{n=1}^{\infty} \alpha_n k^{-n},\\
\beta &= \sum_{n=1}^{\infty} \beta_n k^{-n},\\
\gamma &= \sum_{n=1}^{\infty} \gamma_n k^{-n},\\
W &= E^{(-2)}k + \sum_{n=0}^{\infty}\sum_{m=0}^{n+2} W_{n,m}x^m(\sqrt{k})^{-n},\\
\mathscr{U} &= \sum_{n=0}^{\infty}\sum_{m=0}^{n+1}\left\{\frac{D_m^n}{2m}x^{2m} + \frac{C_m^n k^{-1/2}x^{2m+1}}{2m+1}\right\}k^{-n},\\
D_0^n &= 0.
\end{aligned} \tag{A4}$$

The combination $\mathscr{U}'^2 + \mathscr{U}'' - 2W + 2\mathscr{E}$ occurs for all states, and it is convenient to expand it in a power series as well:

$$\mathscr{U}'^2 + \mathscr{U}'' - 2W + 2\mathscr{E} = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\{T^{nm}k^{-n}x^{2m} + S^{nm}k^{-(n+\frac{1}{2})}x^{2m+1}\}. \tag{A5}$$

For all three states, the coefficients in $\mathscr{U}$ and $\mathscr{E}$ are determined from T^{nm}, S^{nm} by the following relations:

$$D^{nm} = \frac{1}{2D_1^0}\Big\{T^{nm} + 2W_{2n,2m} - (2m+1)D^n_{m+1} - \sum_{\nu=1}^{n-1}\sum_{\mu} D^\nu_\mu D^{n-\nu}_{m+1-\mu} - \sum_{\nu=0}^{n-1}\sum_{\mu} C^\nu_\mu C^{n-\nu-1}_{m-\mu}\Big\},$$

$$C^{nm} = \frac{1}{2D_1^0}\Big\{S^{nm} + 2W_{2n+1,2m+1} - 2(m+1)C^n_{m+1} - 2\sum_{\nu=1}^{n}\sum_{\mu} D^\nu_\mu C^{n-\nu}_{m+1-\mu}\Big\},$$

$$C^n_{n+2} \equiv D^n_{n+2} \equiv 0, \tag{A6}$$

$$E^{(n-1)} = \frac{1}{2}\Big(T^{n0} - D^n_1 + 2W_{2n,0} - \sum_{\nu=0}^{n-1} C_0^{n-1-\nu} C_0^\nu\Big),$$

$$D_1^0 = -\sqrt{2W_{02}}.$$

The limits on the sums over μ in Eq. (A6) are such that only those coefficients defined in Eq. (A4) appear. The T^{nm} and S^{nm} depend on the state. For the ground state

$$T^{nm} = S^{nm} = 0. \tag{A7}$$

For the first excited state,

$$T^{nm} = \sum_{k=1}^{n-m} \alpha_k S^{n-k,m} - 2D^n_{m+1},$$

$$S^{nm} = \sum_{k=1}^{n-m} \alpha_k T^{n-k+1,m+1} - 2C^n_{m+1}, \tag{A8}$$

$$\alpha_n = \frac{\Big(2C_0^{n-1} - \sum_{k=1}^{n-1} \alpha_k T^{n-k,0}\Big)}{T^{00}}.$$

For the second excited state,

$$T^{mm} = -\sum_{k=1}^{n-m} (2\gamma_k C^{n-k}_{m+1} + \beta_k T^{n-k+1,m+1} + \gamma_k S^{n-k,m}) - 4D^n_{m+1},$$

$$S^{nm} = -\sum_{k=1}^{n-m} \{2\gamma_k D^{n-k+1}_{m+2} + \gamma_k T^{n-k+1,m+1} + \beta_k S^{n-k+1,m+1}\} - 4C^n_{m+1}, \tag{A9}$$

$$\beta_k = \frac{-\sum_{k=1}^{n-1}(2\gamma_k C_0^{n-k-1} + \beta_k T^{n-k,0}) - 2\delta_{n,1}}{T^{00}}, \quad \gamma_k = \frac{-\sum_{k=1}^{n-1}(2D^{n-k} + T^{n-k,0})\gamma_k - \sum_{k=1}^{n}\beta_k S^{n-k,0} - 4C_0^{n-1}}{T^{00} + 2D_1^0}.$$

[1]P. Anderson, Phys. Rev. **86**, 694 (1952).
[2]R. Kubo, Phys. Rev. **87**, 568 (1952).
[3]C. Kittle, *Quantum Theory of Solids* (Wiley, New York, 1963).
[4]L. D. Mlodinow, Berkeley thesis (1981).
[5]R. A. Ferrell and D. J. Scalapino, Phys. Rev. A **9**, 846 (1974).
[6]A. J. Bray, J. Phys. A 7, 2144 (1974).
[7]S. Hikami and E. Brézin, J. Phys. A **12**, 759 (1979).
[8]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. **128**, 314 (1980).
[9]L. D. Mlodinow and N. Papanicolaou, Ann. Phys. **131**, 1 (1981).
[10]L. D. Mlodinow, in *Progress in Particle and Nuclear Physics, Quarks and the Nucleus*, edited by D. Wilkenson, Proc. of 1981 Int. School of Nuclear Physics, Erice, Italy (Pergamon, New York 1982), Vol. 8.
[11]C. Bender, L. D. Mlodinow, and N. Papanicolaou, Phys. Rev. A **25**, 1305 (1982).
[12]R. Garstang, Rep. Prog. Phys. **40**, 105 (1977).
[13]J. E. Avron, I. W. Herbst, and B. Simon, Phys. Lett. A **62**, 214 (1977).
[14]J. E. Avron, Ann. Phys. **131**, 73 (1981).
[15]L. D. Landau and E. M. Lifschitz, *Quantum Mechanics: Non Relevatistic Theory* (Pergamon, New York, 1965), pp. 158–170.
[16]C. Quigg and J. L. Rosner, Phys. Rep. **56**, 206 (1979).
[17]M. Cl. Dumont-Lepage, N. Gani, J. P. Gazeau, and A. Ronveaux, J. Phys. A **13**, 1243 (1980).
[18]C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York, 1978), pp. 369–375.
[19]E. Combes and R. Seiler, in *Spectral Properties of Atomic and Molecular Systems*, edited by G. Wooley (1980).
[20]J. Cizek and E. Vrscay, Int. J. Quantum Chem. **21**, 27 (1982).

Unitary equivalence and scattering theory for Stark-like Hamiltonians

Matania Ben-Artzi
Department of Mathematics, Technion—Israel Institute of Technology, Haifa 32000, Israel

(Received 25 March 1983; accepted for publication 24 June 1983)

Let $H_0 = -\Delta + V_0(x_1)$, $H = H_0 + V(x)$ be self-adjoint in $L^2(R^n)$. V_0 depends only on one coordinate and tends monotonically to $\mp\infty$ as $x_1 \to \pm\infty$. V is a real H_0-compact potential, short range with respect to V_0. In particular, the cases $V_0(x_1) = -(\operatorname{sgn} x_1)|x_1|^\alpha$, $0<\alpha\leqslant 2$ and $|V(x)|\leqslant C|x|^{-1}$ are included ($\alpha = 1$ being the Stark effect). It is shown that (a) H is spectrally absolutely continuous over the entire real axis apart from a possible discrete sequence of eigenvalues of finite multiplicity and rapidly decaying eigenfunctions (H_0 has no eigenvalues) and (b) the wave operators $W_\pm(H,H_0)$ exist and are complete.

PACS numbers: 03.65.Nk, 02.30. + g

I. INTRODUCTION

In this work we study some spectral and scattering properties of the operator $H = H_0 + V$, $H_0 = -\Delta + V_0(x_1)$, in $L^2(R^n)$. Here $V_0(x_1)$ is a monotone decreasing real function depending on one coordinate only and tending to $\mp\infty$ as $x_1 \to \pm\infty$. The (real) function V is a multiplication operator in $L^2(R^n)$ (depending in general on all coordinates), which is assumed to be relatively compact with respect to H_0.

The terminology "Stark-like Hamiltonian" is derived from the quantum-mechanical Stark effect (for the hydrogen atom) where (with suitable scaling) $V_0(x_1) = -x_1$ and $V(x) = 1/|x|$, corresponding to the effect of the application of an external uniform electric field to the atom.[1] Naturally, the Stark effect is always included in our admissible class of Hamiltonians H.

Assuming the self-adjointness of H_0,H, two basic questions arise in this connection.[2]

Q 1: Impose conditions on V_0,V so that H_0,H have no singularly continuous spectrum, i.e., apart from eigenvalues the spectrum is absolutely continuous.

The wave operators $W_\pm(H,H_0)$ are defined by

$$W_\pm(H,H_0) = \operatorname*{s-lim}_{t\to\pm\infty} \exp(itH)\exp(-itH_0)P_0,$$

where P_0 is the projection on the subspace of absolute continuity of H_0 in $L^2(R^n)$. When they exist, they are partial isometries with domain $P_0L^2(R^n)$ and range contained in $PL^2(R^n)$, the subspace of absolute continuity of H. The second question is therefore:

Q 2: Find conditions on V_0,V so that

(a) $W_\pm$ exist,

(b) $W_\pm$ are complete in the sense that $\mathrm{Range}(W_\pm) = \mathrm{PL}^2(R^n)$.

It is well known that when $W_\pm$ are complete, they intertwine H,H_0 in the sense that

$$W_\pm^* PHW_\pm = P_0H_0$$

and thus establish a unitary equivalence between the absolutely continuous parts of H_0,H.

In the following sections we shall deal with both Q1 and Q2. To illustrate our results, we take the special case

$$H_{0,\alpha} = -\Delta + V_{0,\alpha}(x_1),\quad H_\alpha = H_{0,\alpha} + V,$$
$$V_{0,\alpha}(x_1) = (-\operatorname{sgn} x_1)|x_1|^\alpha,\quad 0<\alpha\leqslant 2.$$

Let

$$x\in R^n,\quad x=(x_1,x'),\quad x'=(x_2,\ldots,x_n).$$

$H^2(R^n)$ are functions having L^2 distributional derivatives up to second order.

Suppose that $V(x) = V_1(x) + V_2(x)$, where $V_1(x)$ is compactly supported and compact from $H^2(R^n)$ into $L^2(R^n)$ and $V_2(x)$ is bounded and satisfies

$$|V_2(x)|\leqslant CM(x')\cdot\begin{cases}(1+|x_1|)^{\alpha-\beta}, & x_1\leqslant 0,\\ (1+|x_1|)^{\alpha/2-1-\beta}, & x_1\geqslant 0,\end{cases}$$

where $\beta>0$ and $M(x')\to 0$ as $|x'|\to\infty$.

We shall prove:

Theorem A: $H_{0,\alpha},H_\alpha$ are essentially self-adjoint when defined on $C_0^\infty(R^n)$. Retaining the same notation for their unique sef-adjoint extensions, it follows that:

(a) $H_{0,\alpha}$ is absolutely continuous and its spectrum extends over the entire real axis.

(b) The set $\mathscr{N}$ of eigenvalues of H_α is at most a discrete sequence. Each eigenvalue is of finite multiplicity and the corresponding eigenfunctions u are rapidly decaying in the sense that, for every real s, $(1+|x_1|)^s u\in L^2(R^n)$.

(c) H_α is absolutely continuous on the orthogonal complement of the subspace spanned by its eigenfunctions, and its spectrum extends over the entire real axis. In particular, there is no singularly continuous spectrum.

(d) The wave operators exist and are complete, thereby establishing a unitary equivalence between the absolutely continuous parts of $H_{0,\alpha},H_\alpha$.

Our method of proof utilizes the so-called "limiting absorption principle," namely, the observation that the resolvent operator remains bounded when viewed as an element of $B(X,Y)$, where X,Y are Hilbert spaces such that $X\subseteq L^2(R^n)\subseteq Y\subseteq X^*$ and X dense in $L^2(R^n)$. In the case discussed above we define, for $\sigma>\frac{1}{2}$, $\tau>1$, $u\in C_0^\infty(R^n)$,

$$\|u\|_{\tau,\sigma}^2 = \int_{-\infty}^0\int_{R^{n-1}}(1+|x_1|)^{2\tau-\alpha}|u(x_1,x')|^2\,dx_1\,dx' + \int_0^\infty\int_{R^{n-1}}(1+|x_1|)^{2\sigma-\alpha/2}|u(x_1,x')|^2\,dx_1\,dx',$$

$$|||u|||^2_{\tau,\sigma} = \int_{-\infty}^0 \int_{R^{n-1}} (1+|x_1|)^{2\tau+\alpha}|u(x_1,x')|^2\,dx_1\,dx' + \int_0^\infty \int_{R^{n-1}} (1+|x_1|)^{2\sigma+\alpha/2}|u(x_1,x')|^2\,dx_1\,dx'.$$

Denote by $X^2_{\tau,\sigma}, Y_{\tau,\sigma}$ the completions of $C_0^\infty(R^n)$ under $\|\ \|_{\tau,\sigma}, |||\ |||_{\tau,\sigma}$, respectively.

For nonreal z we set $R_{0,\alpha}(z) = (H_{0,\alpha} - z)^{-1}$, $R_\alpha(z) = (H_\alpha - z)^{-1}$, and if K is a compact interval, we denote

$$\Omega^{\pm}(K) = \{z \mid \operatorname{Re} z \in K,\ 0 < \pm \operatorname{Im} z \leqslant 1\},$$

$$\tilde{\Omega}^{\pm}(K) = \Omega^{\pm}(K) \cup K.$$

We shall prove:

Theorem B: Let $\tau > 1$, $\sigma > \frac{1}{2}$, and z be nonreal. Then,

$$R_{0,\alpha}(z) \text{ and } R_\alpha(z) \in B(X_{\tau,\sigma}, Y_{\tau,-\sigma}).$$

Furthermore, if K is compact and $\tau' < \tau$, the operator-valued functions [with values in $B(X_{\tau,\sigma}, Y_{\tau',-\sigma})$],

$$z \to R_{0,\alpha}(z), \quad z \to R_\alpha(z)$$

can be extended to $\tilde{\Omega}^{\pm}(K)$ as continuous maps, where the operator space is endowed with the uniform topology. In the statement for $R_\alpha(z)$ we add the restriction that $K \cap \mathscr{N} = \varnothing$.

Note that our estimates (Sec. II, IV) will be more precise, involving also first-order derivatives. Also, for more general potentials V_0 the norms involved in the definition of X, Y have to be suitably selected.

It is interesting to note that in contrast with the perturbation-theoretical computation of the Stark effect carried out in some physics books[3] we treat the Coulomb potential as a "perturbation" imposed on the electric field.

The absolute continuity for a general Stark-like potential was shown by Rejto–Sinha[4] in the one-dimensional case. Their assumptions on V_0 are quite similar to ours. However, $V(x_1)$ is required to decay faster than $|x_1|^{-1}$. The existence of wave operators in the general case was proved in Ben-Artzi.[5] Indeed, the assumptions there are considerably weaker than those in the present work. In particualr, V_0 is not assumed to be monotone and $V(x_1,x')$ is allowed to explode in directions other than $x_1 \to +\infty$.

The Stark case ($\alpha = 1$ in the notation above) has been treated more thoroughly. Avron and Herbst[6] showed the existence of wave operators as well as the nonexistence of eigenvalues in certain cases and the completeness was proved by Simon.[7] Herbst[8] and Yajima[9] proved in this context Theorems A and B. In particular, the norms and assumptions on V in the latter are very close to ours. In all of these works, special features of the potential $V_0(x_1) = -x_1$ are utilized, notably the expansion by Airy functions and the possibility of removing $-d^2/dx_1^2$ by an explicit transformation.

Our approach is based on studying solutions of

$$(H_0 - z)u = f,$$

which satisfy an appropriate radiation condition. For suitably restricted f, the existence and uniqueness of such solutions is guaranteed by some simple facts about the asymptotic behavior of solutions to ordinary differential equations, which are given, along with the statement of the limiting absorption principle for H_0, in Sec. II. The estimates needed to establish the theorem are proved in Sec. III. We use a representation of H_0 as an ordinary differential operator with operator-valued coefficients. This technique was used for perturbations of the Laplacian by Jäger[10] and Saito.[11] We will have more to say about it in Sec. VII, where some possible extensions of the present work are also discussed.

The limiting absorption principle for H is stated and proved in Sec. IV. For the perturbation-theoretic argument we need to study the invertibility of $I + R_0^{\pm}(\lambda)V$, where $R_0^{\pm}(\lambda)$ are the limiting values of $(H_0 - z)^{-1}$.

In the proof we use Lemma 4.3, which ensures the invertibility of these operators except for a finite (in a compact interval) number of points λ, which represent eigenvalues of finite multiplicity.

Section V is devoted to the proof of this lemma, including the rapid decay of eigenfunctions. This is technically the hardest part of the paper. The idea here is to use a bootstrap procedure for solutions of $(H_0 - \lambda)u = -Vu$, along with a spectral decomposition of H_0. Thus, we upgrade first the decay rate of individual components of u, and then patch them up together, to yield an estimate for u. In Sec. VI the existence and completeness of wave operators are derived from a general theorem of Kato and Kuroda[12] and the limiting absorption principle.

Finally, let us list our assumptions on the basic one-dimensional potential $V_0(t)$:

(V_0) $V_0(t) \in L^2(R)_{\text{loc}}$ and $V_0(t) = V_L(t) + V_S(t)$, where

$(VL\ 1)$ $V_L(t) \in C^2(-\infty, 0] \cap C^2[0,\infty)$ and with some $\delta > \frac{1}{2}$,

$$(1 + |V_L(t)|)^{-1}(d/dt)^j V_L(t) = O(|t|^{-j\delta}) \quad \text{as } |t| \to \infty,\ j = 1,2.$$

Note that we allow a finite jump at $t = 0$.

$(VL\ 2)$ $V_L(t)$ is strictly decreasing on $(-\infty, 0]$, $[0,\infty)$, and there exist constants T, c_0, $\delta_1 > 0$, and $c > 1$ such that

(i) $ct^2 \geqslant |V_L(t)| \geqslant c^{-1}|t|^{\delta_1}$, $\quad |t| \geqslant T$,

(ii) $-V_L'(t) \geqslant c_0(1 + |V_L(t)|)t^{-1}$, $\quad t \geqslant T$.

(VS) For some $\delta_2 > 0$,

$$|V_S(t)|(1 + |V_L(t)|)^{-1/2} = O(|t|^{-1-\delta_2}) \quad \text{as } |t| \to \infty.$$

It follows from Ref. 13 that the restriction of $H_0 = -\Delta + V_0$ to $C_0^\infty(R^n)$ is essentially self-adjoint. Its spectrum extends over the entire real axis and is absolutely continuous.

The detailed assumptions on the potential V are given in Sec. IV [see (4.1) for a simplified version].

II. THE LIMITING ABSORPTION PRINCIPLE FOR H_0

Let $I = [\alpha, \beta]$ be a finite interval on the real line. We set

$$\Omega^{+}(I) = \{z \mid \operatorname{Re} z \in I, 0 < \operatorname{Im} z \leqslant 1\},$$

$$\tilde{\Omega}^{+}(I) = = \{z \mid \operatorname{Re} z \in I, 0 \leqslant \operatorname{Im} z \leqslant 1\}. \tag{2.1}$$

In this section we study solutions of

$$(H_0 - z)u = f, \quad z \in \tilde{\Omega}^{+}(I). \tag{2.2}$$

H_0 is self-adjoint; hence for $\operatorname{Im} z > 0$ and $f \in L^2(R^n)$ the solution $u = (H_0 - z)^{-1} f = R_0(z) f$ is uniquely determined in $L^2(R^n)$. In fact, standard theorems on the regularity of solutions to elliptic equations imply that $u \in H^2(R^n)_{\text{loc}}$, where H^2 is the space of functions having L^2 distributional derivatives up to second order.

We intend to show that, using a suitable topology, $R_0(z)$ can be extended continuously to $\widetilde{\Omega}^+(I)$. To this end, we introduce the following norms for $u \in C_0^\infty(R^n)$. We denote $x = (x_1, x') \in R^n$, $x' = (x_2, ..., x_n)$. Recall that $V_L(x_1)$ is the smooth part of $V_0(x_1)$.

Let $\tau, \sigma \in$ R and $u \in C_0^\infty(R^n)$:

$$\begin{aligned} \|u\|_{\tau,\sigma}^2 = & \int_{-\infty}^0 \int_{R^{n-1}} (1 - x_1)^{2\tau} (1 + |V_L(x_1)|)^{-1} \\ & \times |u(x_1, x')|^2 \, dx_1 \, dx' \\ & + \int_0^\infty \int_{R^{n-1}} (1 + x_1)^{2\sigma} (1 + |V_L(x_1)|)^{-1/2} \\ & \times |u(x_1, x')|^2 \, dx_1 \, dx', \end{aligned} \tag{2.3}$$

$$\begin{aligned} |||u|||_{\tau,\sigma}^2 = & \int_{-\infty}^0 \int_{R^{n-1}} (1 - x_1)^{2\tau} \\ & \times \left[(1 + |V_L(x_1)|) |u(x_1, x')|^2 + \left| \frac{\partial u}{\partial x_1} \right|^2 \right] dx_1 \, dx' \\ & + \int_0^\infty \int_{R^{n-1}} (1 + x_1)^{2\sigma} \\ & \times (1 + |V_L(x_1)|)^{1/2} [|u(x_1, x')|^2 \\ & + (1 + |V_L(x_1)|)^{-1} \left| \frac{\partial u}{\partial x_1} \right|^2] \, dx_1 \, dx', \end{aligned} \tag{2.4}$$

$$|||u|||_{2,\tau,\sigma}^2 = |||u|||_{\tau,\sigma}^2 + \|H_0 u\|_{\tau,\sigma}^2.$$

A norm similar to $\| \ \|_{\tau,\sigma}$ was employed by Yajima[14] for the Stark case $V_L(x_1) = -x_1$.

Denote by $\Psi_{\tau,\sigma}$, $\Phi_{\tau,\sigma}$, $\Phi_{2,\tau,\sigma}$ the completions of $C_0^\infty(R^n)$ under $\| \ \|_{\tau,\sigma}$, $||| \ |||_{\tau,\sigma}$, $||| \ |||_{2,\tau,\sigma}$, respectively.

We shall prove:

Theorem 2.1: Let $f \in C_0^\infty(R^n)$, $z \in \Omega^+(I)$ and $u = R_0(z) f$. Then for every $\tau > 1$, $\sigma > \frac{1}{2}$ there exists a constant C (depending only on σ, τ, I, and the operator) such that

$$|||u|||_{2,\tau,-\sigma} \leqslant C \|f\|_{\tau,\sigma}. \tag{2.5}$$

Thus $R_0(z)$ can be extended by continuity as a map in $B(\Psi_{\tau,\sigma}, \Phi_{2,\tau,-\sigma})$, for every $z \in \Omega^+(I)$.

Furthermore, for every $\tau' < \tau$ the operator-valued map $z \to R_0(z) \in B(\Psi_{\tau,\sigma}, \Phi_{2,\tau',-\sigma})$ can be extended continuously to $\widetilde{\Omega}^+(I)$, where the operator space is equipped with the uniform topology. This extended map will be denoted by $R_0^+(z)$.

In the course of the proof, we shall see that functions of the form $R_0^+(z) f$, $f \in \Psi_{\tau,\sigma}$, satisfy a "radiation condition" in the x_1 direction. In fact, this condition will be employed to prove uniqueness for solutions of (2.2).

Observe that if we establish the theorem for some $\sigma_1 > \frac{1}{2}$, then it holds true for every $\sigma > \sigma_1$. In what follows we therefore assume that σ is always subject to the following restriction.

$$\frac{1}{2} < \sigma < \min\left(\frac{1+\delta}{2}, \frac{1+\delta_1}{2}, \frac{1+\delta_2}{2}, \frac{1+c_0}{2}, 2\delta - \frac{1}{2} \right). \tag{2.6}$$

The constants on the right-hand side are those appearing in the assumptions on V_L, V_S in the Introduction.

The proof of Theorem 2.1 is rather long. It will be obtained in the next section as a result of certain estimates for solutions of (2.2) with specified asymptotic behavior, which we call "radiative." These estimates (and uniqueness and existence) of the radiative solutions will be the contents of Lemmas 2.3, 2.4, 3.1, preceding the proof of the theorem. As a matter of fact, the estimates we obtain are stronger than (2.5) in the sense that they also provide some information on the first-order x' derivatives of u.

Taking the Fourier transform with respect to x' we see that H_0 is unitarily equivalent to the direct integral

$$\begin{aligned} H_0 \simeq & \int_{R^{n-1}} \oplus \left(-\frac{d^2}{dx_1^2} + V_0(x_1) + |\xi'|^2 \right) d\xi': \\ = & \int_{R^{n-1}} \oplus H_{0,\xi'} \, d\xi'. \end{aligned} \tag{2.7}$$

The direct integral is self-adjoint in $\int_{R^{n-1}} \oplus L^2(R) \, d\xi'$.

Theorem 4.1 of Ben-Artzi[15] implies that, for every $\xi' \in R^{n-1}$, $H_{0,\xi'}$ is self-adjoint in $L^2(R)$ and its essential spectrum is R and is totally absolutely continuous.

The direct integral in (2.7) may be viewed as an ordinary differential operator with operator-valued coefficients in $L^2(R, X)$, where $X = L^2(R_{\xi'}^{n-1})$. We shall denote by $\| \ \|_X$ and $(\ , \ )_X$ the norm and scalar product in X.

Equation (2.2) now takes the form:

$$\begin{aligned} & \left(-\frac{d^2}{dx_1^2} + V_0(x_1) + |\xi'|^2 - z \right) u(x_1, \xi') = f(x_1, \xi'), \\ & z \in \widetilde{\Omega}^+(I). \end{aligned} \tag{2.8}$$

[We retain the same notation u, f for the partial transforms. Since the norms in (2.3)–(2.4) are given in terms of $\|u(x_1, \cdot)\|_{L^2(R^{n-1})}$, this will make no difference in view of the Parseval relation.]

Let $R_\pm = \{ x_1 / \pm x_1 \geqslant 0 \}$. For a function f defined on R, denote by $f_\pm$ its restrictions to $R_\pm$. Let $f \in L^2(R, X)_{\text{loc}}$, $r \in R$, and $\eta(x_1) > 0$ be a continuous function on R. We define the following norms:

$$\begin{aligned} \|f_\pm\|_r^2 = & \int_{R_\pm} (1 + |x_1|)^{2r} \|f(x_1)\|_X^2 \, dx_1, \\ \|f_\pm\|_{r,\eta} = & \|\eta^{1/2} f_\pm\|_r. \end{aligned} \tag{2.9}$$

We now set:

$C_0^\infty(R, X)$—X- valued, compactly supported and infinitely strongly differentiable functions on R.

$C_0^\infty(R_\pm, X)$—Restriction of $C_0^\infty(R, X)$ to $R_\pm$.

$L_r^2(R_\pm, X)$—Completion of $C_0^\infty(R_\pm, X)$ with respect to $\|f_\pm\|_r$.

$L_{r,\eta}^2(R_\pm, X)$—Completion of $C_0^\infty(R_\pm, X)$ with respect to $\|f_\pm\|_{r,\eta}$.

$$H^1(R, X) = \left\{ f \,\middle|\, \frac{df}{dx_1} \in L^2(R, X), \quad f \in L^2(R, Y) \right\},$$

where Y is the domain of multiplication by $|\xi'|$ in $X=L^2(R^{n-1}_{\xi'})$.

We suppress the symbol X in the case $X=C$.

Going back to Eq. (2.8), we apply in $R_\pm$ transformations given by a change of variable:

$$dr=Q_\pm\, dx_1,\quad x_1\in R_\pm,$$
$$Q_\pm(x_1,z)=\{\pm[z-V_L(x_1)]\}^{1/2}. \tag{2.10}$$

Here and below we take the branch of the logarithm which is defined and continuous in $C\setminus(-\infty,0]$. Thus, $z^{1/2}$ and $z^{1/4}$ are continuous and positive across the positive real axis.

Note that $Q_\pm$ are well defined with $\operatorname{Re} Q_\pm>0$ for $z\in\widetilde{\Omega}^+(I)$ by rearranging, if necessary, the sum V_S+V_L and allowing for a finite jump of V_L at $x_1=0$.

The new variable r traces out curves $C_\pm(z)$ in the complex plane, parametrized by x_1. Using (2.10), Eq. (2.8) can be rewritten as (see Ben-Artzi[16])

$$\left(-\frac{d^2}{dr^2}+F_\pm+P_\pm\mp 1\right)v_\pm=g_\pm, \tag{2.11}$$
$$F_\pm=Q_\pm^{-2}|\xi'|^2, \tag{2.12}$$
$$v_\pm=Q_\pm^{1/2}u_\pm,\quad g_\pm=Q_\pm^{-3/2}f_\pm, \tag{2.13}$$
$$P_\pm=O(|Q_\pm|^{-1}(1+|x_1|)^{-1-\epsilon}\quad\text{as } |x_1|\to\infty, \tag{2.14}$$
$$\epsilon=\min(2\delta-1,\delta_2)\quad(\text{see the assumptions on } V_L).$$

The differentiation d/dr in (2.11) may be interpreted as performed along $C_\pm(z)$ or else replaced by $Q_\pm^{-1}\,d/dx_1$.

In the sequel we shall need the following weight function

$$\omega(x_1)=[1+|V_L(x_1)|]^{1/2}. \tag{2.15}$$

Note that we have a constant $c>1$ so that uniformly in $z\in\widetilde{\Omega}^+(I)$ and $x_1\in R_\pm$

$$\pm\operatorname{Im} Q_\pm(x_1,z)\geqslant 0, \tag{2.16}$$
$$c^{-1}\omega(x_1)\leqslant|Q_\pm(x_1,z)|\leqslant c\omega(x_1), \tag{2.17}$$
$$\left|\frac{d}{dx_1}Q_\pm(x_1,z)\right|\leqslant c\omega(x_1)(1+|x_1|)^{-\delta} \tag{2.18}$$

[(2.18) follows from assumption (*VL 1*) above].

In what follows we refer to the pair $(v_\pm,g_\pm)$, defined by (2.13), as the one associated with (u,f). They are X-valued functions defined on $R_\pm$.

If $f\in L^2(R,X)_{\rm loc}$, we say that u is a weak solution of (2.8) if for every $\varphi\in C_0^\infty(R_{x_1}\times R^{n-1}_{\xi'})$ we have

$$\int_R(u(x_1,\cdot),(H_{0,\xi'}-z)\varphi(x_1,\cdot))_X\,dx_1 = \int_R(f(x_1,\cdot),\varphi(x_1,\cdot))_X\,dx_1.$$

In the following definition we use the terminology of Saito[17]:

Definition 2.2: Let $f\in L^2(R,X)_{\rm loc}$ and let $u\in H^1(R,X)_{\rm loc}$ be a weak solution of (2.8). Let $(v_\pm,g_\pm)$ be the functions associated with (u,f) by (2.13).

u is called the *radiative function* corresponding to (f,z) [denoted $u=\operatorname{rad}(f,z)$] if

$$\frac{d}{dr}v_+-iv_+\in L^2_{\sigma-1}(R_+,X),\quad\text{some } \sigma>\tfrac12, \tag{2.19}$$
$$\frac{d}{dr}v_-,\quad v_-\in L^2_\tau(R_-,X),\quad\text{some } \tau>1. \tag{2.20}$$

We could, of course, formulate (2.19)–(2.20) in terms of the function u. However, this would lead to a more complicated form, involving $Q_\pm$, and would also complicate the estimates below.

The following two lemmas deal with the uniqueness and existence of the radiative function when f is suitably restricted.

Lemma 2.3 (*Uniqueness*): If $u=\operatorname{rad}(0,z)$, then $u\equiv 0$.

Lemma 2.4 (*Existence*): Let $f\in C_0^\infty(R_{x_1}\times R^{n-1}_{\xi'})$. Then $u=\operatorname{rad}(f,z)$ exists, for every $z\in\widetilde{\Omega}^+(I)$.

If $\operatorname{Im} z>0$, $u=R_0(z)f$.

Proof of Lemma 2.3: For a.e. $\xi'\in R^{n-1}$ the function $u(x_1,\xi')$ satisfies the equation

$$(H_{0,\xi'}-z)u(x_1,\xi') \equiv\left(-\frac{d^2}{dx_1^2}+V_0(x_1)+|\xi'|^2-z\right)u(x_1,\xi')=0 \tag{2.21}$$

and the corresponding functions $v_\pm(x_1,\xi')$ satisfy the one-dimensional radiation conditions

$$\frac{d}{dr}v_+(x_1,\xi')-iv_+(x_1,\xi')\in L^2_{\sigma-1}(R_+), \tag{2.22}$$
$$\frac{d}{dr}v_-(x_1,\xi'),\quad v_-(x_1,\xi')\in L^2_\tau(R_-). \tag{2.23}$$

The asymptotic behavior of solutions of (2.21) is readily derived from Ref. 18. It follows that, for $x_1\geqslant a>0$, a large enough, there are two linearly independent solutions φ_1,φ_2 such that

$$\varphi_1(x_1,\xi',z)=[z-|\xi'|^2-V_L(x_1)]^{-1/4}\times\exp\left\{i\int_a^{x_1}[z-|\xi'|^2-V_L(s)]^{1/2}\,ds\right\}\times[1+h(x_1,\xi',z)], \tag{2.24}$$

$$\frac{d\varphi_1}{dx_1}(x_1,\xi',z)=\{i[z-|\xi'|^2-V_L(x_1)]^{1/2}+\tfrac14 V_L'(x_1)[z-|\xi'|^2-V_L(x_1)]^{-1}\}\times\varphi_1\cdot[1+h_1(x_1,\xi',z)],$$
$$|h(x_1,\xi',z)|+|h_1(x_1,\xi',z)|=O((1+x_1)^{-1-\epsilon})$$
$$\text{as } x_1\to+\infty,\quad \epsilon>0.$$

φ_2 is obtained from φ_1 by changing i to $-i$ and replacing h,h_1 by functions with similar properties.

It follows easily from the assumptions on V_L that

$$[z-V_L(x_1)]^{1/4}[z-|\xi'|^2-V_L(x_1)]^{-1/4}-[z-|\xi'|^2-V_L(x_1)]^{1/4}[z-V_L(x_1)]^{-1/4}\in L^2_{\sigma-1}(x_1\geqslant a)$$

if $\sigma-\frac12$ is small enough.

This implies that

$$\frac{1}{Q_+}\frac{d}{dx_1}(Q_+^{1/2}\varphi_1) - iQ_+^{1/2}\varphi_1 \in L^2_{\sigma-1}(R_+).$$

A similar computation shows that φ_2 does not satisfy (2.22) for any $\sigma > \frac{1}{2}$ so that for some constant $\alpha = \alpha(\xi',z)$,

$$u(x_1,\xi') = \alpha\varphi_1(x_1,\xi',z). \tag{2.25}$$

Turning now to the asymptotic behavior of solutions of (2.21) for $x_1 \leqslant -a$, it follows again from Ben-Artzi[19] that there exist two linearly independent solutions ψ_1, ψ_2 such that, for $x_1 \leqslant -a$,

$$\psi_1(x_1,\xi',z) = [V_L(x_1) + |\xi'|^2 - z]^{-1/4}\exp\left(\int_{-a}^{x_1}(V_L(s) + |\xi'|^2 - z)^{1/2}\,ds\right)[1 + g(x_1,\xi',z)], \tag{2.26}$$

$$\frac{d\psi_1}{dx_1}(x_1,\xi',z) = \left\{[V_L(x_1) + |\xi'|^2 - z]^{1/2} - \frac{1}{4}V_L'(x_1)[V_L(x_1) + |\xi'|^2 - z]^{-1}\right\}\psi_1\cdot[1 + g_1(x_1,\xi',z)],$$

$$|g(x_1,\xi',z)| + |g_1(x_1,\xi',z)| = O((1-x_1)^{-1-\epsilon}) \quad \text{as } x_1 \to -\infty,\ \epsilon > 0.$$

ψ_2 is obtained from ψ_1 by multiplying the exponent and the first term in $\{\cdots\}$ by (-1) and replacing g, g_1 by functions with similar decay properties.

Thus ψ_1 and its derivative decay exponentially as $x_1 \to -\infty$, whereas ψ_2 blows up exponentially. The same considerations as those leading to (2.25) now imply

$$u(x_1,\xi') = \beta\psi_1(x_1,\xi',z). \tag{2.27}$$

Assume first that $\operatorname{Im} z > 0$. It follows from (2.25), (2.27), and the asymptotic behavior of φ_1,ψ_1 that $u(x_1,\xi')$ is an eigenfunction of $H_{0,\xi'}$, with a nonreal eigenvalue z, unless $u(x_1,\xi') \equiv 0$.

Assume now that $z = \lambda \in I$. As is well known, the Wronskian $W(u,\bar{u})$ is independent of x_1. However, using the asymptotic expressions above, we find

$$W(u,\bar{u}) = W(\alpha\varphi_1, \overline{\alpha\varphi_1}) = -2i|\alpha|^2,$$

$$W(u,\bar{u}) = W(\beta\psi_1, \overline{\beta\psi_1}) = 0.$$

This leads again to $u(x_1,\xi') \equiv 0$.

Q.E.D.

Proof of Lemma 2.4: Set

$$\Lambda = \{\xi'/(x_1,\xi') \in \operatorname{supp} f \text{ for some } x_1 \in R\}$$

For $\xi' \notin \Lambda$ set $u(x_1,\xi') = 0$.

For $\xi' \in \Lambda$ we solve (2.21) using the kernel $K(s,t,\xi',z)$ of $(H_{0,\xi'} - z)^{-1}$. Since Λ is compact, it follows from Ben-Artzi[20] that K can be extended as a continuous function on $R \times R \times \Lambda \times \widetilde{\Omega}^+(I)$. Furthermore, for some continuous function $w(\xi',z)$ that never vanishes on $\Lambda \times \widetilde{\Omega}^+(I)$ we have

$$K(s,t,\xi',z) = K(t,s,\xi',z) = w(\xi',z)\varphi_1(s,\xi',z)\psi_1(t,\xi',z), \quad s > t.$$

The lemma now follows from the fact that φ_1,ψ_1 satisfy (2.22), (2.23), respectively, as observed in the process of the previous proof. Q.E.D.

III. ESTIMATE FOR THE RADIATIVE FUNCTION AND PROOF OF THEOREM 2.1

We now give an estimate for $u = \operatorname{rad}(f,z)$. However, we formulate it in terms of the associated pair $(v_\pm, g_\pm)$ [(2.13)]. We use the weight function $\omega(x_1)$ of (2.15)

Lemma 3.1: Let $\tau > 1$, σ as in (2.6) and $z \in \widetilde{\Omega}^+(I)$. Let $f \in \Psi_{\tau,\sigma}$. Then:

(a) $u = \operatorname{rad}(f,z)$ exists and is unique. The radiation conditions (2.22), (2.23) are satisfied with the same τ,σ.

(b) There exists a constant C, depending only on τ,σ, I, and the operator, such that

$$\|v_+\|_{-\sigma} + \|v_-\|_{\tau,\omega} + \left\|\frac{d}{dr}v_+ - iv_+\right\|_{\sigma-1} + \left\|\frac{d}{dr}v_-\right\|_{\tau,\omega}$$
$$+ \|\,|\xi'|v_+\|_{\sigma-1,\omega^{-2}} + \|\,|\xi'|v_-\|_{\tau,\omega^{-1}}$$
$$\leqslant C(\|g_+\|_{\sigma,\omega^2} + \|g_-\|_{\tau,\omega}). \tag{3.1}$$

Proof: In the process of the proof, we denote by C a generic constant that depends only on σ,τ, I, and the operator.

The uniqueness of $\operatorname{rad}(f,z)$ was proved in Lemma 2.3.

Assume first that $f \in C_0^\infty(R_{x_1} \times R_{\xi'}^{n-1})$. The existence of $\operatorname{rad}(f,z)$ was proved in Lemma 2.4. In the proof of that lemma it was shown that $u(x_1,\xi')$ is a scalar multiple of $\varphi_1(x_1,\xi',z)$ [resp. $\psi_1(x_1,\xi',z)$] for $x_1\geqslant x_1^0$ (resp. $x_1\leqslant -x_1^0$). Thus u and its derivative decay exponentially as $x_1\to-\infty$, whereas the validity of (2.22) with the same σ follows from (2.24) and (2.6). We now prove that

$$\left\|\frac{d}{dr}v_+ - iv_+\right\|_{\sigma-1} + \|\,|\xi'|v_+\|_{\sigma-1,\omega^{-2}} \leqslant C(\|v_+\|_{-\sigma} + \|g_+\|_{\sigma,\omega^2}), \tag{3.2}$$

$$\left\|\frac{d}{dr}v_-\right\|_{\tau,\omega} + \|v_-\|_{\tau,\omega} + \|\,|\xi'|v_-\|_{\tau,\omega^{-1}} \leqslant C\left(\|g_-\|_{\tau,\omega} + \left|\int_0^a\int_{R^{n-1}}|v_-(x_1,\xi')|^2\,dx_1\,d\xi'\right|^{1/2}\right). \tag{3.3}$$

Here $a<0$ is independent of v,g. To establish (3.2), write Eq. (2.11) in the form

$$-\frac{d}{dr}(v'_+ - iv_+) - i(v'_+ - iv_+) + F_+v_+ + P_+v_+ = g_+ \quad \left(v'_+ = \frac{dv_+}{dr}\right).$$

Multiplying this equation by Q_+ and taking the real part of its scalar product (in X) with $v'_+ - iv_+$, we get

$$\begin{aligned}-\frac{1}{2}\frac{d}{dx_1}\|v'_+ - iv_+\|_X^2 &+ (\operatorname{Im}Q_+)\|v'_+ - iv_+\|_X^2\\ &+ \operatorname{Re}Q_+^{-1}(|\xi'|^2v_+,v'_+ - iv_+)_X\\ &+ \operatorname{Re}(P_+Q_+v_+,v'_+ - iv_+)_X\\ &= \operatorname{Re}(Q_+g_+,v'_+ - iv_+)_X.\end{aligned}$$

$\operatorname{Im}Q_+\geqslant 0$ follows from (2.10) and

$$\begin{aligned}\operatorname{Re}Q_+^{-1}(|\xi'|^2v_+,-iv_+)_X &= |Q_+|^{-2}(\operatorname{Im}Q_+)(|\xi'|^2v_+,v_+)_X\geqslant 0,\\ \operatorname{Re}Q_+^{-1}(|\xi'|^2v_+,v'_+)_X &= \frac{1}{2}|Q_+|^{-2}\frac{d}{dx_1}\|\,|\xi'|v_+\|_X^2\end{aligned}$$

so that

$$-\frac{1}{2}\frac{d}{dx_1}\|v'_+ - iv_+\|_X^2 + \frac{1}{2}|Q_+|^{-2}\frac{d}{dx_1}\|\,|\xi'|v_+\|_X^2 \leqslant |(P_+Q_+v_+,v'_+ - iv_+)_X| + |(Q_+g_+,v'_+ - iv_+)_X|.$$

Multiplying this inequality by $(1+x_1)^{2\sigma-1}$, integrating over $[0,x_1]$, and noting (2.14) (where $2\sigma<1+\epsilon$), we have

$$\begin{aligned}&\left(\sigma-\frac{1}{2}\right)\int_0^{x_1}(1+t)^{2(\sigma-1)}\|v'_+ - iv_+\|_X^2dt\\ &\quad-\frac{1}{2}\int_0^{x_1}\frac{d}{dt}[(1+t)^{2\sigma-1}|Q_+|^{-2}]\|\,|\xi'|v_+\|_X^2\,dt\\ &\leqslant \tfrac{1}{2}(1+x_1)^{2\sigma-1}\|v'_+ - iv_+\|_X^2\\ &\quad+\tfrac{1}{2}|Q_+(0)|^{-2}\|\,|\xi'|v_+(0)\|_X^2\\ &\quad+C(\|v_+\|_{-\sigma}\|v'_+ - iv_+\|_{\sigma-1}\\ &\quad+\|g_+\|_{\sigma,\omega^2}\|v'_+ - iv_+\|_{\sigma-1}),\end{aligned}$$

where on the right-hand side we have used the Cauchy–Schwartz inequality and (2.17).

Our assumptions on V_L and (2.6) imply that for some $a\geqslant 0$,

$$\frac{d}{dt}[(1+t)^{2\sigma-1}|Q_+|^{-2}] \leqslant -k(1+t)^{2(\sigma-1)}\omega(t)^{-2},\quad k>0,\ t\geqslant a.$$

Standard elliptic estimates imply that since $u\in H^2(R^n)_{\rm loc}$ the trace $|\xi'|v_+(x_1)$ can be estimated by

$$\|\,|\xi'|v_+(x_1)\|_X\leqslant C(\|v_+\|_{-\sigma}+\|g_+\|_{\sigma,\omega^2}),\quad 0\leqslant x_1\leqslant a.$$

Also, the radiation condition (2.19) yields

$$\liminf_{x_1\to+\infty}(1+x_1)^{2\sigma-1}\|v'_+ - iv_+\|_X^2 = 0.$$

Incorporating these considerations in the last inequality and letting $x_1\to+\infty$ through an appropriate subsequence yields (3.2).

To prove (3.3) multiply Eq. (2.11) by $(1-x_1)^{2\tau}Q_-$ and take the real part of the scalar product (in X) of the resulting equation with v_-. Integrating over $(-\infty,x_1)$, $x_1<0$, we obtain

$$\begin{aligned}\operatorname{Re}\int_{-\infty}^{x_1}(1-t)^{2\tau}\Big[-\Big(\frac{d^2}{dt\,dr}v_-,v_-\Big)_X &+ (Q_-^{-1}|\xi'|^2v_-,v_-)_X\\ &+(P_-Q_-v_-,v_-)_X + Q_-\|v_-\|_X^2\Big]dt\\ = \operatorname{Re}\int_{-\infty}^{x_1}(1-t)^{2\tau}(Q_-g_-,v_-)_X\,dt.\end{aligned}$$

Now,

$$\begin{aligned}\operatorname{Re}\int_{-\infty}^{x_1}-(1-t)^{2\tau}\Big(\frac{d^2v_-}{dt\,dr},v_-\Big)_X dt &= \int_{-\infty}^{x_1}(\operatorname{Re}Q_-)(1-t)^{2\tau}\left\|\frac{dv_-}{dr}\right\|_X^2dt\\ &\quad-2\tau\operatorname{Re}\int_{-\infty}^{x_1}(1-t)^{2\tau-1}\Big(\frac{dv_-}{dr},v_-\Big)_X dt\\ &\quad-(1-x_1)^{2\tau}\operatorname{Re}\Big(\frac{dv_-}{dr},v_-\Big)_X.\end{aligned}$$

Note that, for some $c>0$,

$$\operatorname{Re}Q_-\geqslant c|Q_-|\to\infty\quad\text{as}\quad x_1\to-\infty$$

and for every $\eta>0$ small we have some $a<0$ such that

$$\begin{aligned}&|(P_-Q_-v_-,v_-)_X|\leqslant\eta\|v_-\|_X^2,\quad t<a,\\ &\left|(1-t)^{2\tau-1}\Big(\frac{dv_-}{dr},v_-\Big)_X\right| \leqslant\eta(1-t)^{2\tau}\left(\|v_-\|_X^2+\left\|\frac{dv_-}{dr}\right\|_X^2\right),\quad t<a,\\ &|(Q_-g_-,v_-)_X|\leqslant\eta|Q_-|\,\|v_-\|_X^2+\frac{1}{\eta}|Q_-|\,\|g_-\|_X^2.\end{aligned}$$

For $a<x_1\leqslant 0$, the trace norms $\|v_-\|_X,\|dv_-/dr\|_X$, and $\|\,|\xi'|v_-\|_X$ can be estimated as before by the right-hand side of (3.3). Inequality (3.3) now follows from the above considerations and (2.17).

Having established (3.2)–(3.3) and still with $f\in C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$ we proceed to show that, for $x_1\geqslant a>0$,

$$\int_{x_1}^{\infty}(1+t)^{-2\sigma}\Big(\|v_+(t)\|_X^2+\Big\|\frac{dv_+}{dr}\Big\|_X^2\Big)\,dt$$

To prove (3.4), we use that

$$\Big\|\frac{dv_+}{dr}-iv_+\Big\|_X^2=\Big\|\frac{dv_+}{dr}\Big\|_X^2+\|v_+\|_X^2-2\,\mathrm{Im}\Big(\frac{dv_+}{dr},v_+\Big)_X. \tag{3.5}$$

From Eq. (2.11) we have

$$\mathrm{Im}\int_0^r\Big(-\frac{d^2v_+}{dr^2},v_+\Big)_X\,dr+\mathrm{Im}\int_0^r[(F_+v_+,v_+)_X+(P_+v_+,v_+)_X-\|v_+\|_X^2]\,dr=\mathrm{Im}\int_0^r(g_+,v_+)\,dr,$$

where $\int_0^r$ means integration along $C_+(x_1,z)$, as determined by (2.10).

Now,

$$\int_0^r-\Big(\frac{d^2v_+}{dr^2},v_+\Big)_X\,dr=\int_0^{x_1}\Big(-\frac{d}{dx_1}\frac{d}{dr}v_+,v_+\Big)_X\,dx_1$$
$$=\int_0^{x_1}Q_+^{-1}\Big\|\frac{dv_+}{dx_1}\Big\|_X^2\,dx_1-\Big(\frac{dv_+}{dr},v_+\Big)_X+\Big(\frac{dv_+(o)}{dr},v_+(0)\Big)_X;$$

hence

$$\mathrm{Im}\Big(\frac{dv_+}{dr},v_+\Big)_X=\mathrm{Im}\int_0^{x_1}\Big[Q_+^{-1}\Big\|\frac{dv_+}{dx_1}\Big\|_X^2+Q_+^{-1}(|\xi'|^2v_+,v_+)_X-Q_+\|v_+\|_X^2\Big]\,dx_1$$
$$+\mathrm{Im}\int_0^{x_1}[(P_+Q_+v_+,v_+)_X-(g_+Q_+,v_+)_X]\,dx_1+\mathrm{Im}\Big(\frac{dv_+(0)}{dr},v_+(0)\Big)=I_1+I_2+I_3.$$

$\mathrm{Im}\,Q_+\geqslant 0$ implies $I_1\leqslant 0$.

Also (2.14), (2.17), and the trace estimates above imply

$$|I_2|+|I_3|\leqslant C(\|v_+\|_{-\sigma}^2+\|g_+\|_{\sigma,\omega^2}^2).$$

From (3.5) we now have

$$\Big\|\frac{dv_+}{dr}\Big\|_X^2+\|v_+\|_X^2\leqslant\Big\|\frac{dv_+}{dr}-iv_+\Big\|_X^2+C(\|v_+\|_{-\sigma}^2+\|g_+\|_{\sigma,\omega^2}^2). \tag{3.5}$$

Multiplying by $(1+t)^{-2\sigma}$ and integrating over (x_1,∞) yields (3.4), when (3.2) is noted.

We can prove now (3.1) for $f\in C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$. In view of (3.2)–(3.3) we only have to prove that

$$\|v_+\|_{-\sigma}+\Big|\int_0^a\int_{R^{n-1}}|v_-(x_1,\xi')|^2\,dx_1\,d\xi'\Big|^{1/2}\leqslant C(\|g_+\|_{\sigma,\omega^2}+\|g_-\|_{\tau,\omega}). \tag{3.6}$$

Suppose (3.6) is false. Let $\{f_k\}\subseteq C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$, $\{z_k\}\subseteq\tilde\Omega^+(I)$ be sequences such that the corresponding functions $\{v_k,g_k\}$ satisfy

$$\|g_{k+}\|_{\sigma,\omega^2}+\|g_{k-}\|_{\tau,\omega}\leqslant 1/k,$$
$$\|v_{k+}\|_{-\sigma}+\Big|\int_0^a\|v_{k-}\|_X^2\,dx_1\Big|^{1/2}=1.$$

Without loss of generality we assume $z_k\to z$. Noting (3.4), we have

$$\Big(\int_0^b\|v_{k+}\|_X^2\,dx_1\Big)^{1/2}+\Big(\int_a^0\|v_{k-}\|_X^2\,dx_1\Big)^{1/2}\geqslant\frac{1}{2},$$

where $b>0$ is independent of k.

Let $\chi(B)$ be the characteristic function of B. It follows from (3.2)–(3.3) that

$$\|\chi(|\xi'|\geqslant\lambda)v_-\|_{\tau,\omega^{-1}}+\|\chi(|\xi'|\geqslant\lambda)v_+\|_{\sigma-1,\omega^{-2}}\leqslant 2C/\lambda.$$

Hence, the last two estimates imply that for some fixed positive b,λ,

$$\Big[\int_0^b\|\chi(|\xi'|\leqslant\lambda)v_{k+}\|_X^2\,dx_1\Big]^{1/2}+\Big[\int_a^0\|\chi(|\xi'|\leqslant\lambda)v_{k-}\|_X^2\,dx_1\Big]^{1/2}\geqslant\frac{1}{4}.$$

However, in the Appendix we prove that for every bounded set $D\subseteq R_{\xi'}^{n-1}$ and every pair of bounded intervals $J_\pm\subseteq R_\pm$ we have

$$\int_{J_\pm}\|\chi(\xi'\in D)v_\pm\|_X^2\,dx_1\leqslant C(\|g_+\|_{\sigma,\omega^2}^2+\|g_-\|_{\tau,\omega}^2), \tag{3.7}$$

where C depends only on $J_\pm,D$. With $v=v_k$, $g=g_k$, (3.7) is in obvious contradiction to the previous inequality.

It now follows from the definition (2.3) of the norm $\|\ \|_{\tau,\sigma}$, definition (2.13) of $g_\pm$, and relation (2.17) that the right-hand side of (3.1) is less than $C\|f\|_{\tau,\sigma}$. If $\{f_j\}\subseteq C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$ is an approximating sequence to f in $\Psi_{\tau,\sigma}$ and $u_j=\mathrm{rad}(f_j,z)$, it follows from standard elliptic estimates that $\{u_j\}$ is bounded in $H^2(R^n)_{\mathrm{loc}}$. Rellich's compactness theorem guarantees the existence of a subsequence, say, $\{u_l\}$, which converges to a limit u in $H^1(R^n)_{\mathrm{loc}}$. u is a weak solution of (2.8), and, applying (3.1) to (u_l,f_l), we see that it satisfies the radiation condition. Hence $u=\mathrm{rad}(f,z)$ and the uniqueness of the radiative function implies that in fact $u_j\to u$ in $H^1(R^n)_{\mathrm{loc}}$. This proves the existence of $\mathrm{rad}(f,z)$ and the validity of (3.1) for $f\in\Psi_{\tau,\sigma}$ and concludes the proof of the lemma. Q.E.D.

Proof of Theorem 2.1: Let $f\in\Psi_{\tau,\sigma}$ and denote by $u=R_0^+(z)f$ the function rad(f,z) obtained in Lemma 3.1. Taking into account only the first four terms on the left-hand side of (3.1) and noting (2.4), (2.13), (2.15), (2.18), and the Parseval relation, we obtain

$$|||u|||_{\tau,-\sigma}\leqslant C\|f\|_{\tau,\sigma}.$$

Obviously, $u\in H^2(R^n)_{\rm loc}$ and $H_0u=zu+f$, which yields (2.5) since $\|u\|_{\tau,-\sigma}\leqslant|||u|||_{\tau,-\sigma}$.

It follows from the construction of rad(f,z) that if $\operatorname{Im} z>0$ and $f\in C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$, then

$$(H_0-z)^{-1}f=R_0(z)f=R_0^+(z)f. \tag{3.8}$$

This extends to any $f\in\Psi_{\tau,\sigma}\subseteq L^2(R^n)$ by continuity.

Finally, we prove the stated continuity of the map $z\to R_0^+(z)$ on $\widetilde{\Omega}^+(I)$. Due to $H_0u=zu+f$ it suffices to prove with $\Phi_{2,\tau',-\sigma}$ replaced by $\Phi_{\tau',-\sigma}$.

Thus, if $z_k\to z\in\widetilde{\Omega}^+(I)$ and $f\in C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$, we need to show that, for every $\epsilon>0$,

$$|||[R_0^+(z)-R_0^+(z_k)]f|||_{\tau',-\sigma}<\epsilon\|f\|_{\tau,\sigma},\quad k>K(\epsilon). \tag{3.9}$$

Observe that by (3.4), (3.1), and $\tau'<\tau$ we have

$$|||\chi(|x_1|\geqslant b)[R_0^+(z)-R_0^+(z_k)]f|||_{\tau',-\sigma}<\tfrac{1}{3}\epsilon\|f\|_{\tau,\sigma}, \tag{3.10}$$

where $b>0$ does not depend on k.

Using again standard elliptic estimates for the slab $|x_1|\leqslant b$ we have, uniformly in $z\in\widetilde{\Omega}^+(I)$,

$$\begin{aligned}\int_{-b}^{b}&\left\|\,|\xi'|\left[u(x_1,\xi')+\frac{\partial u}{\partial x_1}(x_1,\xi')\right]\right\|_X^2 dx_1\\ &\leqslant C(\|f\|_{\tau,\sigma}^2+\|u\|_{\tau',-\sigma}^2)\\ &\leqslant C_1\|f\|_{\tau,\sigma}^2\quad[u=R_0^+(z)f].\end{aligned} \tag{3.11}$$

Indeed, u is a solution of $(H_0-z)u=f$ and the slab $|x_1|\leqslant b$ can be divided up into a disjoint union of finite rectangular boxes in each of which the L^2 norms of derivatives can be estimated in terms of the norms of u,f. The constants are all equal since H_0 is invariant under translation in x' directions.

From (3.11) we obtain

$$\begin{aligned}&|||\chi(|\xi'|\geqslant\lambda)\chi(|x_1|\leqslant b)[R_0^+(z)-R_0^+(z_k)]f|||_{\tau',-\sigma}\\ &\quad<\tfrac{1}{3}\epsilon\|f\|_{\tau,\sigma},\quad z\in\widetilde{\Omega}^+(I),\end{aligned} \tag{3.12}$$

where $\lambda>0$ is sufficiently large, independent of z,f.

In the Appendix we prove that if $D\subseteq R_{\xi'}^{n-1}$ is bounded and $J\subseteq R_{x_1}$ is a bounded interval, then, uniformly in $z\in\widetilde{\Omega}^+(I)$,

$$\begin{aligned}&|||\chi(\xi'\in D)\chi(x_1\in J)[R_0^+(z)-R_0^+(z_k)]f|||_{\tau',-\sigma}\\ &\quad<\tfrac{1}{3}\epsilon\|f\|_{\tau,\sigma}\end{aligned} \tag{3.13}$$

if $k>K(\epsilon)$.

Combining (3.10), (3.12), and (3.13), we obtain (3.9). Q.E.D.

In the next section we shall need the following simple lemma, which asserts that the solution u of $(H_0-z)u=f$, $\operatorname{Im} z>0$, decays at $-\infty$ "faster" than f. Recall that by assumption $(VL\,2)$ (i) $\omega(t)\geqslant|t|^\epsilon$ as $|t|\to\infty, 0<\epsilon<\frac{1}{2}\delta_1$. Also, we let $\varphi(t)\in C^\infty(R)$, $\varphi(t)=1$ [resp. $\varphi(t)=0$] for $t\leqslant 0$ [resp. $t\geqslant 1$]. Note that the norm in $L_\tau^2(R_-,X)$ is defined in (2.9).

Lemma 3.2: Let $f\in L^2(R^n)\cap L_\tau^2(R_-,X)$, some $\tau\in R$, and for $z\in\Omega^+(I)$ let $u=R_0(z)f$. Then $\varphi u\in\Phi_{\tau+\epsilon,0}$ and

$$|||\varphi u|||_{\tau+\epsilon,0}\leqslant C(\|f_-\|_\tau+(\operatorname{Im} z)^{-1}\|f\|_{L^2(R^n)}). \tag{3.14}$$

If $f\in L^2(R^n)\cap\Psi_{\tau,0}$, then $\varphi u\in\Phi_{\tau,0}$.

Proof: Assume first that $f\in C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$. Observe that (3.3) was proved for every $\tau\in R$. In particular

$$\begin{aligned}&\left\|\frac{\partial u_-}{\partial x_1}\right\|_{\tau+\epsilon}+\|u_-\|_{\tau+\epsilon,\omega^2}\\ &\leqslant C\left(\left\|\frac{d}{dr}v_{-\tau+\epsilon,\omega}\right\|+\|v_-\|_{\tau+\epsilon,\omega}\right)\leqslant C\left(\|g_-\|_{\tau+\epsilon,\omega}\right.\\ &\quad\left.+\left|\int_0^a\int_{R^{n-1}}|v_-(x_1,\xi')|^2\,dx_1\,d\xi'\right|^{1/2}\right)\\ &\leqslant C(\|f_-\|_\tau+(\operatorname{Im} z)^{-1}\|f\|_{L^2(R^n)}),\end{aligned}$$

where the first term was estimated in view of (2.13) and the second by $\|R_0(z)\|_{L^2(R^n)}\leqslant(\operatorname{Im} z)^{-1}$.

Also, since the support of φ in R_+ is bounded, we get from standard elliptic estimates

$$\begin{aligned}&\|\varphi u_+\|_{0,\omega}+\left\|\varphi\frac{\partial u_+}{\partial x_1}\right\|_{0,\omega^{-1}}\leqslant C(\|u\|_{L^2(R^n)}+\|f\|_{L^2(R^n)})\\ &\leqslant C(\operatorname{Im} z)^{-1}\|f\|_{L^2(R^n)}.\end{aligned}$$

Combining the two inequalities (3.14) follows for $f\in C_0^\infty(R_{x_1}\times R_{\xi'}^{n-1})$. A general f can be approximated in $\|\ \|_{L^2(R^n)}+\|\ \|_{L_\tau^2(R_-,X)}$ by such functions (truncate and mollify) $\{f_k\}$. Applying (3.14) to $u_k=R_0(z)f_k$ and using that $u_k\to u$ in $H^1(R^n)_{\rm loc}$, (3.14) follows.

The last statement is just (3.3) as observed above for $\tau=0$. Q.E.D.

IV. THE LIMITING ABSORPTION PRINCIPLE FOR H

Let $H=H_0+V$, where $H_0=\Delta+V_0(x_1)$ and $V(x)=V(x_1,\ldots,x_n)$ is a real function depending on all coordinates.

For $r,s\in R$ let $\rho_{r,s}(x_1)\neq 0$ be a smooth function, such that

$$\rho_{r,s}(x_1)=\begin{cases}(1+x_1^2)^{r/2}, & x_1\leqslant-1,\\ (1+x_1^2)^{s/2}, & x_1\geqslant 1.\end{cases}$$

Our assumptions on the multiplication operator V are the following:

$(V1)$ V is relatively compact with respect to H_0, i.e., $V(H_0+i)^{-1}$ is compact in $L^2(R^n)$.

$(V2)$ For some $\beta>0$, the multiplication operator $\rho_{\beta,1+\beta}V$ is compact as a map from $\Phi_{2,0,0}$ into $\Psi_{0,0}$ and $\rho_{0,(1+\beta)/2}V$ maps $D(H_0)$ into $\Psi_{0,0}$.

Remark 4.1: It is an easy matter to check that multiplication by $\rho_{r,s}(x_1)$ is continuous (in fact, bijective) from $\Phi_{2,\tau,\sigma}\to\Phi_{2,\tau-r,\sigma-s}$ and $\Psi_{\tau,\sigma}\to\Psi_{\tau-r,\sigma-s}$. [This is why du/dx_1 was thrown into the definitions (2.3)–(2.4) to avoid interpolation estimates.] Thus, assumption $(V2)$ implies that for every τ,σ, multiplication by V is compact from $\Phi_{2,\tau,\sigma}\to\Psi_{\tau+\beta,\sigma+1+\beta}$. Note that $(V1)$ implies that

$H/C_0^\infty(R^n)$ is essentially self-adjoint (Kato[21]). We retain the notation H for its unique self-adjoint extension in $L^2(R^n)$.

A sufficient condition for $(V1)$–$(V2)$ is provided by the following (see also Herbst,[22] Yajima[23]):
$V(x) = V_1(x) + V_2(x)$, where $V_1(x)$ is compactly supported and compact from $H^2(R^n)$ into $L^2(R^n)$. $V_2(x)$ is bounded and satisfies

$$|V_2(x)| \leqslant CM(x')\{\chi(x_1 \leqslant 0)(1 + |V_L(x_1)|)(1 + x_1^2)^{-\beta'/2} + \chi(x_1 \geqslant 0)[1 + |V_L(x_1)|]^{1/2}(1 + x_1^2)^{-(1+\beta')/2}\},$$
$$\beta' > \beta, \quad M(x') \to 0 \text{ as } |x'| \to \infty. \tag{4.1}$$

For the Stark case ($V_L(x_1) = -x_1$) this is in accordance with the assumption of Yajima.[24]

By a well-known theorem of Weyl[25] the relative compactness of V implies that the essential spectrum of H is equal to that of H_0, namely, the entire real line. However, in general we cannot expect in this case the absence of eigenvalues.

Let $I, \widetilde{\Omega}^+(I)$ be as in (2.1). By Theorem 2.1 and Remark 4.1

$$T(z) := R_0^+(z)V \in B(\Phi_{2,\tau,-\sigma}, \Phi_{2,\tau',\sigma})$$

is compact for $\tau' < \tau + \beta$, $1 < 2\sigma < 1 + \beta$, and the operator-valued map $z \to T(z)$ is continuous in $\widetilde{\Omega}^+(I)$ (the uniform topology is used for the operator space). For $\operatorname{Im} z > 0$ let $R(z) = (H - z)^{-1}$. The resolvent equation implies

$$R(z) = [I + T(z)]^{-1}R_0(z). \tag{4.2}$$

Observe that $[I + T(z)]^{-1}$ exists in $B(\Phi_{2,\tau,-\sigma})$, $1 < 2\sigma < 1 + \beta$, $\tau > 1$ by the Fredholm–Riesz theory, since

$$C_0^\infty(R^n) \subseteq \text{Range } R_0(z) \subseteq \text{Range}[I + T(z)].$$

Analogously to $R_0(z)$, we would like to extend $R(z)$ continuously to $\widetilde{\Omega}^+(I)$, in a suitable topology. The following theorem states that this is possible, if we avoid finitely many eigenvalues in I.

Theorem 4.2: Let $\mathscr{N}$ be the set of eigenvalues of H in I. Then $\mathscr{N}$ is at most finite (including multiplicity). If u is an eigenfunction associated with $\lambda \in \mathscr{N}$, then $u \in \Phi_{2,\eta,\theta}$ for every real η, θ.

The operator-valued function $z \to R(z) \in B(\Psi_{\tau,\sigma}, \Phi_{2,\tau',-\sigma})$ can be extended continuously to $\widetilde{\Omega}^+(I) \setminus \mathscr{N}$, if $\sigma > \frac{1}{2}$ and $1 < \tau' < \tau$. The operator space is equipped with the uniform topology.

Proof: By (4.2) and Theorem 2.1 it suffices to study the invertibility of $I + T(\lambda)$ in $B(\Phi_{2,\tau,-\sigma})$. Observe that proving the statement for some $\sigma_0 > \frac{1}{2}$ implies its validity for all $\sigma > \sigma_0$. Thus we shall assume that σ satisfies (2.6) and $2\sigma < 1 + \beta$.

Suppose that $[I + T(\lambda)]^{-1}$ exists for some $\lambda \in I$. (4.2) implies that $R(z)$ can be extended as stated to a neighborhood of λ in $\widetilde{\Omega}^+(I)$. In particular $(R(z)f, f)$ is continuous near λ for $f \in C_0^\infty(R^n)$. Using the well-known formula[26]

$$(E(\lambda_1, \lambda_2)f, f) = \lim_{\epsilon \to 0+} \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} \operatorname{Im}(R(\mu + i\epsilon)f, f)\, d\mu \tag{4.3}$$

[$E(\cdot)$ being the spectral family associated with H], it follows that λ lies in an interval which is entirely contained in the absolutely continuous spectrum of H. Hence

$$\mathscr{N} \subseteq \{\lambda \mid \lambda \in I,\ I + T(\lambda) \text{ not invertible in } B(\Phi_{2,\tau,-\sigma})\}. \tag{4.4}$$

Since $T(\lambda)$ is compact, $I + T(\lambda)$ is noninvertible if and only if there exists a function $u \in \Phi_{2,\tau,-\sigma}$ such that

$$u = -T(\lambda)u = -R_0^+(\lambda)Vu. \tag{4.5}$$

We state now a lemma, the proof of which is postponed to the next section.

Lemma 4.3: Let $u \in \Phi_{2,\tau,-\sigma}$ satisfy (4.5) for some $\tau > 1$ and σ satisfying (2.6) and $2\sigma < 1 + \beta$. Then $u \in \Phi_{2,\eta,\theta}$ for every $\eta, \theta \in R$. In particular, u is an L^2 eigenfunction of H, $Hu = \lambda u$. Furthermore, there exists a constant C, depending only on η, θ, I, and the operator, such that

$$|||u|||_{2,\eta,\theta} \leqslant C \|u\|_{L^2(R^n)}. \tag{4.6}$$

Assuming this lemma, we can conclude the proof of the theorem, applying an argument similar to that of Agmon.[27]

Let $\{u_k\}$ be a sequence of solutions of (4.5) for some $\lambda_k \in I$, $\|u_k\|_{L^2(R^n)} = 1$. By (4.6), $\{u_k\}$ is bounded in $\Phi_{2,\eta,\theta}$, $\eta, \theta > 1$; hence $\{Vu_k\}$ is precompact in $\Psi_{\eta,\theta}$, and, as in the proof of (3.9), $\{u_k = -R_0^+(\lambda_k)Vu_k\}$ is precompact in $L^2(R^n)$. Since they are all eigenfunctions of a self-adjoint operator, the set must be finite. By (4.4) $\mathscr{N}$ is finite.

Finally, let $u \in L^2(R^n)$ be an eigenfunction of H associated with $\lambda \in \mathscr{N}$. If $z = \lambda + i\delta$, $\delta > 0$, we have

$$u + R_0(z)Vu = -i\delta R_0(z)u. \tag{4.7}$$

We want to show that $Vu \in \Psi_{\tau,\sigma}$, $\tau > 1$, $\sigma > \frac{1}{2}$. Observe that for some $r, s \in R$ we have $u \in \Phi_{2,r,s}$. Indeed, $V_0(x_1) = O(x_1^2)$ so that $V_0 u \in \Psi_{-2,-2}$ and $-\Delta u = H_0 u - V_0 u \in \Psi_{-2,-2}$. Hence $Vu \in \Psi_{r+\beta, s+1+\beta}$. Applying Lemma 3.2 we have from (4.7) that $u \in \Phi_{r+\eta,s}$ and $H_0 u = -Vu + \lambda u \in \Psi_{r+\eta,s}$, where $\eta = \min(\epsilon, \beta)$.

Thus

$$u \in \Phi_{2,r+\eta,s}.$$

Repeating this procedure we obtain $u \in \Phi_{2,\tau,s}$ for $\tau > 1$. Since $u \in D(H) = D(H_0)$, assumption $(V2)$ implies $Vu \in \Psi_{0,\sigma}$, $\sigma > \frac{1}{2}$; hence $Vu \in \Psi_{\tau,\sigma}$. The limiting absorption principle can therefore be applied to $R_0(z)Vu$.

As $\delta \to 0^+$, the left-hand side of (4.7) converges to $\varphi = u + R_0^+(\lambda)Vu \in \Phi_{2,\tau,\theta}[\theta = \min(s, -\sigma)]$. The right-hand side is bounded in $L^2(R^n)$ ($\|R_0(z)\|_{L^2(R^n)} \leqslant 1/\delta$) and thus has a weakly convergent subsequence which necessarily converges to φ. Hence $\varphi \in L^2(R^n)$. But $(H_0 - \lambda)\varphi = (H - \lambda)u = 0$, which implies $\varphi = 0$ (H_0 has no eigenvalues), i.e., $u = -R_0^+(\lambda)Vu \in \Phi_{2,\tau,-\sigma}$. Lemma 4.3 can now be applied to conclude that u decays of any order and verifies (4.6). This concludes the proof of the theorem.

Q.E.D.

Note that by (4.2) the functions in the range of $R^+(z)$, $z \in \widetilde{\Omega}^+(I)$, are radiative functions in the sense of Definition 2.2.

Corollary 4.4: The essential spectrum of H is the entire real axis. It is absolutely continuous apart from a possible discrete sequence of eigenvalues of finite multiplicity. The corresponding eigenfunctions are rapidly decaying as $|x_1| \to \infty$.

V. RAPID DECAY OF EIGENFUNCTIONS

In this section we prove Lemma 4.3. As observed by Herbst,[28] the proof of this "division" lemma constitutes the hardest part in the proof of the limiting absorption principle. Our method is similar to the one used in Ben-Artzi[29] and exploits a generalized one-dimensional eigenfunction expansion associated with $H_{0,\xi'}$, rather than an expansion corresponding to the complete operator H_0. This enables one to treat more general cases of the basic operator on the underlying domain, where an eigenfunction expansion is not readily available (see Sec. VII below).

We shall find it convenient to use here, along with the spaces $\Psi_{\tau,\sigma}, \Phi_{\tau,\sigma}, \Phi_{2,\tau,\sigma}$, also their one-dimensional analogs, denoted, respectively, by $\tilde{\Psi}_{\tau,\sigma}, \tilde{\Phi}_{\tau,\sigma}$, and $\tilde{\Phi}_{2,\tau,\sigma}$. Their norms are obtained from (2.3)–(2.4) by omitting x' and $\int_{R^{n-1}}$.

Proof of Lemma 4.3: Set $F = -Vu$. By assumption $(V2)$ we have $F \in \Psi_{\tau+\beta,-\sigma+1+\beta} \subseteq \Psi_{\tau+\beta,\sigma}$. Hence Eq. (4.5) and Theorem 2.1 imply $u \in \Phi_{2,\tau+\beta,-\sigma}$. Continuing in this fashion we see that $u \in \Phi_{2,\eta,-\sigma}$ for every $\eta \in R$. Without loss of generality we therefore assume $u \in \Phi_{2,\tau,-\sigma}$, $\tau > 1$. It remains to show that $-\sigma$ can be replaced by any $\theta \in R$.

Throughout this section we use the representation (2.8) for u,F. Let $J \subseteq R^{n-1}_{\xi'}$. For a function $u(x_1,\xi')$ we denote $u_J = \chi(J)u$. The proof of the lemma follows from the following two propositions.

Proposition 5.1: If $u \in \Phi_{2,\tau,\theta}$ satisfies (4.5) with $\theta \geqslant -\sigma$, then $u_J \in \Phi_{2,\tau,\theta+\gamma}$ for every bounded $J \subseteq R^{n-1}_{\xi'}$, where $0<\gamma<\beta$ is fixed independently of θ, J.

Proposition 5.2: If $u \in \Phi_{2,\tau,\theta}$ satisfies (4.5) with $\theta \geqslant -\sigma$, and if $u_J \in \Phi_{2,\tau,\theta+\gamma_0}$ for every bounded J, then $u \in \Phi_{2,\tau,\theta+\gamma}$, for some fixed $0<\gamma \leqslant \gamma_0$, and

$$|||u|||_{2,\tau,\theta+\gamma} \leqslant C\,|||u|||_{2,\tau,\theta}, \tag{5.1}$$

where γ does not depend on θ.

Proof of Proposition 5.1: Let $H_{0,\xi'}$ be the self-adjoint operator in $L^2(R)$ defined by (2.7). We recall some facts from Ben-Artzi[30] concerning the spectral representation of $H_{0,\xi'}$.

Let $R^{+}_{0,\xi'}(\lambda) = \lim_{\delta\to 0^+}(H_{0,\xi'} - \lambda - i\delta)^{-1}$, where the limit exists in $B(\tilde{\Psi}_{\tau+\beta,\sigma}, \tilde{\Phi}_{2,\tau,-\sigma})$, $\lambda \in I$. As already mentioned in the proof of Lemma 2.4, this operator is represented by a kernel $K(s,t,\xi',\lambda)$, which is continuous on $R \times R \times J \times I$, and given explicitly by

$$K(s,t,\xi',\lambda) = \begin{cases} w(\xi',\lambda)\varphi_1(s,\xi',\lambda)\psi_1(t,\xi',\lambda), & s>t, \\ w(\xi',\lambda)\psi_1(s,\xi',\lambda)\varphi_1(t,\xi',\lambda), & s \leqslant t, \end{cases} \tag{5.2}$$

where φ_1, ψ_1 are given by (2.24), (2.26).

$H_{0,\xi'}$ is absolutely continuous and of multiplicity one over R, and can be diagonalized by means of an eigenfunction expansion. Specifically, let $E_{0,\xi'}(\cdot)$ be the spectral family associated with $H_{0,\xi'}$. There exists a unitary map of $E_{0,\xi'}(I)L^2(R)$ onto $L^2(I)$ such that, for $g \in L^2(R)$,

$$UE_{0,\xi'}(I)g(\mu) := \hat{g}_{\xi'}(\mu) = \lim_{T\to\infty}\int_{-T}^{T} g(s)\nu(s,\xi',\mu)\; ds, \; \mu \in I, \tag{5.3}$$

$$(UH_{0,\xi'}E_{0,\xi'}(I)U^{-1}\hat{g}_{\xi'})(\mu) = \mu\hat{g}_{\xi'}(\mu),$$

where

$$\nu(s,\xi',\mu) = k(\mu)\psi_1(s,\xi',\mu), \quad \mu \in I, \tag{5.4}$$

and $k(\mu) \neq 0$ is continuous in I.

Using the asymptotic behavior of ψ_1 (which near $+\infty$ is a linear combination of $\varphi_1, \bar{\varphi}_1$) we see that for $g \in \tilde{\Psi}_{\tau,\sigma}$, $\tau>1$, $\sigma > \frac{1}{2}$, $\hat{g}_{\xi'}(\mu)$ is continuous in I [the integral in (5.3) is absolutely convergent and continuous in μ by dominated convergence].

Using the formula (4.3), we have, if $g \in \tilde{\Psi}_{\tau,\sigma}$,

$$\begin{aligned}\frac{d}{d\mu}(E_{0,\xi'}(-\infty,\mu)g,g) &= \frac{1}{\pi}\,\mathrm{Im}(R^{+}_{0,\xi'}(\mu)g,g) \\ &= |\hat{g}_{\xi'}(\mu)|^2.\end{aligned} \tag{5.5}$$

Note that the middle term is well defined since $g \in \tilde{\Psi}_{\tau,\sigma} \Rightarrow R^{+}_{0,\xi'}(\lambda)g \in \tilde{\Phi}_{\tau,-\sigma}$.

An application of the representation (2.7) and Eq. (4.5) gives

$$\begin{aligned}\mathrm{Im}\int_{R^{n-1}} &(R^{+}_{0,\xi'}(\lambda)F(x_1,\xi'),F(x_1,\xi'))\,d\xi' \\ &= \mathrm{Im}(R^{+}_{0}(\lambda)F,F) \\ &= \mathrm{Im}(u,{}^{-}Vu) = 0\end{aligned}$$

(the scalar product is well defined since $u \in \Phi_{2,\tau,\theta} \Rightarrow Vu \in \Psi_{\tau,\theta+1+\beta}$).

Noting (5.5), we conclude

$$\mathrm{Im}(R^{+}_{0,\xi'}(\lambda)F(x_1,\xi'),F(x_1,\xi')) = |\hat{g}_{\xi'}(\lambda)|^2 = 0, \tag{5.6}$$

which holds for a.e. $\xi' \in J$. (5.3), (5.4), and (5.6) imply that $\int_R F(s,\xi')\psi_1(s,\xi',\lambda)\,ds = 0$ so that, from (5.2),

$$\begin{aligned}u(x_1,\xi') = w(\xi',\lambda)\Big(&\psi_1(x_1,\xi',\lambda)\int_{x_1}^{\infty}F(s,\xi')\varphi_1(s,\xi',\lambda)\,ds \\ &-\varphi_1(x_1,\xi',\lambda)\int_{x_1}^{\infty}F(s,\xi')\psi_1(s,\xi',\lambda)\,ds\Big)\end{aligned} \tag{5.7}$$

and, by differentiation,

$$\begin{aligned}\frac{\partial u}{\partial x_1}(x_1,\xi') = w(\xi',\lambda)\Big(&\frac{d\psi_1}{dx_1}\int_{x_1}^{\infty}F(s,\xi')\varphi_1(s,\xi',\lambda)\,ds \\ &-\frac{d\varphi_1}{dx_1}\int_{x_1}^{\infty}F(s,\xi')\psi_1(s,\xi',\lambda)\,ds\Big).\end{aligned} \tag{5.8}$$

It follows from our assumptions that $F_J \in \Psi_{\tau,\theta+1+\beta}$ and the asymptotic estimates (2.24), (2.26) hold uniformly in $\xi' \in J$, so that, with $\omega(x_1)$ defined by (2.15),

$$\begin{aligned}&\int_0^{\infty}\int_J (1+x_1)^{2(\theta+\gamma)}\omega(x_1)|u_J(x_1,\xi')|^2\,dx_1\,d\xi' \\ &\leqslant C\int_J\int_0^{\infty}(1+x_1)^{2(\theta+\gamma)}\Big(\int_{x_1}^{\infty}\omega(s)^{-1/2}|F(s,\xi')|\,ds\Big)^2\,dx_1\,d\xi' \\ &\leqslant C\int_0^{\infty}(1+x_1)^{2(\theta+\gamma)}\Big(\int_{x_1}^{\infty}\int_J\omega(s)^{-1}(1+s)^{2(\theta+1+\beta)} \\ &\qquad |F(s,\xi')|^2\,ds\,d\xi'\int_{x_1}^{\infty}\times(1+s)^{-2(\theta+1+\beta)}\,ds\Big)\,dx_1 \\ &\leqslant C\,\|F_J\|^2_{\tau,\theta+1+\beta} \quad \text{since } \gamma<\beta.\end{aligned}$$

A similar computation using (5.8) shows that $u_J \in \Phi_{\tau,\theta+\gamma}$. Also, by (4.5), $H_0u_J = \lambda u_J + F_J \in \Psi_{\tau,\theta+\gamma}$; hence $u_J \in \Phi_{2,\tau,\theta+\gamma}$.

This concludes the proof of Proposition 5.1. Note that in the last inequality C depends on J, so that we cannot deduce from it that $u\in\Phi_{2,\tau,\theta+\gamma}$. This is what we show next.

Proof of Proposition 5.2: Since $H_0u = \lambda u + F$ and $F\in\Psi_{\tau,\theta+1+\beta}$, it clearly suffices to prove that $u\in\Phi_{\tau,\theta+\gamma}$, with $|||\ |||_{2,\tau,\theta+\gamma}$ replaced by $|||\ |||_{\tau,\theta+\gamma}$ on the left-hand side of (5.1).

Note that, as has already been observed in the beginning of this section, τ can be taken arbitrarily large, so that all integrations over R_- are justified. In such a case, assume that in (2.8) z is real, take the scalar product (in X) with u, and integrate over $(-\infty,x_1)$ to obtain

$$\operatorname{Im}(u,du/dx_1)_x = \operatorname{Im}(f,u)_X \quad (\operatorname{Im} z = 0). \tag{5.9}$$

In particular, in our case $f = F = -Vu$, and we have the important relation

$$\operatorname{Im}\left(u, \frac{du}{dx_1}\right)_X = 0. \tag{5.10}$$

Transforming the equation $(H_0 - \lambda)u = F$ as in (2.11), let (v,g) be the pair corresponding to (u,F). When relation (5.10) is incorporated into inequality (3.1) (note that $Q_\pm$ are real), we get

$$\left|\left|\frac{dv_+}{dr}\right|\right|_{\sigma-1} + \|v_+\|_{\sigma-1} = \left|\left|\frac{dv_+}{dr} - iv_+\right|\right|_{\sigma-1} \leqslant C\,\|F\|_{\tau,\sigma} \leqslant C\,\|F\|_{\tau,\theta+1+\beta},$$

which implies, by (2.15), (2.17) that $u\in\Phi_{\tau,\sigma-1} \Rightarrow u\in\Phi_{2,\tau,\sigma-1}$. Furthermore, the last inequality implies, in fact

$$|||u|||_{\tau,\sigma-1} \leqslant C\,|||u|||_{2,\tau,-\sigma}. \tag{5.11}$$

We can now start our proof with $\theta = \sigma - 1 + s$, $s\geqslant 0$. Let $w = (1+x_1)^s u$, $x_1 \geqslant 0$. In what follows we work only on R_+. Thus, the norms defined in (2.9) will be used. For simplicity, we suppress the index $+$.

Clearly, w satisfies the following equation:

$$\left(-\frac{d^2}{dx_1^2} + V_0(x_1) + |\xi'|^2 - \lambda\right)w = G, \tag{5.12}$$

$$G = [(1+x_1)^s F - s(s-1)(1+x_1)^{s-2}u] + \left(-2s(1+x_1)^{s-1}\frac{d}{dx_1}u\right) = G_1 + G_2.$$

Since $u_J\in\Phi_{2,\tau,\theta+\gamma_0}$, we have

$$w_J\in L^2_{\sigma-1+\gamma_0,\omega}(R_+,X),$$

$$\frac{d}{dx_1}w_J\in L^2_{\sigma-1+\gamma_0,\omega^{-1}}(R_+,X) \tag{5.13}$$

and also

$$G_1\in L^2_{\sigma+\beta,\omega^{-1}}(R_+,X), \quad G_2\in L^2_{\sigma,\omega^{-1}}(R_+,X). \tag{5.14}$$

Clearly we are done if the following claim is proved:

Claim: $w\in L^2_{\sigma-1+\gamma,\omega}(R_+,X)$, $dw/dx_1 \in L^2_{\sigma-1+\gamma,\omega^{-1}}(R_+,X)$, and

$$\|w\|_{\sigma-1+\gamma,\omega} + \left|\left|\frac{dw}{dx_1}\right|\right|_{\sigma-1+\gamma,\omega^{-1}} \leqslant C\,|||u|||_{2,\tau,\theta}, \tag{5.15}$$

where $0<\gamma\leqslant\gamma_0$ is independent of s (i.e., θ).

Now, to prove the claim, we multiply Eq. (5.12) by $\chi(J)$ and transform as in (2.11) to obtain, with $y = Q^{1/2}w$, $y' = dy/dr$,

$$-\frac{d}{dx_1}(y_J' - iy_J) - iQ(y_J' - iy_J) + Q^{-1}|\xi'|^2 y_J + PQy_J = Q^{-1/2}G_J. \tag{5.16}$$

Let $t = \sigma - 1 + \gamma$, where $0<\gamma\leqslant\gamma_0$ is still to be selected.

It follows from (5.13) that

$$y_J, y_J' \in L^2_t(R_+,X). \tag{5.17}$$

Also, from (2.14), if $\gamma\leqslant\epsilon$,

$$PQy\in L^2_{\sigma+\epsilon}(R_+,X) \subseteq L^2_{1+t}(R_+,X). \tag{5.18}$$

We now take the scalar product (in X) of (5.16) with $(1+x_1)^{2t+1}(y_J' - iy_J)$, then take the real part and integrate over $[0,T]$ to get

$$\begin{aligned}
&\left(t+\frac{1}{2}\right)\int_0^T (1+x_1)^{2t}\|y_J' - iy_J\|_X^2\,dx_1 \\
&\quad + \frac{1}{2}\int_0^T Q^{-2}(1+x_1)^{2t+1}\frac{d}{dx_1}\|\,|\xi'|y_J\|_X^2\,dx_1 \\
&\leqslant (1+T)^{2t+1}\|(y_J' - iy_J)(T)\|_X^2 + \int_0^T (1+x_1)^{2t+1} \\
&\quad \times |(-PQy_J + Q^{-1/2}G_{1,J}, y_J' - iy_J)_X|\,dx_1 \qquad (5.19)\\
&\quad + \operatorname{Re}\int_0^T (1+x_1)^{2t+1}(Q^{-1/2}G_{2,J}, y_J' - iy_J)_X\,dx_1 \\
&= I_1(T) + I_2(T) + I_3(T).
\end{aligned}$$

We now impose our restriction on γ. Our assumptions on V_L imply for some $\mu>0$, $(d/dt)[(1+t)^\mu Q^{-2}]\leqslant 0$ if $t\geqslant t_0$. We assume now that $0<2t+1\leqslant\mu$, namely, $0<\gamma<\frac{1}{2}(1+\mu)-\sigma$. Note that by the remarks in the beginning of the proof of Theorem 4.2, we can take σ arbitrarily close to $\frac{1}{2}$, adding still another restriction to (2.6).

Integrating by parts in the second integral on the left-hand side of (5.19), and using (3.1) to estimate the integral over the finite interval $[0,t_0]$, we have

$$\int_0^T (1+x_1)^{2t}\|y_J' - iy_J\|_X^2\,dx_1 \leqslant C(I_1(t) + I_2(t) + I_3(t) + |||u|||_{2,\tau,\theta}).$$

By (5.17) we have

$$\liminf_{T\to\infty} I_1(T) = 0. \tag{5.20}$$

Noting (5.14), (5.18), we have

$$\|-PQy_J + Q^{-1/2}G_{1,J}\|_{1+t} \leqslant C(\|y\|_{\sigma-1} + |||u|||_{2,\tau,\theta}) \leqslant C\,|||u|||_{2,\tau,\theta}.$$

In order to estimate $Q^{-1/2}G_1$ we need that $1+t\leqslant\sigma+\beta$, i.e., $\gamma\leqslant\beta$. Then we note that the above estimate follows directly from the expression for G_1 since $\|(1+x_1)^s F\|_{\tau,1+t} \leqslant \|F\|_{\tau,\theta+1+\beta} \leqslant C\,|||u|||_{2,\tau,\theta}$. The constant C is, of course, independent of J. Hence, using the Cauchy–Schwartz inequality in $I_2(T)$,

$$|I_2(T)|\leqslant\frac{C}{\eta}\,|||u|||^2_{2,\tau,\theta}+\eta\int_0^T(1+x_1)^{2t}\|y_J'-iy_J\|^2_X\,dx_1, \tag{5.21}$$

where $\eta>0$ is arbitrarily small.

The estimate of $I_3(T)$ is considerably more subtle. The reason is that we cannot proceed as in the case of $I_2(T)$, since then we would need an estimate of $\|Q^{-1/2}G_{2,J}\|_{1+t}$, whereas by (5.14) we only have $Q^{-1/2}G_2\in L^2_\sigma(R_+,X)$ (note that $1+t>\sigma$).

Let us therefore estimate the expression

$$\begin{aligned}
\mathrm{Re}(1&+x_1)^{2t+1}(Q^{-1/2}G_{2,J},y_J'-iy_J)_X\\
&=-2s\Big[\mathrm{Re}(1+x_1)^{2t+s}\Big(Q^{-1/2}\frac{du_J}{dx_1},y_J'\Big)_X\\
&\quad-\mathrm{Im}(1+x_1)^{2t+s}\Big(Q^{-1/2}\frac{du_J}{dx_1},y_J\Big)_X\Big]\\
&=-2s[L_1^J(x_1)+L_2^J(x_1)],
\end{aligned}$$

$$\begin{aligned}
|L_2^J(x_1)|&=(1+x_1)^{2(t+s)}\Big|\mathrm{Im}\Big(\frac{du_J}{dx_1},u_J\Big)_X\Big|\\
&=(1+x_1)^{2(t+s)}|\mathrm{Im}(u_J,F_J)_X|\\
&\leqslant(1+x_1)^\theta\|u\|_X(1+x_1)^{\theta+2\gamma}\|F\|_X=h_1(x_1),
\end{aligned}$$

where obviously $h_1(x_1)\in L^1(R_+)$ if $2\gamma<1+\beta$. The passage to $\mathrm{Im}(u_J,F_J)_X$ was done as in (5.9) (but it does not vanish). Now,

$$\begin{aligned}
\frac{dy_J}{dr}&=\frac{1}{Q}\frac{d}{dx_1}[Q^{1/2}(1+x_1)^su_J]\\
&=p(x_1)u_J+Q^{-1/2}(1+x_1)^s\frac{du_J}{dx_1},
\end{aligned}$$

where, using the assumption on V_L,

$$|p(x_1)|\leqslant C(1+x_1)^{s-\delta}.$$

Hence

$$L_1^J(x_1)\geqslant\mathrm{Re}(1+x_1)^{2t+s}\Big(Q^{-1/2}\frac{du_J}{dx_1},p(x_1)u_J\Big)_X=M^J(x_1),$$

and, if $2\gamma<\delta$,

$$\begin{aligned}
|M^J(x_1)|&\leqslant C\omega(x_1)^{-1/2}(1+x_1)^\theta\Big\|\frac{du_J}{dx_1}\Big\|_X\\
&\quad\cdot(1+x_1)^\theta\|u\|_X=h_2(x_1),
\end{aligned}$$

where again $u\in\Phi_{\tau,\theta}\Rightarrow h_2(x_1)\in L^1(R_+)$.

Since $s\geqslant0$, we have

$$\begin{aligned}
I_3(T)&\leqslant2s\int_0^T[M^J(x_1)+|L_2^J(x_1)|]\,dx_1\\
&\leqslant2s\int_0^T[h_1(x_1)+h_2(x_1)]\,dx_1.
\end{aligned}\tag{5.22}$$

Substituting (5.20)–(5.22) into (5.19) and letting $T\to\infty$ along a suitable sequence, we get

$$\begin{aligned}
\int_0^\infty&(1+x_1)^{2t}\|y_J'-iy_J\|^2_X\,dx_1\\
&\leqslant C\Big(|||u|||^2_{2,\tau,\theta}+\int_0^\infty[M^J(x_1)+|L_2^J(x_1)|]\,dx_1\Big).
\end{aligned}$$

C is independent of J.

Take now a sequence of bounded sets $J_k\uparrow R^{n-1}_\xi$. By (5.10) we have, pointwise,

$$\lim_{k\to\infty}L_2^{J_k}(x_1)=\lim_{k\to\infty}M^{J_k}(x_1)=0$$

so that by dominated convergence the last inequality yields:

$$\int_0^\infty(1+x_1)^{2t}\|y'-iy\|^2_X\,dx_1\leqslant C\,|||u|||^2_{2,\tau,\theta}.$$

By (5.10) $\mathrm{Im}(y',y)_X=0$, so that the last inequality implies

$$\begin{aligned}
\int_0^\infty&(1+x_1)^{2(\sigma-1+\gamma)}\\
&\times\Big[\omega(x_1)\|w(x_1)\|^2_X+\omega(x_1)^{-1}\Big\|\frac{dw}{dx_1}\Big\|^2_X\Big]dx_1\\
&\leqslant C\,|||u|||^2_{2,\tau,\theta},
\end{aligned}$$

which is exactly (5.15). As was observed in the beginning of the proof of this proposition, inequality (5.1) is an immediate consequence.

We can now finish up the proof of Lemma 4.3. Using the remarks preceding the statement of the proposition, we now have

$$|||u|||_{2,\eta,\theta}\leqslant C\,|||u|||_{2,\tau,-\sigma},$$

where θ,η are arbitrary and $C=C(\tau,\theta,\eta,I)$. In particular, u, $Hu\in L^2(R^n)$. Let $\tau<\eta$ and $-\sigma<\theta$, and set $B=\{x\in R_n/|x_1|\leqslant b\}$. If b is sufficiently large, we obviously have

$$|||\chi(R^n\setminus B)u|||_{2,\tau,-\sigma}\leqslant(1/2C)|||u|||_{2,\eta,\theta},$$

and, since V_0 depends on x_1 only, standard elliptic estimates imply

$$|||\chi(B)u|||_{2,\tau,-\sigma}\leqslant C_1(\|u\|_{L^2(R^n)}+\|H_0u\|_{L^2(R^n)})$$

[see the reasoning following (3.11)].

Now assumption ($V1$) (Sec. IV) implies that the graph norm of H_0 is equivalent to that of H. Since $Hu=\lambda u$, the last inequality yields $|||\chi(B)u|||_{2,\tau,-\sigma}\leqslant C_2\|u\|_{L^2(R^n)}$.

Inequality (4.6) is now a direct consequence of these considerations. This concludes the proof of Lemma 4.3.

VI. UNITARY EQUIVALENCE AND WAVE OPERATORS

In Ben-Artzi[31] the existence of the wave operators $W_\pm=\text{s-lim}_{t\to\pm\infty}e^{itH}e^{-iH_0}$ was proved under very mild growth conditions on V. However, if V satisfies the more restrictive assumptions of Sec. IV, we obtain the following:

Theorem 6.1: Let V satisfy assumptions ($V1$)–($V2$) and $H=H_0+V$. Let $L^2(R^n)_p$ be the closed subspace spanned by eigenfunctions of H and $L^2(R^n)_{ac}$ the subspace of absolute continuity with respect to H. Then:

(a) The point spectrum of H constitutes at most a discrete set of eigenvalues of finite multiplicity.

(b) The absolutely continuous spectrum of H covers the entire real axis and $L^2(R^n)=L^2(R^n)_p\oplus L^2(R^n)_{ac}$ (i.e., there is no singularly continuous spectrum).

(c) The wave operators $W_\pm$ exist and are complete in the sense that $\mathrm{Range}(W_\pm)=L^2(R^n)_{ac}$. Furthermore, if H_{ac}

is the restriction of H to $L^2(R^n)_{ac}$, then H_{ac} is unitarily equivalent to H_0 and $W_\pm^* H_{ac} W_\pm = H_0$.

Proof: (a) and (b) follow from Corollary 4.4. As for part (c), let $e(H)$ denote the discrete set of eigenvalues of H. Let $I \subseteq R \setminus e(H)$ be a compact interval. Obviously, Theorem 4.2 applies equally well to $\widetilde{\Omega}^+(I)$ and $\widetilde{\Omega}^-(I)$, where the latter is the complex conjugate of the former. Thus $R_0(z)$, $R(z)$ can be extended continuously to $\widetilde{\Omega}^\pm(I)$ in $B(\Psi_{\tau,\sigma}, \Phi_{2,\tau',-\sigma})$ and $VR(z)$, $VR_0(z)$ can be extended continuously to $B(\Psi_{\tau,\sigma})$. Since $\tau > 1$ and $\sigma > \frac{1}{2}$, it follows from (2.3) that $\Psi_{\tau,\sigma} \subseteq L^2(R^n)$ is a Hilbert space, dense in $L^2(R^n)$. These remarks allow us to apply a general theorem of Kato and Kuroda,[32] which yields all the assertions of (c).

Q.E.D.

Corollary 6.2: Let $V_0(x_1) = -(\operatorname{sgn} x_1)|x_1|^\alpha$, $0 < \alpha \leqslant 2$, and $V(x) = V_1(x) + V_2(x)$, where $V_1(x)$ is compactly supported and compact from $H_2(R^n)$ into $L^2(R^n)$ and $V_2(x)$ is bounded and satisfies, for some $\beta > 0$,

$$|V_2(x_1,x')| \leqslant CM(x')\{\chi(x_1 \leqslant 0)(1-x_1)^{\alpha-\beta} + \chi(x_1 \geqslant 0)(1+x_1)^{\alpha/2-1-\beta}\}, \tag{6.1}$$

where $M(x') \to 0$ as $|x'| \to \infty$.

Then all the conclusions of Theorem 6.1 hold true for

$$H_0 = -\Delta + V_0(x_1), \quad H = H_0 + V.$$

Proof: V satisfies $(V1)$–$(V2)$ in view of (4.1).

VII. SOME ADDITIONAL REMARKS

Our basic operator H_0 was given by [see (2.7)] $H_0 = -d^2/dx_1^2 + |\xi'|^2 + V_0(x_1)$. Under a suitable transformation it takes the form [see (2.11)]

$$H_0 = -\frac{d^2}{dt^2} + B(t) + C(t), \tag{7.1}$$

viewed as an operator in $L^2(R,X)$, where $B(t) = f(t)B$, B is positive (self-adjoint) in X, and $f(t)$ is monotonically decreasing [see (2.12)], whereas $C(t)$ is bounded in X for every $t \in R$ and $\|C(t)\|$ decays faster than $|t|^{-1-\epsilon}$, $\epsilon > 0$, as $|t| \to \infty$. It is in this context that a limiting absorption principle for H_0 is established, where weighted $L^2(R,X)$ norms are used. This approach to the spectral study of short-range perturbations of the Laplacian was started by Jäger[33] and further utilized by Saito[34] for long-range perturbations, where $C(t)$ includes a smooth part that decays only as $|t|^{-\epsilon}$ (with better decay of its derivative). The authors study the perturbed operator $H = H_0 + V$ in the form (7.1), admitting $V = V(t,x)$ as part of $C(t)$. This approach introduces restrictive boundedness requirements on V and at the same time complicates considerably some of the proofs (e.g., the equivalent of the Uniqueness Lemma 2.3).

In our treatment we split the study of H into two phases. First we establish the limiting absorption principle for H_0, using the form (7.1). In this setup, the basic facts about uniqueness and existence of solutions are simple consequences of the spectral representation of B and asymptotic estimates for one-dimensional solutions. Then we proceed to analyze H, using a perturbation-theoretic argument. Here again the spectral structure of B is exploited, allowing a bootstrap procedure on one-dimensional components of H_0, which is used in proving the rapid decay of eigenfunctions and hence the discreteness of the set of eigenvalues. It should be noted that only minimal information is needed concerning the asymptotic behavior of the kernel $[-d^2/dx_1^2 + |\xi'|^2 + V_0(x_1) - z]^{-1}$ in $L^2(R)$. In particular, only values of ξ', z restricted to bounded sets should be considered.

Our method was used in Ben-Artzi[35] to give a proof of Agmon's theorem concerning the limiting absorption principle for a general short-range perturbation of the Laplacian as well as some cases of "exploding" potentials.

Also, since the "global" Fourier transform techniques are replaced by "local" estimates on the spectrum of B, the underlying domain need not be the whole of R^n. To illustrate such a possibility, consider the following case.

Let $E = [0,\infty) \times D \subseteq R_{x_1} \times R_{x'}^{n-1}$ be a cylinder in R_n, where D is a bounded domain in R^{n-1} having a smooth boundary. The operator $H_0 = -\Delta + V_0(x_1)$ is self-adjoint on

$$D(H_0) = \{u \mid u, H_0 u \in L^2(E),\ u = 0 \text{ on } \partial E\};$$

in fact, it is unitarily equivalent to the direct sum $\Sigma_j \oplus [-d^2/dx_1^2 + \mu_j + V_0(x_1)]$, where $\mu_j \uparrow +\infty$ is the sequence of eigenvalues of $-\Delta_{x'}$ in D (with zero Dirichlet data) and each component is self-adjoint on $L^2(R_+)$ (with zero data at $x_1 = 0$). Let $V(x)$ be bounded in E and satisfy

$$|V(x_1,x')| \leqslant C(1 + |V_L(x_1)|^{1/2})(1+x_1)^{-1-\epsilon}, \quad \epsilon > 0,$$

then it follows from Rellich's compactness theorem that V is relatively compact with respect to H_0 and so $H = H_0 + V$ is self-adjoint in $D(H_0)$. Our arguments in the preceding sections are applicable to the present situation word for word, with $|\xi'|^2$ replaced by $\{\mu_j\}$. Defining $L_s^2(E)$ to be the closure of $C_0^\infty(E)$ under the norm

$$\|f\|_s^2 = \int_E (1+x_1)^{2s}|f(x)|^2\,dx,$$

we get the following therorem.

Theorem 7.1: (a) H_0 is spectrally absolutely continuous and its spectrum is the entire real axis.

(b) Let $I = [\alpha,\beta]$ be a finite interval, $\Omega^+(I)$, $\widetilde{\Omega}(I)$ as in (2.1), and $R_0(z) = (H_0 - z)^{-1}$. Let $\sigma > \frac{1}{2}$. There exists a constant C such that

$$\|R_0(z)f\|_{-\sigma} \leqslant C\|f\|_\sigma, \quad z \in \Omega^+(I).$$

Furthermore, the map $z \to R_0(z) \in B(L_\sigma^2, L_{-\sigma}^2)$ can be extended as a continuous map on $\widetilde{\Omega}^+(I)$, where the uniform topology is used for the operator space.

(c) If $R(z) = (H-z)^{-1}$, the map $z \to R(z)$ can also be extended to $\widetilde{\Omega}^+(I)$ continuously in $B(L_\sigma^2, L_{-\sigma}^2)$, excluding, however, a possible finite set of eigenvalues of finite multiplicity. If u is an eigenfunction associated with such an eigenvalue, it is rapidly decaying in the sense that $u \in L_s^2(E)$ for every $s \in R$.

Observe that the spectrum of each component of H_0 extends over the entire real axis. This follows from a well-known theorem in ordinary differential operators.[36]

Again, the cylinder E can be replaced by a conic domain with apex at the origin, in which case $\mu_j = \mu_j(x_1)$ are monotone decreasing functions.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor A. Devinatz, Professor J. C. Guillot, and Professor P. Rejto for useful discussions concerning various aspects of this work.

APPENDIX

Let $H_0 = -d^2/dx^2 + V_0(x)$, viewed as an operator in $L^2(R)$. Its restriction to $C_0^\infty(R^n)$ is essentially self-adjoint and thus admits a unique self-adjoint extension, still denoted by H. Let $R_0(z) = (H_0 - z)^{-1}$, $z \in \Omega^+(I)$. It is known from Ben-Artzi[37] that $R_0(z)$ is an integral operator having a kernel $\mathscr{K}(s,t,z)$, which can be extended as a continuous function on $R \times R \times C^+$ ($C^+ = \{z | \operatorname{Im} z \geqslant 0\}$). Furthermore, if $I = [\alpha, \beta]$, $\widetilde{\Omega}^+(I)$ as in (2.1), there exists a constant $C = C(I)$ such that

$$|\mathscr{K}(s,t,z)| \leqslant C\omega(s)^{-1/2}\omega(t)^{-1/2}, \quad (s,t,z) \in R \times R \times \widetilde{\Omega}^+(I), \tag{A1}$$

where $\omega(s)$ is defined by (2.15).

Proof of (*3.7*): As in the proof of Lemma 2.4, $u(x_1,\xi')$ solves the equation $(H_0 + |\xi'|^2 - z)u(x_1,\xi') = f(x_1,\xi')$.

The set $\{z - |\xi'|^2 | z \in \widetilde{\Omega}^+(I), \xi' \in D\}$ is contained in some $\widetilde{\Omega}^+(L)$, where L is again a compact interval. Hence,

$$u(s,\xi') = \int_R \mathscr{K}(s,t,z - |\xi'|^2) f(t,\xi')\, dt, \tag{A2}$$

and, using (A1),

$$\begin{aligned} \omega(s)|u(s,\xi')|^2 \leqslant C \Bigg[&\left(\int_{R_+} |f(t,\xi')|^2 \omega(t)^{-1} (1+t)^{2\sigma} dt \right) \\ &\times \left(\int_{R_+} (1+t)^{-2\sigma} dt \right) \\ &+ \left(\int_{R_-} |f(t,\xi')|^2 \omega(t)^{-2} (1-t)^{2\tau} dt \right) \\ &\times \left(\int_{R_-} \omega(t)(1-t)^{-2\tau} dt \right) \Bigg]. \end{aligned} \tag{A3}$$

Note that the last integral converges since $\tau > 1$ and $\omega(t) = O(|t|)$. Integrating over ξ' and using (2.13), (2.17), and the definition (2.9) of the norms on the right-hand side, we obtain (3.7).

Proof of (*3.13*): From (A2) we have

$$\begin{aligned} u(s,\xi',z) &= \int_{|t|>b} + \int_{|t|<b} \mathscr{K}(s,t,z - |\xi'|^2) f(t,\xi')\, dt \\ &= u_1(s,\xi',z) + u_2(s,\xi',z). \end{aligned}$$

As in (A3), we have, uniformly in $z \in \widetilde{\Omega}^+(I)$,

$$\int_D \omega(s)|u_1(s,\xi',z)|^2\, d\xi' \leqslant C\epsilon(b)\|f\|_{\tau,\sigma}^2,$$

$$\epsilon(b) = \max((1+b)^{-2\sigma+1}, (1+b)^{-2\tau+2}).$$

The continuity of $\mathscr{K}(s,t,z - |\xi'|^2)$ as a function of all arguments implies

$$\int_D |u_2(s,\xi',z_k) - u_2(s,\xi',z)|^2\, d\xi' < \epsilon \|f\|_{\tau,\sigma}^2 \quad \text{if } k > K(\epsilon), \; s \in J.$$

[1]E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1961).
[2]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed., Chap. X.
[3]Ref. 1.
[4]P. A. Rejto and K. Sinha, Helv. Phys. Acta **49**, 389–413 (1976).
[5]M. Ben-Artzi, "An application of asymptotic techniques to certain problems of spectral and scattering theory of Stark-like Hamiltonians," Trans. Am. Math. Soc. **278**, 817–839 (1983).
[6]J. E. Avron and I. W. Herbst, Commun. Math. Phys. **52**, 239–54 (1977).
[7]B. Simon, Duke Math. J. **46**, 119–68 (1979).
[8]I. W. Herbst, Math. Z. **155**, 55–70 (1977).
[9]K. Yajima, J. Fac. Sci. Univ. Tokyo 1A **26**, 377–89 (1979).
[10]W. Jäger, Math. Z. **113**, 68–98 (1970).
[11]Y. Saito, *Spectral Representations for Schrödinger Operators with Long-Range Potentials*, Lecture Notes in Mathematics No. 727 (Springer, Berlin, 1979).
[12]T. Kato and S. T. Kuroda, Rocky Mount. J. Math. **1**, 127–71 (1971).
[13]Ref. 5.
[14]Ref. 9.
[15]Ref. 5.
[16]Ref. 5, Appendix.
[17]Ref. 11.
[18]Ref. 5.
[19]Ref. 5.
[20]Ref. 5, Theorem 4.1.
[21]Ref. 2, Chap. IV.
[22]Ref. 8.
[23]Ref. 9.
[24]Ref. 9.
[25]Ref. 2, Chap. IV.
[26]N. Dunford and J. Schwartz, *Linear Operators II* (Interscience, New York, 1963), p. 1202.
[27]S. Agmon, Ann. Scuol. Norm. Sup. Pisa **2**, 151–218 (1975).
[28]Ref. 8, Proposition 4.2.
[29]M. Ben-Artzi, "Resolvent estimates for a certain class of Schrödinger operators with exploding potentials," J. Diff. Eqs. (to be published).
[30]Ref. 5.
[31]Ref. 5.
[32]Ref. 12, Theorem 6.3.
[33]Ref. 10.
[34]Ref. 11.
[35]Ref. 29.
[36]Ref. 26, Theorem XIII, 7.16.
[37]Ref. 5, Sec. 4.

The minimum potential strength for the existence of bound states

B. G. Sidharth[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 29 July 1982; accepted for publication 24 June 1983)

Bounds are deduced for both the minimum potential strength for the existence of bound states and the scattering length for everywhere attractive potentials which fall off faster than $(1/r^3)$ for large r. These bounds are the second Born approximations. A variational calculation is then invoked to prove that, interestingly, these bounds are actually good estimates when a specified sufficiency condition is satisfied. The estimates are then shown to compare favorably with known results for a number of potentials of practical interest, viz., the square-well, Morse, exponential, and Yukawa potentials.

PACS numbers: 03.65.Nk, 03.65.Ge

INTRODUCTION

We consider the radial Schrödinger equation, when the potential $\lambda U(r)$ is attractive, λ being a parameter. It is known that[1] if $\lambda U(r) \to 0$ like r^{-n}, as $r \to \infty$, $n < 2$, then there are an infinite number of bound states for any $\lambda > 0$.

In general there may exist a $\lambda_0 > 0$ such that there are bound states when $\lambda > \lambda_0$, but not when $\lambda < \lambda_0$. We shall call λ_0 the minimum potential strength. The question that arises is, what is the value of λ_0?

In what follows, we shall consider potentials that are nonsingular; that is, $|r^2 U(r)| < M$ as $r \to 0$. Otherwise the question of a minimum potential strength does not arise.[2,3]

We shall first show that if the potential falls off as fast as or faster than r^{-2} as $r \to \infty$, then the minimum potential strength λ_0 is not arbitrarily small. Next we shall deduce an estimate for λ_0, when $r^{-m} U(r) \to 0$ as $r \to \infty$ for some $m > 3$. In the process we shall also obtain an estimate for the scattering length.

THE MINIMUM POTENTIAL STRENGTH

For the radial Schrödinger equation,

$$u'' + [K^2 - l(l+1)/r^2 - \lambda U(r)]u = 0, \quad u(0) = 0, \quad (1)$$

we shall, in the sequel, require the following results:

(A) The ground state for a spherically symmetrical potential is an $l = 0$ state[4];

(B) The number of bound states for a particular l equals the number of nodes of the zero energy, that is, $K^2 = 0$, wave function in $0 < r \leqslant \infty$[5];

(C) For an attractive potential $\lambda U(r)$, with $l = 0$, if the phase shift $\delta(K;\lambda) \to 0$ as $K \to 0$, there are no bound states. The first bound state appears when the scattering length $C_0 \equiv -\lim_{K\to 0} (\tan\delta/K) = \infty$.[6,7]

Now, let $\lambda U(r)$ fall off as fast as or faster than r^{-2} as $r \to \infty$. We shall prove that λ_0 cannot be arbitrarily small.

By (A), the first bound state is an $l = 0$ state. Assume $\lambda_0 = 0$. Then, bound states exist for every $\lambda > 0$, when $l = 0$.

From (1), the $l = 0$ zero-energy wave function V satisfies,

$$V'' + \lambda |U(r)| V = 0,$$

remembering that

$$-\lambda U(r) = \lambda |U(r)|$$

because the potential is attractive.

By (B) it follows that there lies at least one zero, r_0 say, of V in $r > 0$.

Now let W be *any* solution of

$$W'' + (1/4r^2) W = 0.$$

As $U(r)$ is nonsingular, we can choose λ suitably small, such that

$$\lambda |U(r)| \leqslant 1/4r^2$$

everywhere. Hence, by Sturm's lemma,[8] there lies at least one zero of W in $0 < r < r_0$. This is impossible if we choose $W = r^{1/2}$.

So bound states cannot exist for arbitrarily small λ when $U(r)$ falls off as fast as or faster than r^{-2} as $r \to \infty$.

We next proceed to obtain an estimate for λ_0. For this, we assume that

$$\lim_{r\to\infty} r^m U(r) = 0 \text{ for some } m > 3. \quad (2)$$

We now use the inequalities[9,10]:

$$\frac{\partial}{\partial\lambda}\left[\frac{1}{\lambda} + KB \cot\delta(\lambda)\right] \geqslant 0, \quad (3)$$

$$1/\lambda + KB \cot\delta(\lambda) \geqslant C/B, \quad (4)$$

where

$$B = \frac{1}{K^2}\int_0^\infty \sin^2 Kr U(r) dr,$$

$$C = -\frac{1}{K^3}\int_0^\infty U(r)\sin Kr\left[\cos Kr \int_0^r \sin^2 Kr' U(r') dr' + \sin Kr \int_r^\infty \sin Kr' \cos Kr' U(r') dr'\right] dr. \quad (5)$$

The inequalities (3) and (4) are valid when $0 < |\delta| < \pi$ and for potentials which are everywhere attractive only or repulsive only.

It can be easily shown that

$$\tan\delta(\lambda) = -\lambda KB + O(\lambda^2), \quad (6)$$

$$\tan\delta(\lambda) = -\lambda KB - \lambda^2 KC + O(\lambda^3). \quad (7)$$

Equation (6) shows that "$-\lambda KB$" is the first Born estimate for $\tan\delta$ while (7) shows that B and C are related to the second Born estimate.

a) Permanent address: Birla Planetarium & Astronomical Research Centre, 96, Jawaharlal Nehru Road, Calcutta 700071, India.

Now, from (5), we get,

$$\lim_{K\to 0} B \equiv \overline{B} = \int_0^\infty r^2 U(r)dr,$$

$$\lim_{K\to 0} C \equiv \overline{C} = -\left[\int_0^\infty U(r)r\left\{\int_0^r r'^2 U(r')dr' + r\int_r^\infty r' U(r')dr'\right\}dr\right]. \tag{8}$$

The justification for passing to the limit within the integral is that sin Kr, etc. are uniformly continuous with respect to K and r, while, owing to (2), the integrals are uniformly convergent with respect to K.[11]

So, taking the limit $K\to 0$ in (4), we get

$$\frac{1}{\lambda} - \frac{\overline{B}}{C_0(\lambda)} \geqslant \frac{\overline{C}}{\overline{B}}, \tag{9}$$

where $C_0(\lambda)$ is the scattering length defined in (C).

From (C) we know that at the minimum potential strength λ_0, $1/C_0(\lambda_0) = 0$. So, from (9) we get

$$\frac{1}{\lambda_0} \geqslant \frac{\overline{C}}{\overline{B}} \quad \text{or} \quad \lambda_0 \leqslant \frac{\overline{B}}{\overline{C}} \equiv \lambda_1, \tag{10}$$

where $\overline{B}$ and $\overline{C}$ are given by (8).

The inequality (10) has also been obtained from other considerations.[12] It is interesting that λ_1 is the second Born estimate for the minimum potential strength in that it follows from the expression (7), when λ^3 and higher powers can be neglected.

We shall now show that the inequality (10) is an approximate equality under certain conditions, which are valid in a number of cases of practical interest. Firstly we note that from (3) it follows that

$$\frac{1}{\lambda_0} \leqslant \frac{1}{\lambda'} + \lim_{K\to 0} KB \cot \delta(\lambda') = \frac{1}{\lambda'} - \frac{\overline{B}}{C_0(\lambda')}, \tag{11}$$

where $C_0(\lambda')$ is the scattering length at the potential strength $\lambda' > \lambda_0$.

We next use the fact that[7] the scattering length is given by

$$-C_0(\lambda) = a - \int_0^\infty \phi(r) H \phi(r) dr, \tag{12}$$

where the function $\phi(r)$ is such that the expression on the right side is stationary for arbitrary variations of the function $\phi(r)$, and where

$$\phi(0) = 0, \quad \phi(r) \to r + a \quad \text{as } r \to \infty \tag{13}$$

and

$$H \equiv -\frac{d^2}{dr^2} + \lambda U(r).$$

Let us find a wave function ϕ for which the right side of (12) becomes stationary. For this, we choose

$$\phi(r) = \psi(r) + \alpha\eta(r),$$

where

$$\psi(r) = r + \lambda\int_0^r r'^2 U(r')dr' + \lambda r\int_r^\infty r' U(r')dr' \equiv r - \lambda\theta(r), \tag{14}$$

so that

$$\psi(0) = 0, \quad \psi(r) \to r + a \quad \text{as } r \to \infty,$$

where

$$a = \lambda\int_0^\infty r^2 U(r)dr.$$

ψ is the first Born estimate; that is, the estimate obtained by neglecting λ^2 and higher powers in the expression for the wave function u defined in (1). α is a parameter and $\eta(r)$ is arbitrary, except that

$$\eta(0) = 0, \quad \eta(\infty) = b, \tag{15}$$

This ensures that $\phi(r)$ satisfies (13), viz.,

$$\phi(0) = 0, \quad \phi(r) \to r + (a + \alpha b) \quad \text{as } r \to \infty.$$

We substitute (14) on the right side of (12) and vary α about the value $\alpha = 0$. This yields

$$\frac{d}{d\alpha}\left[(a + \alpha b) - \int_0^\infty \psi H\psi\, dr - \alpha\int_0^\infty \{\psi H\eta + \eta H\psi\} dr - \alpha^2\int_0^\infty \eta H\eta\, dr\right] = 0 \quad \text{at } \alpha = 0,$$

or

$$b + \int_0^\infty \psi\eta''\, dr + \int_0^\infty \eta\psi''\, dr - 2\lambda\int_0^\infty U(r)\psi\eta\, dr = 0. \tag{16}$$

From (14),

$$\psi' = 1 + \lambda\int_r^\infty rU(r)dr, \quad \psi'' = -\lambda r U(r).$$

Also, integrating by parts, the integral,

$$\int_0^\infty \psi\eta''\, dr = -\lambda\int_0^\infty rU(r)\eta\, dr - b,$$

where we have used (15) and the fact that $\psi'(r) \to 1$ as $r \to \infty$.

So, (16) becomes

$$\int_0^\infty U(r)\eta\{\psi(r) + r\} dr = 0. \tag{17}$$

This means that, if

$$\psi(r) + r = 0 \text{ everywhere,} \tag{18}$$

then substitution of $\phi(r) = [\psi(r) + \alpha\eta(r)]_{\alpha=0} = \psi(r)$ in (12) will give the value of $C_0(\lambda)$. Actually, if $U(r)$ falls off rapidly, so that $U(r) \approx 0$ for $r > R$ then (18) need be true only for $r < R$, because

$$\int_R^\infty U(r)\eta(\psi + r)dr \approx 0.$$

We can see roughly, why we can expect that (18) will hold, at least when r is not large.

From (14)

$$\psi(r) = r\left(1 + \lambda\int_0^\infty rU(r)dr\right) + \lambda\left(\int_0^r r'^2 U(r')dr' - r\int_0^r r' U(r')dr'\right).$$

If $U(r) = \text{const} + O(r)$ as $r \to 0$, then for small r,

$$\psi = a_0 r + O(r^3).$$

If $U(r) \sim 1/r$ as $r \to 0$, then

$$\psi = a_0 r + O(r^2).$$

In either case, at least for small r,

$$\psi \approx a_0 r,$$

though in the latter case, this is true for small values of r as compared to the former case.

Moreover, a suitable choice of λ would ensure that $a_0 = -1$, so that (18) would be valid for sufficiently small values of r. If $U(r)$ falls off rapidly, therefore, (17) would be valid for this value of λ.

The rigorous condition which ensures that (18) will hold, i.e., that

$$\lambda\theta = 2r,$$

where θ is defined in (14), can be obtained in a form convenient for the subsequent work:

$$\left(\int_0^\infty rU\theta\, dr\right)^2 = \left(\int_0^\infty r^2 U\, dr\right)\left(\int_0^\infty \theta^2 U\, dr\right). \tag{19}$$

This is so because it is well known that the Cauchy–Schwarz inequality,

$$\left(\int fg\, dr\right)^2 \leqslant \left(\int f^2\, dr\right)\left(\int g^2\, dr\right),$$

reduces to an equality only when $f \propto g$.

The condition (19) ensures that $\theta \propto r$. We can easily compute the value λ' for which

$$\lambda'\theta = 2r \tag{20}$$

in the following manner: In the expression (10) for λ_1, viz.,

$$\lambda_1 = \frac{-\int_0^\infty r^2 U(r) dr}{\int_0^\infty rU(r)\left[\int_0^r r'^2 U(r')dr' + r\int_r^\infty r' U(r')dr'\right]dr}$$

$$\equiv \frac{\int_0^\infty r^2 U(r) dr}{\int_0^\infty rU(r)\theta(r) dr}, \tag{21}$$

we multiply numerator and denominator by λ' and substitute Eq. (20). This gives

$$\lambda' = 2\lambda_1. \tag{22}$$

[Equation (22) can be obtained alternatively, directly from the variational problem.]

Next, we rewrite the condition (19) in a more compact form:

$$\left(\frac{\int_0^\infty rU\theta\, dr}{\int_0^\infty r^2 U\, dr}\right)^2 \equiv \frac{1}{\lambda_1^2} = \frac{\int_0^\infty U\theta^2\, dr}{\int_0^\infty Ur^2\, dr} \equiv D, \tag{23}$$

where we have used (21).

What we have proved is, that if (19) or (23) is satisfied, then, for λ' given by (22), the scattering length is obtained by substituting (14) in (12). This gives

$$-C_0(\lambda') = \lambda'\overline{B} - \left[\int_0^\infty \psi\{\lambda' rU + \lambda' U\psi\} dr\right]$$

as $\psi'' = -\lambda rU$, $\overline{B}$ being defined in (8), or

$$\frac{C_0(\lambda')}{\lambda'} = -\overline{B} + \int_0^\infty U(\psi - r)^2\, dr + 2\int_0^\infty r^2 U\, dr + 3\int_0^\infty rU(\psi - r) dr.$$

So,

$$\frac{C_0(\lambda')}{\lambda'\overline{B}} = 1 + \lambda'^2 \frac{\int_0^\infty U\theta^2\, dr}{\overline{B}} - 3\lambda' \frac{\int_0^\infty rU\theta\, dr}{\overline{B}}.$$

On using the condition (19) and Eq. (21) we get

$$\frac{C_0(\lambda')}{\lambda'\overline{B}} = 1 + \left(\frac{\lambda'}{\lambda_1}\right)^2 - 3\frac{\lambda'}{\lambda_1}.$$

Remembering that $\lambda' = 2\lambda_1$ by (22), we have finally,

$$C_0(\lambda')/\lambda'\overline{B} = -1.$$

Substituting this in (11), we get

$$\frac{1}{\lambda_0} \leqslant \frac{2}{\lambda'} = \frac{1}{\lambda_1} \quad \text{or} \quad \lambda_0 \geqslant \lambda_1. \tag{24}$$

The inequality (24) is valid for potentials which are not necessarily everywhere attractive. But the inequality (10) is valid for everywhere attractive potentials only.

Combining (24) with (10) we get

$$\lambda_0 = \lambda_1, \tag{25}$$

where λ_1 is given by (21).

In summary, for everywhere attractive potentials, whenever the sufficiency condition (23) is satisfied, then (25) is valid.

Further, taking the limit $K \to 0$ in inequality (3) and combining it with inequality (9), we get for any $\lambda < \lambda_0$,

$$\frac{1}{\lambda_0} \geqslant \frac{1}{\lambda} - \frac{\overline{B}}{C_0(\lambda)} \geqslant \frac{\overline{C}}{\overline{B}} = \frac{1}{\lambda_1}.$$

When (23) is satisfied as $\lambda_0 = \lambda_1$, this shows that

$$\frac{1}{\lambda} - \frac{\overline{B}}{C_0(\lambda)} = \frac{1}{\lambda_0}$$

or

$$C_0(\lambda) = \frac{\overline{B}}{1/\lambda - 1/\lambda_0}, \quad \lambda < \lambda_0 \tag{26}$$

which is an expression for the scattering length.

If (23) is satisfied only approximately, then the ensuing steps show that (25) and (26) are approximate equalities. We have seen that (23) can be expected to be valid. Let us now verify in some examples that this is so, and that consequently (25) holds.

(*i*) *The attractive square-well potential:* $U(r) = -V < 0$, $r \leqslant a$, $U(r) = 0$, $r > a$.

In this case

$$D\lambda_1^2 = 1.0125,$$

so that

$$D \approx 1/\lambda_1^2;$$

that is, (23) is nearly satisfied.

Formula (25) gives

$$\lambda_0 Va^2 \approx 2.5.$$

From the theory, it is known[13] that bound states exist only if

$$\lambda Va^2 > 2.46.$$

(*ii*) *The Morse potential:* $U(r) = \lambda \exp(-2ar) - 2\lambda \exp(-ar)$.

This time

$$D\lambda_1^2 = 1.081\,8014$$

and condition (23) is approximately satisfied.

Formula (25) gives

$$\sqrt{\lambda_0}/a \approx 0.942.$$

From the theory, it is known[14] that bound states begin to exist for

$$\sqrt{\lambda_0}/a \approx 1.$$

(iii) *The exponential potential:* $U(r) = -\lambda \exp(-2ar)$. Here

$$D\lambda_1^2 = 1.0889.$$

So condition (23) is approximately satisfied.

Formula (25) gives

$$\sqrt{\lambda_0}/a \approx 2.529.$$

It is known from theory[15] that the minimum potential strength is given by

$$\sqrt{\lambda_0}/a \approx 2.405.$$

(iv) *The Yukawa potential:* $U(r) = -(\lambda/r)\exp(-ar)$. By a suitable transformation, we can set $a = 1$.

In this case,

$$D\lambda_1^2 = 1.150\,728.$$

Here, $U(r) \sim 1/r$ as $r \to 0$. Though condition (23) holds approximately, as remarked earlier, the agreement is not as good as in the preceding cases where $U(r) \to \text{const}$ as $r \to 0$.

Formula (25) gives

$$\lambda_0 = 2.$$

A numerical integration and a variational calculation give, respectively,[16]

$$\lambda_0 = 1.68 \text{ and } \lambda_0 = 1.74.$$

Thus in all the above cases, formula (25) gives a credible result.

CONCLUDING REMARKS

We have seen that the second Born estimates for the minimum potential strength and the scattering length, viz., formulas (25) and (26), respectively, are valid for everywhere attractive potentials when condition (23) is satisfied. We would expect that the second Born estimates would be valid only for weak potentials when the cube and higher powers of the potential strength can be neglected, and not for potentials which are so strong that they can produce bound states. Indeed, in the examples considered, the minimum potential strength has not been small. We have shown that, surprisingly, the condition (23) ensures that these second Born estimates are valid right up to the minimum potential strength.

ACKNOWLEDGMENTS

The author is grateful to Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste, Italy, and also to the referee, in the light of whose constructive comments, the paper has been rewritten.

[1] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, New York, 1965), pp. 54 and 55.
[2] W. M. Frank, D. J. Land, and M. S. Spector, Rev. Mod. Phys. **43**(1), 36 (1971).
[3] B. G. Sidharth, Act. Phys. Pol. B **9**(4), 359 (1978).
[4] B. W. Downs, Am. J. Phys. **31**, 277 (1963).
[5] T. Regge and V. de Alfaro, *Potential Scattering* (North-Holland, Amsterdam, 1965), p. 73.
[6] P. Roman, *Advanced Quantum Theory* (Addison-Wesley, Reading, MA, 1965), pp. 175–179.
[7] R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1965), p. 308ff.
[8] E. C. Titchmarsh, *Eigenfunction Expansion (Part I)* (Oxford U. P., New York, 1962), p. 107.
[9] B. G. Sidharth, Bull. Cal. Math. Soc. **71**(6), 370 (1979).
[10] B. G. Sidharth and A. Abdel-Hafez, Act. Phys. Pol. A **56**(4), 577 (1979).
[11] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge U. P., New York, 1952), pp. 73 and 74.
[12] W. Kohn, Rev. Mod. Phys. **26**(3), 292 (1954).
[13] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1949), p. 77.
[14] E. U. Condon and P. M. Morse, *Quantum Mechanics* (McGraw-Hill, New York, 1929), p. 71.
[15] D. Ter Haar, *Selected Problems in Quantum Mechanics* (Academic, New York, 1964), p. 218.
[16] H. M. Schey and J. L. Schwartz, Phys. Rev. **B139**, 1428 (1965).

On quantum solitons and their classical relatives: Reducible quantum fields and infinite constituent "elementary" systems

Piotr Garbaczewski
Institute of Theoretical Physics, University of Wrocław, 50-205 Wrocław, Poland

(Received 30 November 1982; accepted for publication 1 April 1983)

We demonstrate that the emergence of translation modes in the quantization of some at least nonlinear field theory models (like, e.g., ϕ^4 or the sine–Gordon systems) implies a specific structure of their state spaces namely this of the direct integral Hilbert space, which follows from the reducibility of the involved quantum field canonical commutation relations (CCR) algebras. As a special manifestation of this structure, one recovers infinite constituent "elementary" quantum systems living in the commutant of the CCR algebra, which appear as the Schrödinger or the two level ones. The corresponding Hamiltonians are derived. In addition, we propose a modification of the standard infrared Hilbert (photon field) space construction employed in quantum electrodynamics. We demonstrate that, in principle, Fermi (CAR) generators, carrying the spin–charge–momentum labels of Dirac particles, can be defined as operators in the electromagnetic (photon field) Hilbert space. The photon field (CCR) algebra is highly reducible, and in the present case fermions arise in the commutant of it, playing the role of intertwining operators.

PACS numbers: 03.70. + k, 11.10.Lm

1. MOTIVATION

Usually one quantizes the classical field theory models under a tacit assumption that the field ϕ exhibits at most a space-time functional dependence: $\phi = \phi(\mathbf{x}, t)$, $\mathbf{x} \in R^3$ (or rather that the $\mathbf{x}, t$ dependence only is relevant for the quantization). An example of a classical field theory where the configuration space variable $\mathbf{x}$ is augmented by an auxiliary real variable w was considered in Ref. 1:

$$\mathscr{L} = \mathscr{L}(x, w) = \tfrac{1}{2}[\partial_\mu \phi(x, w)]^2 - \tfrac{1}{2}m^2\phi^2(x, w) - V[\phi, x, w] \tag{1.1}$$

with $x \in M^4$, ∂_μ being the space-time derivative, $x = (\mathbf{x}, t)$.

One insists there on a dynamical independence for all space and time of fields with distinct w value. This w-ultralocality of the field (1.1) has been exploited in Ref. 1 to construct a relatively simple quantized model. However, under an assumption that the quantum field CCR algebra is defined as follows:

$$[A_l(w), A^*_{l'}(w')]_- = \delta_{ll'}\delta(w - w'),$$
$$[A_l(w), A_l(w')]_- = 0, \quad A_l(w)\Omega = 0 \quad \forall l, w. \tag{1.2}$$

Here $l = 1, 2, \cdots$ enumerates the oscillatorlike degrees of freedom of the wth field, and Ω is the Fock state. Let us recall that a conventional quantization procedure for the neutral scalar field theory model [omit w in (1.1)] would result in the commutation relations

$$[A_l, A^*_{l'}]_- = \delta_{ll'} \quad [A_l, A_{l'}]_- = 0 = [A^*_l, A^*_{l'}]_-$$
$$A_l\Omega = 0 \quad \forall l \tag{1.3}$$

and, as noticed, for example, in Refs. 2–4 still remains successful for models of the type (1.1). An oversimplified model with local fields which are at most bilinear in the generators (1.2) does not occur in case (1.3).

On the other hand, the quantization procedures developed during the 1970s for, for example, the sine–Gordon model, did ignore the fact that classical solutions of its field equation, at least in the soliton sector, exhibit a parametrization additional to x,t. In the simplest 1-soliton case it reduces to the real w-parametrization of (1.1), while for N-solitons the situation becomes more complex. The underlying parametrization enters soliton solutions $\phi = \phi(x, t)$ of the $(1 + 1)$-dimensional field equation:

$$\Box\phi(x, t) = m^2 \sin\phi(x, t) \quad m > 0, \tag{1.4}$$

via nonplane wave solutions $\varphi(x, t)$ of the free field equation

$$(\Box - m^2)\varphi(x, t) = 0,$$
$$\varphi(x, t) = \varphi_a(x, t) = \exp[m\gamma_a(x - v_a t) + \delta],$$
$$v_a = \frac{|a|^2 - 1}{|a|^2 + 1}, \quad \gamma_a = (\operatorname{sgn} a)(1 - v_a^2)^{-1/2}, \tag{1.5}$$
$$a \in R^1, \quad |a| \in (0, \infty)$$

so that (see Refs. 2–5)

$$\phi = \phi(x, t) = \phi[\varphi_a](x, t) = \phi_a(x, t) = \tfrac{1}{4}\tan^{-1}\varphi_a(x, t) \tag{1.6}$$

for one soliton, while

$$\phi(x, t) = \phi[\varphi_{a_1}, \ldots, \varphi_{a_N}](x, t) = \phi_{a_1 \cdots a_N}(x, t) \tag{1.7}$$

for N solitons, in the absence of "breathing" components. When the latter are present the parameters may become complex, and then are required to appear in complex conjugate pairs as, e.g., in the N-soliton (one breather) case of

$$\phi(x, t) = \phi_{aa_2a_3,\ldots,a^*,\ldots,a_N}(x, t), \tag{1.8}$$

where N-2 parameters are real and there is a complex conjugate pair.

Let us mention that in (1.5)–(1.8) one has still a freedom of choice of the phases $\delta_1, \ldots, \delta_N$. For 1-solitons, the sine–

Gordon Lagrangian density has the form (1.1) with the parameter w replaced by a. In the N-soliton case the parameter w of (1.1) may be replaced by a sequence $(a_1, \ldots, a_N) = (a)$.

As mentioned before, the (a) parametrization is insufficient for a unique characterization of the sine–Gordon soliton fields: the (δ) parametrization should also be taken into account. In the notation

$$\varphi_a(x, t) = \exp m\gamma_a(x + q) = \varphi_{aq}(x),$$
$$q = q_a = v_a(t_0 - t) \Rightarrow \delta = \delta_a = m\gamma_a v_a t_0,$$

we have absorbed both δ and time t dependence of φ_a in a new parameter q. The a dependence of $q = q_a$ can be ignored due to the freedom of choice of the initial time instant t_0. Hence (1.7) can be rewritten as

$$\phi(x, t) = \phi_{(a, q)}(x) = \phi_{a_1 \cdots a_N q_1 \cdots q_N}(x), \tag{1.9}$$

thus leading to the $(2N + 1)$-parameter family of N-solitons, where both (a) and (q) are viewed as independent parametric families. Notice that the time variable completely disappears from the formalism, which is obviously an effect of the translation freedom of (1.4).

Let $\widehat{\varphi}_{\rm in}(x, 0), \widehat{\pi}_{\rm in}(x, 0)$ be a canonical pair generating the CCR algebra of the mass m neutral scalar field in $1 + 1$ dimensions:

$$[\widehat{\varphi}_{\rm in}(x, 0), \widehat{\pi}_{\rm in}(y, 0)]_- = i\delta(x - y),$$
$$a(k) = \int dx \exp(-ikx)\left[\sqrt{k^2 + m^2}\,\widehat{\varphi}_{\rm in}(x, 0) + i\widehat{\pi}_{\rm in}(x, 0)\right], \tag{1.10}$$
$$a(k)|0) = 0 \quad \forall k.$$

We introduce the coherent states (coherent soliton states of Refs. 2–4) as these states of the field (1.10) which satisfy

$$a(k)|\varphi) = \alpha(k)|\varphi),$$
$$(\varphi|\widehat{\varphi}_{\rm in}(x)|\varphi) = (0)\{\widehat{\varphi}_{\rm in}(x) + \varphi(x)\}|0) = \varphi(x), \tag{1.11}$$
$$(\Box - m^2)\varphi(x) = 0.$$

Here $\alpha(k)$ is not a square-integrable function, since we make an identification of $\varphi(x)$ with either $\varphi_{aq}(x)$, (1.8) in the sine–Gordon soliton case, or with $\Sigma_{i=1}^N \varphi_{a_i q_i}(x)$ as required in the N-soliton case (without breathers).[2–4] Coherent states (2.11) should be viewed as continuous generalizations of the more familiar direct product coherent states

$$|\varphi) = \mathrm{U}_\varphi^*|0) = \prod_k^{\otimes}\{\exp(\alpha a^* - \overline{\alpha} a)f_0\}_k \tag{1.12}$$

with an ultraviolet cutoff implicit, f^0 being a Fock state of the Schrödinger representation of the CCR algebra. If one takes a coherent soliton state $|\varphi)$, then applies polynomials in creation and annihilation operators (1.10) to $|\varphi)$, and finally closes the resulting set of vectors, one arrives at the Hilbert space IDPS $(|\varphi))$, which is separable and carries an irreducible representation of the CCR algebra. IDPS $(|\varphi))$ is the one among infinitely many separable subspaces of the complete (von Neumann's) Hilbert space $\mathscr{H}$ of the quantum field (1.10).

Since we use coherent product states as the generating vectors for IDPS $(|\varphi))$, the theory of Ref. 6 can be applied to classify the unitarily (in) equivalent representations of the CCR algebra affiliated with different coherent states. In Ref. 2 we have done it for the soliton coherent states of the sine–Gordon fields; compare also Ref. 7.

Because the above-introduced parametrization (a, δ) characterizes uniquely both the soliton field (1.9) and its free field constituents φ_{aq}, we arrive at the unambiguous labeling of quantum soliton (Hilbert space) sectors:

$$|\varphi) = |a_1, \ldots, a_N, q_1, \ldots, q_N) \Rightarrow \mathrm{IDPS}(|\varphi)) := \mathscr{H}^N_{(a,q)}. \tag{1.13}$$

It happens because the free fields $\Sigma_{i=1}^N \varphi_{a_i q_i}(x)$ are the boson transformation parameters of (1.11). Suppose we have two distinct coherent soliton states $|\varphi)$ and $|\varphi')$ in $\mathscr{H}$. If the respective boson transformation parameters $\varphi(x)$ and $\varphi'(x)$ satisfy

$$\int_{R^1} dx\, [\varphi(x) - \varphi'(x)]^2 \not< \infty, \tag{1.14}$$

then the representations of the CCR algebra in IDPS $(|\varphi))$ and IDPS$(|\varphi'))$ respectively, are unitarily inequivalent,[3,6] and the scalar product $(\varphi|\varphi') = 0$ is conventionally introduced in $\mathscr{H}$ for $\varphi \neq \varphi'$ while $\|\varphi\| = \|\varphi'\| = 1$ and $\|\varphi\|^2 = (\varphi|\varphi)$. Obviously, the function $\varphi_{aq}(x)$ of (1.8) is not square integrable on the real line R^1 and the same concerns both $\Sigma_{i=1}^N \varphi_{a_i q_i}(x)$ and any $(\varphi - \varphi')(x)$ with $\varphi \neq \varphi'$.

For an example of the one-soliton boson transformation parameter $\varphi_{aq}(x)$, let us notice that even if $q = q'$ and $a = -a'$ (i.e., $|a| = |a'|$), we still have $(\varphi|\varphi') = 0$. More generally, classical N-soliton sine–Gordon fields give rise to the rich (nondenumerable) family of normalized and pairwise-inequivalent vectors together with the related CCR algebra carrier spaces of (1.13) $\{\mathscr{H}^N_{(aq)}\}^{N = 1, 2, \ldots}_{a, q \in R^1}$, where each $\mathscr{H}^N_{(a,q)} = \mathrm{IDPS}\,(|a_1, \ldots, a_N, q_1, \ldots, q_N))$ is by definition separable. We get in fact the field $\mathscr{H}_t$ of separable Hilbert spaces labeled by a continuous index set $T \ni t$. The present paper is the fourth one in the series of investigations devoted to the problem of quantum solitons and their classical relatives. We continue the discussion of different aspects of the quantization of nonlinear fields, with emphasis on the $(1 + 1)$-dimensional models (Refs. 2–4; see also Ref. 5). The presence of the soliton solutions is known to complicate the traditional local quantization program by giving rise to zero-energy modes (related to translations) and then necessitating the "collective coordinates."

In the present paper we show that the translation freedom implies a very specific form of the state space of the quantized nonlinear system, namely, this of the direct integral Hilbert space, which we describe in Sec. 2 together with the derivation of the (inherent) infinite constituent "elementary" quantum systems. In Sec. 3 we introduce the "elementary" two-level systems, and investigate possible forms of their interaction. In Sec. 4 we propose a modification of the standard infrared Hilbert space construction employed in quantum electrodynamics, to allow the direct integral procedures of Sec. 2. Then, we construct Fermi (CAR) generators and represent them as operators in the electromagnetic field Hilbert space. The fermions carry the spin–charge–momentum labels of the Dirac particles and belong to the commutant of the photon field algebra (this is a consequence of the reducibility of the latter).

Remark: Our knowledge of soluble models in $1+1$ dimensions is not broad enough to state that the above sine–Gordon structure (1.4)–(1.15) may always be found. Since, however, some of these features were observed to be valid for ϕ^4 and appeared also in the quantization of the c-number massive Thirring model, we feel that, even if not quite general, the sine–Gordon structure assumption may be of some use for the construction of the nontrivial quantum field theory models, preserving a correct classics–quanta relationship.

2. A LITTLE BIT OF THE OLD-FASHIONED MATHEMATICS: CONTINUOUS DIRECT SUMS OF HILBERT SPACES AND "ELEMENTARY" QUANTUM SYSTEMS

A. Let $\{h_k\}_{k=1,2,\ldots}$ be a countable sequence of separable Hilbert spaces. By h we denote the set of all sequences $\xi = \{\xi_k\}_{l=1,2,\ldots}$, $\xi_k \in h_k$, subject to the restriction $\Sigma_k \|\xi_k\|^2 < \infty$ and satisfying the linearity properties $\xi + \eta = \{\xi_k + \eta_k\} \in h, \alpha\xi = \{\alpha\xi_k\} \in h, \quad \alpha \in C, h$ is a Hilbert space called a direct sum of separable Hilbert spaces h_k. To deal with a continuous generalization of this concept, we shall follow Refs. 8 and 9.

Let T be a μ-measurable set, where μ is a positive measure. We introduce a field of separable Hilbert spaces labeled by elements of the index set T: $\{h_t, t \in T\}$.

By h we denote the set of all vector valued functions: ξ: $t \to \xi_t, \xi \equiv \{\xi_t \in h_t, \quad t \in T\}$ such that

$$\left| \int_T (\xi_t, \eta_t)\, d\mu(t) \right| < \infty \tag{2.1}$$

for any two functions $\xi, \eta \in h$. In particular, from (2.1) there follows the requirement:

$$\int_T \|\xi_t\|^2 \, d\mu(t) < \infty \quad \forall \xi \in h. \tag{2.2}$$

In addition to (2.1) the linearity is introduced via

$$\xi + \eta = \{\xi_t + \eta_t\}, \quad \alpha\xi = \{\alpha\xi_t\}. \tag{2.3}$$

The scalar product formula in h is given by (2.1):

$$(\xi, \eta) = \int_T (\xi_t, \eta_t)\, d\mu(t);$$

h is a Hilbert space called a direct integral of Hilbert spaces h_t with respect to the measure μ:

$$h = \int_T^{\oplus} h_t \, d\mu(t). \tag{2.4}$$

Let $\varphi(t)$ be a continuous (together with derivatives) function $T \to L(T)$ with the property

$$\int_T |\varphi(t)|^2 \|\xi_t\|^2 \, d\mu(t) < \infty \tag{2.5}$$

for all $\xi \in h$. We shall define a linear operator L_φ in h:

$$L_\varphi: \quad \xi \to L_\varphi \xi = \{\varphi(t)\xi_t\}. \tag{2.6}$$

Here upon $\int_T |\varphi(t)|^2 \, d\mu(t) < \infty$ we have

$$\begin{aligned} \|L_\varphi \xi\|^2 &= \int_T |\varphi(t)|^2 \|\xi_t\|^2 \, d\mu(t) \\ &\leqslant \left[\int_T |\varphi(t)|^2 \, d\mu(t) \right] \cdot \|\xi\|^2 \\ &= \|\varphi\|_\mu^2 \|\xi\|^2, \end{aligned} \tag{2.7}$$

and, consequently, L_φ is a bounded operator in h.

By taking a set of suitable functions $\{\varphi\}$ we get a corresponding set of bounded linear operators in h, which form a weakly closed commuting ring of bounded in h linear operators with unity. Notice that if $L_\varphi, L_\psi \in R$ then also $L_\varphi + L_\psi = L_\delta \in R$ and $L_\varphi \cdot L_\psi = L_\gamma \in R$.

Moreover, if φ is a real function, then it gives rise to a self-adjoint operator in h. If $|\varphi(t)| = 1$, the corresponding operator is a unitary one.

In this way to each direct integral of Hilbert spaces we have assigned a commuting ring of bounded operators. An inverse problem of the decomposition of a Hilbert space into a direct integral, with respect to a given commuting ring, is much more involved,[7–9] albeit useful for the solution of the reduction problem once an operator algebra is given in a Hilbert space. Let us mention that for an example of the sine–Gordon system we have a detailed knowledge about the state space of the system, but no clear understanding of the sine–Gordon field algebra.[2,3,10,11]

B. Let us choose a one-parameter index set $T \in R^1$ and a related one-parameter family $\{h_\lambda, \lambda \in R^1\}$ of Hilbert spaces. Let U_t be a unitary element of the commuting ring in $h = \int^{\oplus} h_\lambda \, d\mu(\lambda)$:

$$U_t \xi = U_t\{|\lambda)\} = \{\exp(i\lambda t) \cdot |\lambda)\} = \xi_t. \tag{2.8}$$

Each h_λ is separable Hilbert space; hence it can be equipped with an orthonormal complete basis system

$$\{|n, \lambda)\}_{n=0,1,\ldots} \quad (n, \lambda | m, \lambda) = \delta_{nm},$$
$$\sum_n |n, \lambda)(n, \lambda| = 1_\lambda, \quad 1_\lambda h_\lambda = h_\lambda. \tag{2.9}$$

A definition of U_t equivalent to (2.8) can be given as follows:

$$U_t = \int_{R^1}^{\oplus} \exp(it\lambda) \sum_n |n, \lambda)(n, \lambda| \, d\mu(\lambda). \tag{2.10}$$

Let us introduce an operator V_s which is *not an element of the commuting ring*:

$$V_s = \int_{R^1}^{\oplus} \sum_n |n, \lambda - s) \exp\left(-s \frac{\partial}{\partial\lambda}\right)(n, \lambda| \, d\mu(\lambda), \tag{2.11}$$

$$\begin{aligned} V_s \xi &= V_s\{|\lambda)\} \\ &= \left\{\sum_n |n, \lambda - s) \exp\left(-s \frac{\partial}{\partial\lambda}\right)(n, \lambda| \sum_m |m, \lambda)(m, \lambda|\lambda)\right\} \\ &= \left\{\sum_n |n, \lambda - s) \exp\left(-s \frac{\partial}{\partial\lambda}\right)(n, \lambda|\lambda)\right\} \\ &= \left\{\sum_n |n, \lambda - s)(n, \lambda - s|\lambda - s)\right\} = \{|\lambda - s)\}. \end{aligned}$$

Consequently,

$$V_s U_t \xi = V_s\{\exp(it\lambda)|\lambda)\}$$
$$= \{\exp(it\lambda)|\lambda - s)\}\exp(-its), \tag{2.12}$$
$$U_t V_s \xi = U_t\{|\lambda - s)\} = \{\exp(it\lambda)|\lambda - s)\},$$

which yields

$$V_s U_t = \exp(-its) U_t V_s, \tag{2.13}$$

i.e., a typical definition of the CCR algebra as given in the Weyl form.

Notice that

$$(V_s \xi, V_s \xi) = \int_{R^1} (\lambda - s|\lambda - s)\, d\mu(\lambda) = (\xi, \xi); \tag{2.14}$$

hence V_s is a unitary operator. Moreover, in addition to $U_t U_{t'} = U_{t+t'}$ we have a semigroup property for V_s as well:

$$V_s V_{s'} \xi = V_s\{|\lambda - s')\} = \{|\lambda - s - s')\} = V_{s+s'}\xi. \tag{2.15}$$

A strong continuity for both U_t and V_s is apparent; hence, by an application of the Stone theorem, the infinitesimal (self-adjoint) generators Q and P of U_t and V_s are recovered,

$$Q = \int_{R^1}^{\oplus} \lambda \sum_n |n, \lambda)(n, \lambda|\, d\mu(\lambda),$$
$$P = \int_{R^1}^{\oplus} \sum_n |n, \lambda)\Big(-i\frac{\partial}{\partial\lambda}\Big)(n, \lambda|\, d\mu(\lambda) \tag{2.16}$$

such that

$$QP\xi = \{\lambda \sum_n |n, \lambda)\Big(-i\frac{\partial}{\partial\lambda}\Big)(n, \lambda|\lambda)\},$$
$$PQ\xi = P\{\lambda|\lambda)\} = \{\sum_n |n, \lambda)(-i)(n, \lambda|\lambda)$$
$$+ \lambda \sum_n |n, \lambda)\Big(-i\frac{\partial}{\partial\lambda}\Big)(n, \lambda|\lambda)\} \tag{2.17}$$
$$= -i\xi + QP\xi,$$

which implies

$$[Q, P]_- \xi = i\xi, \tag{2.18}$$

provided an appropriate domain is chosen to guarantee that both PQ and QP have a meaning in it (see Remark 1 below).

In this way we have demonstrated that, in addition to the commuting ring, the noncommuting pair can be introduced, which generates the CCR algebra representation in the direct integral $h = \int_{R^1}^{\oplus} h_\lambda\, d\mu(\lambda)$ of separable Hilbert spaces. The respective creation and annihilation operators read as follows:

$$A = \frac{1}{\sqrt{2}}(Q + iP)$$
$$= \int_{R^1}^{\oplus} \sum_n |n, \lambda)\frac{1}{\sqrt{2}}\Big(\lambda + \frac{\partial}{\partial\lambda}\Big)(n, \lambda|\, d\mu(\lambda),$$
$$A^* = \frac{1}{\sqrt{2}}(Q - iP) \tag{2.19}$$
$$= \int_{R^1}^{\oplus} \sum_n |n, \lambda)\frac{1}{\sqrt{2}}\Big(\lambda - \frac{\partial}{\partial\lambda}\Big)(n, \lambda|\, d\mu(\lambda),$$
$$[A, A^*]_- \xi = \xi$$

Let us recall that if one intends (as we do) to choose $h_\lambda = \mathrm{IDPS}(|\lambda))$ as determined by starting from (2.10) and (2.11), i.e., by the canonical generators $\{a^*(k), a(k)\}_{k \in R^1}$, then the received generators (2.19) are essentially *new* quantum objects, since we cannot reconstruct A and A^* *solely* in terms of $a^*(k)$ and $a(k)$. This last feature is rather common for these nontrivial field theory models which lead to reducible field algebras; see, for example, Refs. 12–16, but also Refs. 5, 17–20, where an infrared problem for charged particles involves reducible electromagnetic field algebras. The pair (2.19) gives rise to the Schrödinger representation of the CCR algebra, with the (direct integral) vacuum $\Omega \in h = \int_{R^1}^{\oplus} h_\lambda\, d\mu(\lambda)$ selected by the requirements

$$\Omega = \{|\Omega, \lambda)\}, \quad \int_{R^1} \sum_n |(n, \lambda|\Omega, \lambda)|^2\, d\mu(\lambda) < \infty,$$
$$\frac{1}{\sqrt{2}}\Big(\lambda + \frac{\partial}{\partial\lambda}\Big)(n, \lambda|\Omega, \lambda) = 0 \quad \forall n. \tag{2.20}$$

Notice that all h_λ can be inequivalent to the Fock space of the field algebra (2.10). Moreover, an index n in $(n, \lambda|\Omega, \lambda)$ identifies the nth basis vector in h which can be received from $|\lambda)$ by applying the nth function of generators $a^*(k)$ and $a(k)$. It has nothing in common with an index N of

$$\frac{1}{\sqrt{N!}} A^{*N} \Omega := \{|N, \lambda)\}$$
$$= \{\sum_n |n, \lambda)(n, \lambda|N, \lambda)\} \tag{2.21}$$

which corresponds to the Nth excitation level, but in terms of the secondary quanta (2.19).

Remark 1: As is well known,[8] the direct integral of separable Hilbert spaces with respect to any standard measure is a separable Hilbert space again. Hence properties of the representation (2.13) and its generators (2.17) can be understood on the basis of general results described in Ref. (9).

It is not useless to mention that elements of the direct integral space $\mathscr{H}_\mu = \int^{\oplus} \mathrm{IDPS}(|\lambda))\, d\mu(\lambda)$ have the form

$$\psi(f) = \int^{\oplus} f(\lambda)|\psi, \lambda)\, d\mu(\lambda), \quad |\psi, \lambda) \in \mathrm{IDPS}(|\lambda)),$$
$$(\psi, \lambda|\psi, \lambda) = 1 \quad \forall\lambda, \quad (\psi, \lambda|\psi, \lambda') = 0, \ \lambda \neq \lambda',$$
$$(\psi(f), \psi(f)) = \|\psi(f)\|^2 = \int |f(\lambda)|^2\, d\mu(\lambda) < \infty.$$

Notice that we can formally represent the continuous set of orthonormal (in von Neumann's space) vectors $(\psi, \lambda|\psi, \lambda') = 0$, $\lambda \neq \lambda'$, $(\psi, \lambda|\psi, \lambda) = 1$ as generalized vectors associated with $\mathscr{H}_\mu$:

$$\psi_\lambda(\delta) := \int^{\oplus} \delta(\lambda - \lambda')|\psi, \lambda')\, d\mu(\lambda')$$

so that if considered in the topology of $\mathscr{H}_\mu$ we arrive at

$$(\psi_\lambda(\delta), \psi_{\lambda'}(\delta)) = \delta(\lambda - \lambda').$$

It demonstrates how the continuity of the orthonormal set $\{|\lambda)\}$ is lost while passing to the direct integral Hilbert space $\mathscr{H}_\mu$: $\psi_\lambda(\delta)$ is not a Hilbert space vector.

Because of $\psi(f) = \int^{\oplus} f(\lambda)|\psi, \lambda)\, d\mu(\lambda)$, where $f(\lambda)$ is a function $R^1 \to C$ whose modulus is square μ-integrable, the problem of finding a dense in $\mathscr{H}_\mu$ domain $\mathscr{D} \subset \mathscr{D}(P) \cap \mathscr{D}(Q)$ reduces to this of finding an appropriate domain $\mathscr{D}$ in the set of functions $f(\lambda)$. Notice that if $|\psi,\lambda) = \Sigma_n \psi_n |n,\lambda)$, $\psi_n \in C$, $\psi(f) = \int^{\oplus} f(\lambda)|\psi,\lambda)\, d\mu(\lambda)$, then the action of P, Q, PQ, and QP on $\psi(f)$ reduces to

$$QP\psi(f) = \int^{\oplus} \lambda \sum_n |n,\lambda) \cdot \psi_n \Big(-i \frac{\partial f(\lambda)}{\partial \lambda} \Big)\, d\mu(\lambda)$$
$$= \int^{\oplus} \lambda \Big(-i \frac{\partial f}{\partial \lambda} \Big) \cdot |\psi,\lambda)\, d\mu(\lambda)$$
$$PQ\psi(f) = \int^{\oplus} d\mu(\lambda)$$
$$\times \Big\{ -if(\lambda)|\psi,\lambda) + \lambda \Big(-i \frac{\partial f(\lambda)}{\partial \lambda} \Big) |\psi,\lambda) \Big\},$$

i.e., to the well-known Schrödinger representation problem in the Hilbert space of square μ- integrable functions.

Remark 2: The representation (2.10) of the CCR algebra in the separable Hilbert space is unitarily equivalent to the direct sum of Schrödinger representations. Consequently, the vector Ω of (2.20) would be unique in the irreducible case only. However, this is not the case in $\mathscr{H}_\mu$. Let us define

$$\Omega = \Omega_n := \int^{\oplus} f_0(\lambda)|n,\lambda)\, d\mu(\lambda),$$

where $|n,\lambda)$ is the nth basis vector in IDPS $(|\lambda))$ and $f_0(\lambda)$ satisfies: $(1/\sqrt{2})(\lambda + \partial/\partial\lambda) f_0(\lambda) = 0$. Then

$$A\Omega_n = \int^{\oplus} \Big[\frac{1}{\sqrt{2}} \Big(\lambda + \frac{\partial}{\partial \lambda} \Big) f_0(\lambda) \Big] |n,\lambda)$$
$$\times d\mu(\lambda) = 0 \quad \forall n = 0, 1, \cdots$$

and

$$|N,n) = \frac{1}{\sqrt{N!}} A^{*n} \Omega_n = \int^{\oplus} \frac{1}{\sqrt{N!}} \Big\{ \Big[\frac{1}{\sqrt{2}} \Big(\lambda - \frac{\partial}{\partial \lambda} \Big) \Big]^N f_0(\lambda) \Big\}$$
$$\times |n,\lambda)\, d\mu(\lambda), (N,n|M,m) = \delta_{nm}\delta_{NM}.$$

Hence the representation (2.13) is not irreducible in $\mathscr{H}_\mu$.

C. Since we have in $h = \int^{\oplus}_{R^1} h_\lambda\, d\mu(\lambda)$ a canonical pair P, Q, it seems rather natural to follow a conventional quantum route, and to search for a Hamiltonian system in h. In an abstract scheme, there is no natural choice of the Hamiltonian. However, for a particular quantization procedure for the classical (e.g., sine–Gordon) model H can be determined once time-dependent trajectories are established in the set of parameters $\lambda \in R^1$. This is the case when the (underlying) classical dynamics is taken into account.

Let us consider a family of sine–Gordon 1-solitons, each one with a fixed a value, but differing in the choice of $q = q_a = \lambda$ of (1.8). In fact,

$$\lambda = \lambda(t) = v_a(t_0 - t) = \lambda - v_a t = \lambda + \dot{\lambda} t, \tag{2.22}$$

and $\lambda = v_a t_0 \in R^1$ is quite arbitrary. We look for the time development generator H such that

$$Q\xi = \{\lambda|\lambda)\} \to Q(t)\xi$$
$$= \exp(iHt) Q \exp(-iHt)\xi$$
$$= \{\lambda'(\lambda,t)|\lambda'(\lambda,t))\}, \quad \lambda'(\lambda,t) = \lambda + \dot{\lambda} t. \tag{2.23}$$

Notice that for the infinitesimal time variations we have

$$Q(\Delta t) \cong Q + \Delta t \dot{Q}, \quad \dot{Q} = -i[Q,H]_-,$$
$$Q(\Delta t)\xi \cong \{(\lambda + \Delta t \dot{\lambda})|\lambda + \dot{\lambda}\Delta t)\}, \tag{2.24}$$

and, consequently, because of (2.23),

$$\ddot{Q} = -i[\dot{Q},H]_- = 0 \tag{2.25}$$

so that the form (2.24) persists for all $t \in R^1$ in (2.23). Notice that $H = H(\dot{Q})$ solves (2.25).

We shall deduce an explicit form of $H = H(\dot{Q})$ by following the idea of Ref. 21, which, if appropriately modified, can be applied to our case.

For the translation operator P we have the formula (2.17). Let us rewrite it in the form

$$P = \int^{\oplus}_{R^1} d\mu(\lambda)\, T_{10}(\lambda), \quad T_{10}(\lambda) = T_{01}(\lambda), \tag{2.26}$$
$$= \int^{\oplus}_{R^1} \sum_n |n,\lambda) \Big(i \frac{\partial}{\partial \lambda} \Big) (n,\lambda|\, d\mu(\lambda).$$

Analogously for H

$$H = \int^{\oplus}_{R^1} d\mu(\lambda)\, T_{00}(\lambda) \tag{2.27}$$

with $T_{00}(\lambda)$ being still unspecified. Let us now introduce a symmetric stress energy tensor $T_{\mu\nu}(\lambda)$ so that the generator of Lorentz transformations in h can be introduced as follows:

$$M_{10} = \int^{\oplus}_{R^1} d\mu(\lambda)\{\lambda\, T_{00}(\lambda) - tT_{10}(\lambda)\} \tag{2.28}$$
$$= HQ - tP.$$

Upon a standard requirement

$$0 = \frac{dM_{10}}{dt} = \frac{\partial M_{10}}{\partial t} - i[M_{10},H]_-$$
$$= -P + H\dot{Q} - t\dot{P}, \tag{2.29}$$

provided with a momentum conservation demand $\dot{P} = 0$, we arrive at the identity

$$P = H\dot{Q}, \tag{2.30}$$

which yields

$$i = [Q,P]_- = [Q,\dot{Q}]_- H + i\dot{Q}^2. \tag{2.31}$$

Recall now that $H = H(\dot{Q})$; hence formally one can introduce the notion of $\partial H/\partial \dot{Q}$. Suppose that in the domain of H there exists at least one vector on which the following two operator identities hold true (the nature of the constraint will be investigated below):

$$\Big[Q, \frac{\partial H}{\partial \dot{Q}} \Big]_- = i,$$
$$2i\dot{Q} = [Q,2H]_- = \Big[Q, \dot{Q} \frac{\partial H}{\partial \dot{Q}} \Big]_- \tag{2.32}$$

In (2.32) for some vectors the following holds :

$$2i\dot{Q} = [Q,\dot{Q}]_- \frac{\partial H}{\partial \dot{Q}} + i\dot{Q}, \tag{2.33}$$

which by taking into account (2.31) leads to the conclusion that, for such vectors,

$$\frac{\partial H}{\partial \dot{Q}} = \frac{\dot{Q}}{1 - \dot{Q}} H, \tag{2.34}$$

which is satisfied by

$$H = M/(1-\dot{Q}^2)^{1/2}, \tag{2.35}$$

with M being an integration constant, $M \in R^1$. Furthermore,

$$P = M\dot{Q}/(1-\dot{Q}^2)^{1/2} \tag{2.36}$$

and

$$[Q,\dot{Q}]_- = i(1-\dot{Q}^2)H^{-1} = (i/M)(1-\dot{Q}^2)^{3/2}. \tag{2.37}$$

Because of (2.36) a conventional relativistic formula for H (in $1+1$ dimensions) follows:

$$H = (P^2+M^2)^{1/2}. \tag{2.38}$$

Due to $\dot{Q} = \dot{Q}(P)$ and $H = H(P)$, the representation of P in $h = \int_{R^1}^{\oplus} h_\lambda \, d\mu(\lambda)$ as given by (2.17) and (2.26) leads to

$$\dot{Q} = \int_{R^1}^{\oplus} \sum_n |n,\lambda) \frac{i\partial}{(M^2-\partial^2)^{1/2}} (n,\lambda \mid d\mu(\lambda), \tag{2.39}$$

$$H = \int_{R^1}^{\oplus} \sum_n |n,\lambda)(M^2-\partial^2)^{1/2}(n,\lambda \mid d\mu(\lambda).$$

D. Let us now analyze the constraints (2.32) which diminish the arbitrariness in the choice of H, by demanding the existence of suitable vectors in the domain. By making use of the first constraint (2.32) we get formally

$$\frac{\partial H}{\partial \dot{Q}} = \frac{M\dot{Q}}{(1-\dot{Q}^2)^{1/2}} = \frac{PH^2}{M^2} \tag{2.40}$$

with $[P,H]_- = i\dot{P} = 0$. Hence the second one reads

$$\left[Q, \frac{\partial H}{\partial \dot{Q}}\right]_- |\psi) = \frac{1}{M^2}[Q, PH^2]_- |\psi)$$
$$= \left\{ \frac{2P^2}{M^2} + i\frac{H^2}{M^2} \right\} |\psi) = i|\psi), \tag{2.41}$$

where $[\dot{Q}, H]_- = 0$ is taken into account. From (2.41) it follows that $|\psi)$ is a common eigenvector of both P and H:

$$P|\psi) = 0, \quad H|\psi) = \sqrt{P^2+M^2}|\psi) = M|\psi). \tag{2.42}$$

Then we find

$$2i\dot{Q} = \left[Q, \dot{Q}\frac{\partial H}{\partial \dot{Q}}\right]_- = \frac{1}{M^2}[Q, P^2H]_-, \tag{2.43}$$

and hence

$$(P^2+2H^2)/2M^2\,\dot{Q}|\psi) = \dot{Q}|\psi)$$
$$\Rightarrow \dot{Q}|\psi) = 0 \ \text{or} \ [\dot{Q},H]_- = 0 = [\dot{Q},P]_-, \tag{2.44}$$

which may impose a restriction on $\dot{Q}$ if applied to $|\psi)$. Consequently, (2.32) is equivalent to

$$H|\psi) = M|\psi), \quad P|\psi) = 0, \tag{2.45}$$

and thus the parameter M corresponds to the rest mass of the elementary quantum system associated with the direct integral $h = \int_{R^1}^{\oplus} h_\lambda \, d\mu(\lambda)$ of separable Hilbert spaces. The main problem now is to find M while maintaining consistency with the classical field equations (e.g., the sine-Gordon one) which underlies the whole derivation of (2.42). Recall that M appears in (2.35) as an integration constant, but is not at all constrained to be a c-number. The more natural requirement is that M belongs to the commutant of the $\{P,Q\}$ C^*-algebra. Since the $\{P,Q\}$ pair arises in the reducible representation of the primary scalar field, (2.10) algebra, both P and Q do commute with the CCR algebra generators $\{a^*(k), a(k)\}_{k\in R^1}$. If M is a c-number then H does commute also, as being solely constructed in terms of P and Q. However, an integration procedure leading to (2.35) does not exclude the fact that $M = M(a^*, a)$ is an operator element of the neutral scalar field algebra. This route has been followed in Ref. (22); however, the authors start from another assumption: (1) about the Hilbert space structure, which is a direct product of the Fock space for collective modes and the Hilbert space for the primary quantum field; (2) about the existence of the particle number operator for the neutral scalar field in its soliton sectors.

Let us assume to have a rest frame vector $|\psi)$ of (2.45), and let us further assume that $H|\psi) = M_0|\psi)$, $M_0 \in R^1$. By applying to $|\psi)$ polynomials in $a^*(k)$, $a(k)$ and then making a Hilbert space closure of the set obtained, we arrive at the previously introduced notion of IDPS $(|\psi))$. Consequently, if one has any Hamiltonian operator $H_0(a^*, a) = H_0$ generating a unitary in time evolution *in* IDPS$(|\psi))$, it can be safely added to M_0, thus giving rise to the following modification of $H = \sqrt{P^2+M_0^2}$:

$$H = \sqrt{P^2+(M_0+H_0)^2}, \quad H_0 = H_0(a^*, a). \tag{2.46}$$

Recall that $\hbar = c = 1$. In Ref. (22) H_0 is supposed to be a free neutral scalar field Hamiltonian, which, however, has no eigenvectors outside of the Fock space IDPS $(|0))$. Our IDPS$(|\psi))$ is inequivalent to IDPS $(|0))$. However, H_0 can be regarded to be (if specialized to our example) the quantum sine-Gordon Hamiltonian, constrained to the particular soliton sector. Then we can expect[2,3,10,11] that the conventional sine-Gordon spectrum and eigenvectors can be produced in IDPS $(|\psi))$.

Remark: A construction of coherent soliton states for the sine-Gordon system is motivated by the following assumptions: If $\hat{\phi}(x,t)$ is an interacting sine-Gordon field, then it admits the Haag type expansion in terms of (asymptotic-like but *not* asymptotic at all) free mass m neutral scalar field generators $a^*(k)$ and $a(k)$, i.e., $\hat{\phi}(x,t) = \phi(a^*, a, x, t)$. Then, a coherent state expectation value of $\hat{\phi}$ in the tree (zero loop) approximation,

$$(\alpha|{:}\hat{\phi}(a^*, a, x, t){:}|\alpha) = (0|{:}\hat{\phi}(a^*+\overline{\alpha}, a+\alpha, x, t){:}|0)$$
$$= \phi(\overline{\alpha}, \alpha, x, t) = \phi_{cl}(x,t), \tag{2.47}$$

should allow us to restore both the classical sine-Gordon field, its equations of motion, and the Hamiltonian

$$(\alpha|{:}H(\hat{\phi}){:}|\alpha) = (\alpha|{:}H(\hat{\phi})(a^*, a){:}|\alpha)$$
$$= (0|{:}H(\phi)(a^*+\overline{\alpha}, a+\alpha){:}|0)$$
$$= H(\phi)(\overline{\alpha}, \alpha). \tag{2.48}$$

In the above, $|0)$ is the Fock state for the $\{a^*, a\}$ field algebra.

Consequently, the quantum sine-Gordon Hamiltonian, consistent with the above tree approximation mappings, should implement an evolution unitary in time in each of the irreducibility sectors for the $\{a^*, a\}$ field algebra, i.e., in each IDPS $(|\alpha))$. Hence it is a rather natural choice to identify $H_0 = H_0(a^*, a)$ of (2.46) with the sine-Gordon Hamiltonian, while constructed *solely* in terms of $a^*(k)$ and $a(k)$.

Let us add that the original Haag expansions in terms of free asymptotic fields of the model are of the perturbative nature. In the above we admit expansions which formally have the Haag form (infinite power series), but (1) are nonperturbative and (2) can be defined with respect to the free fields which are not asymptotic ones for the model under consideration ("confinement").

The choice of the c-number constant M_0 is to some extent arbitrary, but we are motivated by the fact that the sine-Gordon parameter m of (2.4) gives rise to the classical 1-soliton mass equal $8m$. The 1-soliton momentum $k_a = 8mv_a/\sqrt{1-v_a^2}$ is in a relativistic relationship with $8m$: $E = \sqrt{(8m)^2 + k^2}$. Hence an identification $M_0 = 8m$ is quite natural.

Suppose now that we have solved the spectral problem for H_0 in IDPS $(|\psi))$. If $|E, \psi)$ is an eigenvector of H_0 in IDPS$(|\psi))$, then it is also an eigenvector of H,

$$H\,|E, \psi) = \sqrt{P^2 + (8m+E)^2}\;|E, \psi). \tag{2.49}$$

Hence we are in principle capable of producing quantum corrections (due to the primary field excitations built over the extended particle state) to the 1-soliton mass $8m$. If we adopt the discrete (bound state, WKB) spectrum of the sine-Gordon Hamiltonian: $\{E_i\}_{i=1,2,\dots}$ then the secondary Hamiltonian $H = H(H_0, P, Q) = H(a^*, a, P, Q)$ has a discrete mass spectrum with $H_0|0, \psi) = 0$ corresponding to the $H\,|\psi) = 8m|\psi)$ equation.

Let us emphasize that in contrast to Ref. (22), but in agreement with the observations of Ref. (6), the asymptotic problem for the neutral scalar (sine-Gordon) field cannot be solved in IDPS $(|\psi))$. There is no unitary mappings (*solely* in the scalar field algebra) of the soliton Hilbert space IDPS $(|\psi))$ into a Fock space IDPS $(|0))$. *Consequently, we have a typical "confinement" of the quantum scalar field constituents of the extended particle state (soliton) generating IDPS* $(|\psi))$.

On the other hand, a folk-lore statement is to attribute the notion of a quantum particle to an elementary quantum system (i.e., an irreducible representation of the CCR algebra, which the pair P, Q does indeed generate). Consequently, for each fixed a value, the 1-soliton Hilbert space $h_a = \int_{R^1}^{\oplus} h_{a\lambda}\, d\mu(\lambda)$ can be interpreted as the carrier state space for a quantum particle, which though "elementary" is still an infinite constituent object, see, e.g., in this connection Ref. (23). The constituents can never be seen in the conventional (Fock space) sense, due to the above-mentioned "confinement" property.

Let us emphasize a paradoxical situation: a folk-lore understanding of P, Q is that an elementary particle is structureless. Quite the contrary, our derivation of P, Q is based on the rich infinite constituent structure underlying the construction of the state space for the sine-Gordon field, while soliton sectors are taken into account.

3. "ELEMENTARY" TWO LEVEL SYSTEMS AND THEIR INTERACTION IN 1 + 1 DIMENSIONS

A. The modulus $|a|$ of the 1-soliton parameter a is relevant to the velocity v_a and $v_a = v_{-a}$. Here sgn a is relevant to the topological invariant value:

$$R = \frac{1}{4\pi}\int_{R^1} \frac{\partial\phi}{\partial x}\,dx, \tag{3.1}$$

which for the ath 1-soliton reads $R = R(a) = R(\pm|a|) = \pm\frac{1}{2}$. We assume $a \neq 0$ and then compose a direct sum of the $\pm a$th soliton Hilbert spaces:

$$h(a) = h_a \oplus h_{-a}, \quad a \in R^+. \tag{3.2}$$

One can observe (see, e.g., Ref. 3) that a classical 1-soliton momentum value reads $k = 8m(a^2-1)/2|a|$ and hence the a label can be replaced by a joint (R, k) label: $h_{\pm a} \to h_{Rk}$ so that

$$h(a) = h_k = \oplus_{R=\pm 1/2}\, h_{Rk}. \tag{3.3}$$

It means that now we are able to give a quantum meaning to the topological invariant R. For this purpose, we shall define the spin-$\frac{1}{2}$ SU(2) group raising and lowering operators, as has been previously done in Ref. (3), in a slightly different context:

$$\sigma_k^+ = \int_{R^1}^{\oplus} \sum_n |n, k, +, \lambda)(n, -, k, \lambda|\, d\mu(\lambda),$$
$$\sigma_k^- = \int_{R^1}^{\oplus} \sum_n |n, k, -, \lambda)(n, +, k, \lambda|\, d\mu(\lambda). \tag{3.4}$$

Because of $(n, a, \lambda|m, -a, \lambda) = 0\ \ \forall n, m, \lambda$, we find immediately that

$$\sigma_k^+ \xi_{\mp k} = \xi_{\pm k}, \ \xi_{+k} \in h_{+k}, \xi_{-k} \in h_{-k},$$
$$(\sigma_k^\pm)^2 = 0, \tag{3.5}$$

i.e.,

$$\sigma_k^3 \xi_{\pm k} = (\pm\tfrac{1}{2})\xi_{\pm k}, \ \sigma_k^3 = (-\tfrac{1}{2})1_k + \sigma_k^+\sigma_k^-,$$
$$1_k = [\sigma_k^+, \sigma_k^-]_+ \tag{3.6}$$

In accordance with Ref. 2 and 3 the N-soliton coherent state (without breathers) exhibits the following parametrization:

$$|k_1, \ R_1, \lambda_1, \dots, k_N, R_N, \lambda_N) = |\underline{k}, \underline{R}, \underline{\lambda})_N,$$
$$k_1 < k_2 < \cdots < k_N \tag{3.7}$$

where, due to the classical momentum ordering, R_i is a topological invariant of the k_ith (asymptotic) 1-soliton.

A generalization of the previous direct integral procedure reads now as follows:

$$|\underline{k}, \underline{R}, \underline{\lambda})_N \to \text{IDPS}(|\underline{k}, \underline{R}, \underline{\lambda})_N),$$
$$h_{k_1R_1\cdots k_NR_N} = h^N_{\underline{k}\underline{R}} = \int_{R^1}^{\oplus} d\mu(\lambda_1)\cdots\int_{R^1}^{\oplus} d\mu(\lambda_N)$$
$$\times \text{IDPS}(|\underline{k}, \ \underline{R}, \ \underline{\lambda})_N). \tag{3.8}$$

By our choice of the sharply ordered momentum sequence $k_1 < k_2 < \cdots < k_N$ we have guaranteed a fulfillment of the "classical Pauli exclusion principle" regarding that neither a_i in the parametric sequence $\{a_1, \dots, a_N\}$ can appear more than once. By composing a direct sum with respect to all possible configurations of $\pm$'s in the sequence $\{R_1, \dots, R_N\}$ we finally arrive at the N-soliton analog of the two-level Hilbert space h_k of (3.3):

$$h_{k_1\cdots k_N} = \oplus_{\text{conf}(\underline{R})}\, h_{k_1R_1\cdots k_NR_N}. \tag{3.9}$$

The corresponding set of spin-$\frac{1}{2}$SU(2)^N generators can be constructed by adopting the idea of Ref. (3) to our (direct integral) case:

$$\sigma_i^+ = \sigma_{ik_1\dots k_N}^+ = \sum_{\mathrm{conf}(\underline{R})}^{(i)} \int_{R^1}^{\oplus} \cdots \int_{R^1}^{\oplus} d\mu(\lambda_1)\cdots d\mu(\lambda_N) \times |k_1, R_1, \lambda_1, \dots, R_i = +\tfrac{1}{2}, \dots, k_N, R_N, \lambda_N)(k_1 R_1, \lambda_1, \dots, R_i = -\tfrac{1}{2}, \dots, k_N, R_N, \lambda_N|,$$

$$\sigma_i^- = \sum_{\mathrm{conf}(\underline{R})}^{(i)} \int_{R^1}^{\oplus} \cdots \int_{R^1}^{\oplus} d\mu(\lambda_1)\cdots d\mu(\lambda_N) \times |k_1, R_1, \lambda_1, \dots, R_i = -\tfrac{1}{2}, \dots, k_N, R_N, \lambda_N) \times (k_1, R_1, \lambda_1, \dots, R_i = +\tfrac{1}{2}, \dots, k_N, R_N, \lambda_N|, \tag{3.10}$$

where $\Sigma^{(i)}_{\mathrm{conf}(\underline{R})}$ means that we perform summations over all admissible configurations of $\{R_1,\dots,R_N\}$ under an assumption that R_i is left untouched. Obviously,

$$[\sigma_i^+, \sigma_j^+]_- = 0 = [\sigma_i^-, \sigma_j^-]_-,$$
$$[\sigma_i^-, \sigma_j^+]_- 0, \quad i \neq j,$$
$$[\sigma_i^+, \sigma_i^-]_+ = 1_{k_1 \cdots k_N} = 1_k^N, \tag{3.11}$$
$$1_k^N h_k^N = h_k^N.$$

All mappings changing a configuration of $\{R_1,\dots,R_N\}$ at a fixed choice of $k_1 < \dots < k_N$ leave the Hilbert space $h_{k_1 k_2 \dots k_N}$ invariant. Since with such configuration-to-configuration mappings we have automatically associated the spin-$\frac{1}{2}$ xyz Heisenberg model Hamiltonian (see Refs. 2 and 3), $h_{k_1 k_2 \cdots k_N}$ can be viewed as the carrier Hilbert space for a system of N interacting spins $\frac{1}{2}$ on a linear lattice with sites labeled by $k_1,\dots,k_N$:

$$H_{xyz} = -\sum_{a=1}^{3} \sum_{i=1}^{N} \mathscr{J}_a \sigma_i^a \sigma_{i+1}^a, \tag{3.12}$$

$$\sigma_i^1 = \frac{1}{\sqrt{2}}(\sigma_i^+ + \sigma_i^-), \quad \sigma_i^2 = \frac{i}{\sqrt{2}}(\sigma_i^+ - \sigma_i^-),$$
$$\sigma_i^3 = -\tfrac{1}{2} + \sigma_i^+ \sigma_i^-.$$

Needless to say, the spectrum of the weakly anisotropic xyz model, while going to continuum is mapped into the WKB sine-Gordon/massive Thirring model spectrum. Hence a close relationship with the quantum sine-Gordon model still persists despite the direct integrals involved. Be aware, however, that (3.12) is not a continuum but an (irregular) lattice Hamiltonian.

An explicit construction of the eigenvectors of (3.12), which can be easily reproduced in each $h_{k_1 k_2 \cdots k_N}$, was given together with the derivation of the eigenvalues in Ref. 24; see also Ref. 11. Let us emphasize that the spin-$\frac{1}{2}$ notion arising in the above is of the purely topological (classical topological invariant) origin; hence, with H_{xyz} in mind, we should say rather about an interacting system of topological spins $\frac{1}{2}$.

B. Before, we have introduced the Hamiltonian system in $h_{k_1 \cdots k_N}$ by allowing mappings among configurations $\{R_1,\dots,R_N\}$ at a fixed choice of $k_1 < \cdots < k_N$, which resulted in the spin $\frac{1}{2}$ H_{xyz} operator as the appropriate generator of time translations.

We can, however, define a more restrictive set of mappings, by following the route of Refs. 25 and 26, which aims at the description of particle scattering in $1 + 1$ dimensions, for systems subject to the infinite set of conservation laws. Namely, let us adopt the following scattering principles: (1) absence of particle production; (2) equality of the sets of initial and final momenta $\{p_1,\dots,p_N\} = \{p'_1,\dots,p'_N\}$; (3) if there is more than one type of particles involved, then the numbers n_i of particles of the same type are unchanged in the scattering process.

Now let us make an assumption that the classical soliton momentum indices appearing in $h_{k_1 \cdots k_N}$ are the "particle momenta" of the would-be scattered extended quantum systems. Then $h_{k_1 \cdots k_N}$ appears as a substitute (albeit completely different in its structure) of the conventional N-particle Fock space sector: $\otimes_{i=1}{}^N (h_i)$. Instead of speaking about different (topological $R = \pm\frac{1}{2}$) spin projections for a given "particle," we can view them as completely distinct species, having thus introduced a soliton and antisoliton as the two-"particle" types. Due to requirement (3), the number of soliton or antisoliton labels remains unchanged in any scattering, and, moreover, by (1)–(3) we find that[25] the S operator acts on the states in $h_{k_1 \cdots k_N}$ by a possible permutation of momentum labels. In terms of soliton states

$$S: \quad |k_1, R_1 \dots k_N, R_N) \rightarrow |k_{\pi(1)}, R_1, \dots, k_{\pi(N)}, R_N) = |k_1, R_{\pi(1)}, \dots, k_N, R_{\pi(N)}), \tag{3.13}$$

which means that the scattering process is described by an "exchange" of topological charges at a fixed $k_1 < \cdots < k_N$ sequence, which is subject to our requirements (3). Here

$$h_{k_1 \cdots k_N} \ni |f, k_1, \dots, k_N) = |f):$$
$$= \sum_{\{R_1 \cdots R_N\}} f_{R_1 \cdots R_N} |k_1, R_1, \dots, k_N, R_N) \tag{3.14}$$

so that

$$S: |f) \rightarrow \sum_{\{\underline{R}\}} f_{R_1 \cdots R_N} |k_1, R_{\pi(1)}, \dots, k_N, R_{\pi(N)}) = \sum_{\{\underline{R}\}} f_{R_{\pi(1)} \cdots R_{\pi(N)}} |k_1, R_1, \dots, k_N, R_N) = |f') \tag{3.15}$$

with the normalization

$$(k_1, R_1, \dots, k_N, R_N | k_1, R'_1, \dots, k_N, R'_1) = \delta_{R_1 R'_1} \cdots \delta_{R_N R'_N}. \tag{3.16}$$

Because of (1)–(3) the S operator should be factorized if defined on states in $h_{k_1 \cdots k_N}$ (see Refs. 25 and 26):

$$(g'|f) = (g, k_1 \dots k_N | S | f, k_1 \dots k_N) = (g, k_1, \dots, k_N | \prod_{1 \leq i < j \leq N} S(k_i, k_j) | f, k_1 \dots, k_N), \tag{3.17}$$

where, since the two-particle S operators $S(k_i, k_j)$ in general fail to commute, it is necessary to specify the order of factors occuring in (3.17). A possible choice which corresponds to the ordering of momenta $k_1 < \dots < k_N$ (see Ref. 25) is

$$S_{N-2,N} \cdots (S_{2N} \cdots S_{23})(S_{1N} \cdots S_{12}) = \prod_{1 \leq i < j \leq N} S_{ij}, \tag{3.18}$$

where

$$S_{ij}S_{kl} = S_{kl}S_{ij},$$
$$S_{ij}S_{ik}S_{jk} = S_{jk}S_{ik}S_{ij}, \tag{3.19}$$

i, j, k, l all unequal. Let us once more recall the sharp momentum ordering we use in the above.

Suppose for a while that $h_{k_1 \ldots k_N}$ is formally constructed without any reference to the $k_1 < \ldots < k_N$ demand. Then let us notice that the antisymmetrizing symbol

$$\sigma(k_1,\ldots,k_N) = \prod_{1 \le i < j \le N} [\Theta(k_i - k_j) - \Theta(k_j - k_i)] \tag{3.20}$$

with $\Theta(p) = 1$ for $p > 0$ and $\Theta(p) = 0$ for $p \le 0$ equals 1 except for the case of an odd permutation when it equals -1, and for the case of coinciding momenta ($k_i = k_j$ for some choice of i, j) when it equals 0.

Therefore, σ^2 is nonnegative and equals 1 for any choice of noncoinciding k's in the momentum sequence. It implies that

$$\sigma^2(1-\sigma^2) = 0, \quad \sigma^3 = \sigma,$$
$$\begin{aligned} h_{k_1 \cdots k_N} &= \sigma^2 h_{k_1 \cdots k_N} \oplus (1-\sigma^2) h_{k_1 \cdots k_N} \\ &= h^1_{k_1 \cdots k_N} \oplus h^2_{k_1 \cdots k_N} \end{aligned} \tag{3.21}$$

and $h^1_{k_1 \cdots k_N}$ can always be written in the form (3.14) with $k_1 < \cdots < k_N$. The analysis of Ref. 27 (see also Ref. 11) shows that $h^1_{k_1 \cdots k_N}$ includes Fermi states of the Bose system defined in $h_{k_1 \cdots k_N}$ with k's unrestricted by (3.20).

Because the soliton Hilbert space of (3.14) respects the sharp ordering of momenta, we have $\sigma^2 h_{k_1 \cdots k_N} = h^1_{k_1 \cdots k_N} = h_{k_1 \cdots k_N}$. Consequently, the soliton "particle" scattering is the same as the scattering of particles subject to the Pauli exclusion principle.

Since $R = +\frac{1}{2} \to R = -\frac{1}{2}$ implies the soliton $\to$ antisoliton, i.e., the "particle" $\to$ "antiparticle" mapping, the general features of the scattering (3.15) should coincide with these observed for the conventional Fermi massive Thirring model, irrespective of the fact that we make the whole of the construction for the Bose system only.

The state $|f, k_1 \ldots, k_N)$ is a normalized vector for any N; hence the two-particle S matrices S_{ij} can be studied in more detail:

$$\begin{aligned} &S_{ij} = T_{ij} + R_{ij}, \\ &(f, k_1, k_2|\, T_{ij}\,|f, k_1, k_2) = t_{ij}, \\ &(f, k_2, k_4|R_{ij}\,|f, k_2, k_1) = r_{ij}. \end{aligned} \tag{3.22}$$

In particular, if we consider a particle–antiparticle (i.e., soliton–antisoliton) pair, we have only four functions occurring in (3.22): $t_{\bar f \bar f}$, $t_{f \bar f}$, $t_{\bar f \bar f}$, $r_{f \bar f}$, which are still related by crossing, symmetry and unitarity. If to introduce the rapidity Θ,

$$\begin{aligned} &k_i^0 = m_i \cosh \Theta_i, \quad k_i^1 = m_i \sinh \Theta_i, \\ &(k_i + k_j)^2 = m_i^2 + m_j^2 + 2m_i m_j \cosh \Theta_{ij}, \; \Theta_{ij} = \Theta_i - \Theta_j, \\ &(k_i - k_j)^2 = m_j^2 + m_i^2 + 2m_i m_j \cosh(i\pi - \Theta_{ij}), \end{aligned} \tag{3.23}$$

then, by making use of the crossing and unitarity relations, one arrives[25] at

$$\begin{aligned} &t = t_{\bar f \bar f}, \quad r = r_{f \bar f}, \quad u = t_{ff} = t_{\bar f \bar f} \\ &t(\Theta) = u(i\pi - \Theta), \quad r(\Theta) = r(i\pi - \Theta), \end{aligned} \tag{3.24}$$

and

$$r^2(\Theta) = t^2(\Theta)\{1 - [t(-\Theta)t(\Theta)^{-1}]\}, \tag{3.25}$$

which proves that the scattering is described by only one independent function, $t(\Theta)$, say.

A particular choice of $t(\Theta)$ determines the quantum field theory model which governs the scattering process. If we, for example, decide to follow Refs. 25 and 28, then the choice of the transmission amplitude:

$$t(\Theta) = (-1)^\lambda \prod_{k=1}^{\lambda} \frac{e^{\Theta - i\pi k/\lambda} + 1}{e^{\Theta} + e^{-i\pi k/\lambda}}, \; \lambda \text{ integer}, \tag{3.26}$$

corresponds to the so called soliton–antisoliton scattering, and $t(\Theta)$ has poles at $\Theta_k = i\pi(1 - k/\lambda)$ which is in relation with the WKB sine-Gordon spectrum: $m_k = 2m \sin \pi k/2\lambda$ provided that $\lambda = 8\pi/\gamma = 8\pi/\beta^2(1 - \beta^2/8\pi)$ and the coupling term of the sine-Gordon equation reads $(m^2/\beta)\sin \beta\phi$. Due to the fact that λ is chosen to be an integer, the reflection amplitude $r(\Theta)$ vanishes.

Consequently, we have found it possible to introduce in the soliton Hilbert space $h_{k_1 \cdots k_N}$, a concept of what is usually known as the "soliton-antisoliton scattering" for the quantized sine-Gordon system. If we make one more identification, $\lambda = 1 + 2g/\pi$, the amplitude (3.26) describes the transmission phenomena in the massive Thirring model with the coupling constant g.

4. ON REDUCIBLE FIELD ALGEBRAS IN QUANTUM ELECTRODYNAMICS

A. One says that a representation $\pi(\mathscr{U})$ of some (field) algebra $\mathscr{U}$ is reducible in a Hilbert space $\mathscr{H}$, if there is at least one nontrivial (i.e., different from zero and unity) operator R, which (1) commutes in $\mathscr{H}$ with the whole of $\pi(\mathscr{U})$ and (2) is not an element of $\pi(\mathscr{U})$

In terms of quantum fields, one may assume that $\mathscr{U}$ is generated by (say) the CCR algebra creation–annihilation operators $\{a^*, a\}$ of some Bose field. Then the weakest statement about R is that it cannot be *solely* constructed in terms of $\{a^*, a\}$[12–14]: Obviously, for an irreducible representation, any element of $\pi(\mathscr{U})$ can be given as a function of $\{a^*, a\}$. A particular example of such a situation was considered before in Sec. 2. One of the observations of Ref. 29 was that a quantized solution of the free massive Dirac equation, satisfying the usual CAR algebra commutation relations can be reconstructed in terms of the two potential Maxwell (Bose) operators. Then, upon imposing suitable constraints, one can arrive at the free electromagnetic field (the Bose quantized Coulomb gauge potential $\hat{A}_\mu$) reconstruction of $\hat\psi(x)$, so that $\hat\psi(x) = \psi(\hat{A}_\mu)(x)$.

If by $\mathscr{U}(\hat{A})$ we denote the electromagnetic free field algebra, then $\hat\psi$ is necessarily an element of it, as all its spinor components can be *solely* expressed in terms of $\hat{A}_\mu$. However, then $\hat\psi$ does not in any case commute with $\hat{A}_\mu$: $[\hat\psi(x), \hat{A}_\mu(y)]_- \neq 0$,[5] while the mutual commutativity requirement $[\hat\psi(x), \hat{A}_\mu(y)]_- = 0$ lies at the foundations of the conventional (perturbative) investigations in the infrared quantum electrodynamics.[17–19] On the other hand, this assumption is not at all necessary for the understanding of the QED Hilbert space structure, as described in Ref. 20. At this point, it is worthwhile to recall the fundamental conjecture of Ref. 20 that the asymptotic electron (Dirac) field must not locally commute with the electromagnetic field. Consequently, even if one follows the traditional route by starting from the mutually commuting free fields $[\hat{\psi}(x), \hat{A}_\mu(y)]_- = 0$, the final goal must be a construction of the asymptotic free fields (arising as weak limits of the interpolating ones) which do fail to mutually commute.

In the Kulish–Faddeev–Zwanziger approach,[17,19] one incorporates the radiation/Coulomb phase operator so as to arrive at the asymptotic electron field $\hat{\psi}_{\rm as} = \psi_{\rm as}(\hat{\psi}, \hat{\psi}, \hat{A}_\mu)$, which is a nonlocal function of the mutually commuting plane wave solutions to respective free field equations: nonasympotic electron $\hat{\psi}(x)$ and photon $\hat{A}_\mu$ ones (see also Ref. 30).

While constructing $\hat{\psi}_{\rm as}$, one believes[19] that an appropriate Hilbert space is the direct product of the free field representation space for photons and the traditional Fock representation space for electrons. Let us mention that this kind of assumption has been made by Matsumoto, *et al.*[21,22] to give account of the collective degrees of freedom for soliton fields.

On the other hand, it is well known that the free electromagnetic field algebra has a highly reducible representation in the Hilbert space of infrared states,[6,18,20,30] hence quite a natural way of getting the mutually commuting free fields $\{\hat{\psi}, \hat{A}_\mu\}$ would be to follow the construction of Sec. 2. *Then the nonasymptotic free Fermi field would be in principle identifiable in the commutant of the (reducible) free photon field algebra.*

For this purpose we need, however, an appropriate set of electromagnetic field (coherent) states. In the traditional perturbative framework[6,31,17,19,30] one encounters the coherent photon states, which describe the soft photon clouds accompanying one or more Dirac particles. Like the soliton states $|p_1, R_1, \lambda_1,\ldots, p_N, R_N, \lambda_N)$ of (3.7) the photon coherent states $|p_1, e_1,\ldots,p_N, e_N)$ can be used to generate the respective incomplete direct product spaces, which carry pairwise unitarily inequivalent irreducible representations of the CCR algebra. However, in the photon case the parametrization is not rich enough, and the translation freedom, which is so crucial for the procedures of Sec. 2, is lacking.

B. Let $\{b_s^*(\mathbf{k}), d_s^*(\mathbf{k}), b_s(\mathbf{k}), d_s(\mathbf{k})\}_{=1,2}$ be the CAR algebra generators associated with the free Dirac field,

$$\hat{\psi}(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int \left(\frac{m}{p_0}\right)^{1/2} \sum_s [b_s(\mathbf{p}) w_s(\mathbf{p}) e^{i\mathbf{px}} + d_s^*(\mathbf{p}) v_s(\mathbf{p}) e^{-i\mathbf{px}}]\, d\mathbf{p},$$

$$\hat{\psi}(\mathbf{k}) = \frac{1}{(2\pi)^{3/2}} \int \left(\frac{m}{p_0}\right) \sum_s [b_s^*(\mathbf{p}) \overline{w}_s(\mathbf{p}) e^{-i\mathbf{px}} + d_s(\mathbf{p}) \overline{v}_s(\mathbf{p}) e^{i\mathbf{px}}]\, d\mathbf{p}, \tag{4.1}$$

$$[b_i(\mathbf{p}), b_j^*(\mathbf{q})]_+ = \delta(\mathbf{p}-\mathbf{q})\delta_{ij} = [d_i(\mathbf{p}), d_j^*(\mathbf{p})]_+,$$

and let $\{a_\mu^*(\mathbf{p}), a_\mu(\mathbf{p})\}$ be the CCR algebra ones for the free Maxwell field:

$$\hat{A}_\mu(\mathbf{x}) = \frac{1}{(2\pi)^{3/2}} \int [a_\mu^*(\mathbf{k}) e^{-i\mathbf{kx}} + a_\mu(\mathbf{k}) e^{i\mathbf{kx}}] \frac{d\mathbf{k}}{(2k_0)^{1/2}} \tag{4.2}$$

$$[a\mu(\mathbf{k}), a_\nu^*(\mathbf{p})]_- = -g_{\mu\nu}\delta(\mathbf{k}-\mathbf{p})$$

with $\operatorname{diag} g_{\mu\nu} = (1, -1, -1, -1)$, $p = ((\mathbf{p}^2 + m^2)^{1/2}, \mathbf{p})$, $k = (|\mathbf{k}|, \mathbf{k})$. The mutual commutativity condition

$$[a_\mu^\#(\mathbf{k}), b_s^\#(\mathbf{p})]_- = 0 = [a_\mu^\#(\mathbf{k}), d_s^\#(\mathbf{p})]_- \tag{4.3}$$

is imposed according to convention.

The electric charge operator for the Dirac particles reads

$$Q = -e\int d^4p\, \delta(p^2 - m^2)\Theta(p^0) \sum_s [b_s^*(p) b_s(p) - d_s^*(p) d_s(p)]. \tag{4.4}$$

One knows that the n-particle–antiparticle Dirac state vector

$$b^*_{s_1(\mathbf{p}_1)} \ldots b^*_{s_n}(\mathbf{p}_n) d^*_{t_1}(\mathbf{q}_1) \cdots d^*_{t_m}(\mathbf{q}_m)|0)_F \tag{4.5}$$

induces the infrared coherent photon state describing the associated radiation

$$|\mathbf{p}_1, e_1,\ldots,\mathbf{p}_{n+m}, e_{n+m}) = \exp\left\{\frac{e}{(2\pi)^{3/2}} \int [f^\mu_{n,m}(\cdots|\mathbf{k}) a_\mu^*(\mathbf{k}) - \bar{f}^\mu_{n,m}(\cdots|\mathbf{k}) a_\mu(\mathbf{k})] \frac{d\mathbf{k}}{(2k_0)^{1/2}}\right\}|0)_B. \tag{4.6}$$

Here $|0)_F$ and $|0)_B$ are the respective Fock vacua and one should realize that $|\mathbf{p}_1, e_1,\ldots, \mathbf{p}_{n+m}, e_{n+m})$ is not an element of the Fock space, since

$$f^\mu_{n,m}(\cdots|\mathbf{k}) = \left(\sum_{i=1}^n \frac{p_i^\mu}{p_i \cdot k} - \sum_{i=1}^m \frac{q_i^\mu}{q_i \cdot k}\right) \times \varphi(\mathbf{k}, \mathbf{p}_1,\ldots,\mathbf{p}_n \mathbf{q}_1,\ldots,\mathbf{q}_m), \tag{4.7}$$

where $\varphi(\mathbf{k}, \mathbf{p}_1\ldots, \mathbf{p}_n, \mathbf{q}_1,\ldots,\mathbf{q}_m) = 1$ for $|\mathbf{k}| \leqslant \delta \ll 1$ while φ rapidly vanishes for $|\mathbf{k}| > \delta$. Notice that because of (4.5) no coinciding $(\mathbf{p}, e)$ pairs can appear in (4.6). By introducing the orthonormal transverse polarization vectors $\epsilon_s^\mu(\mathbf{k})$, $s = 1, 2$, $\epsilon_s^0(\mathbf{k}) = 0$, $\boldsymbol{\epsilon}_s \cdot \mathbf{k} = 0$ we can rewrite (4.6) in the (noncovariant) Chung form,[31]

$$|\mathbf{p}_1, e_1,\ldots,\mathbf{p}_{n+m}, e_{n+m}) = \exp\left\{\frac{e}{(2\pi)^{3/2}} \times \int \sum_{s=1}^2 [f^s_{n,m}(\cdots|\mathbf{k}) a_s^*(\mathbf{k}) - \bar{f}^s_{n,m}(\cdots|\mathbf{k}) a_s(\mathbf{k})] \frac{d\mathbf{k}}{(2k_0)^{1/2}}\right\}|0)_B := W_{nm}(\underline{p}, \underline{e})|0)_B, \tag{4.8}$$

where

$$f^s_{n,m}(\cdots|\mathbf{k}) = f_{nm}(\cdots|\mathbf{k}) \cdot \epsilon_s(\mathbf{k}) \tag{4.9}$$

so that

$$[a_i(\mathbf{k}), a_j^*(\mathbf{p})]_- = \sum_{\mu\nu} \epsilon_i^\mu(\mathbf{k})\,\epsilon_j^\nu(\mathbf{p})\,[a_\mu(\mathbf{k}), a_\nu^*(\mathbf{p})]_- = \delta_{ij}\delta(\mathbf{k}-\mathbf{p}) \tag{4.10}$$

provided $\Sigma_{\mu\nu}(-g_{\mu\nu})\epsilon_i^\mu(\mathbf{k})\epsilon_j^\nu(\mathbf{k}) = \delta_{ij}$.

C. In the above discussion we have distinguished a ball Ω with the radius smaller than δ in the momentum space. Let δ be the photon momentum detectability threshold. Then photons with $|\mathbf{k}|\leqslant\delta$ may be called "soft" while those with $|\mathbf{k}|>\delta$ the "hard" ones.[31] The soft photons are described by the previously introduced coherent states, while, for the hard ones, the conventional occupation number representation is adopted in the literature.

Let Ω_s be the momentum space region with $s\cdot\delta\leqslant|\mathbf{k}|<(s+1)\delta$. The previous Ω is Ω_0 in the present notation. Let $\chi_s(\mathbf{p})$ be the characteristic function of the set Ω_s: $\chi_s(\mathbf{p})=1$ for $\mathbf{p}\in\Omega_s$, 0 otherwise, and let V_s be the respective momentum space volume of Ω_s. Then we introduce

$$a_i^\#(s) = \frac{1}{\sqrt{V_s}}\int d\mathbf{p}\; a_i^\#(\mathbf{p})\chi_s(\mathbf{p}),$$
$$[a_i(s), a_j^*(t)]_- = \delta_{ij}\delta_{st},\ i,j=1,2\ \ s,t=0,1,2,\cdots, \tag{4.11}$$

and quite analogously

$$[b_i(s), b_j^*(t)]_+ = \delta_{ij}\delta_{st} = [d_i(s), d_j^*(t)]_+ \tag{4.12}$$

for Fermi operators (4.1).

At this point let us consider a single hard ($s>0$) photon mode $\{a_i^*(s), a_i(s)\}_{i=1,2}$. By $W_s(a^*, a)$ we denote a polynomial in $a_i^*(s)$, $a_i(s)$, $i=1,2$, s fixed. Let us consider the set of all such polynomials $\{W_s(a^*,a)\}$, s fixed. Then the Hilbert space closure of the *set of vectors* $\{W_s(a^*,a)|0)_B\}$ is a Hilbert space $h_s = \overline{\{W_s(a^*,a)|0)_B\}}$ of the sth photon mode.

The lattice index s will be, for simplicity, omitted in below. While in $h_s = h$ we have the two Bose degrees of freedom $\{a_i^*, a_i\}_{i=1,2}$. By using them we can easily construct the infinitesimal generators of the $E(2)$ group Lie algebra in h, namely, the obvious formulas:

$$a = q+ip,\quad a^* = q-ip \tag{4.13}$$

allows us to introduce in h the two translation generators

$$N_j = (1/2i)(a_j - a_j^*) = p_j,\ q_j = \tfrac{1}{2}(a_j + a_j^*),\ j=1,2, \tag{4.14}$$

which together with the rotation generator

$$\mathscr{J}_3 = q_1p_2 - q_2p_1 \tag{4.15}$$

form the $E(2)$ group Lie algebra

$$[\mathscr{J}_3, N_1]_- = iN_2,$$
$$[\mathscr{J}_3, N_2]_- = -iN_1,\ [N_1,N_2]_- = 0. \tag{4.16}$$

In the most obvious Schrödinger representation (4.14)–(4.15) read

$$N_1 = -i\frac{\partial}{\partial u},\ q_1 = u,\ N_2 = -i\frac{\partial}{\partial v},\ q_2 = v,$$
$$\mathscr{J}_3 = -i\Big(u\frac{\partial}{\partial v} - v\frac{\partial}{\partial u}\Big),\ \hbar = c = 1. \tag{4.17}$$

The unitary in h translation operators have the form

$$T_\lambda = \exp i\boldsymbol{\lambda}\mathbf{N} = \exp\sum_{j=1}^{2}\frac{\lambda_j}{2}(a_j - a_j^*); \tag{4.18}$$

hence, if applied to the Fock vacuum, they generate the coherent states for the sth (hard photon) mode, parametrized by real labels λ_1, λ_2:

$$T_\lambda|0) = |\lambda_1,\lambda_2),\ a_j|\lambda_1,\lambda_2) = \tfrac{1}{2}\lambda_j|\lambda_1,\lambda_2). \tag{4.19}$$

If to recall that the coherent soft photon states (4.8) were constructed under an assumption that all entering photon modes belong to Ω_0, we find that each soft photon state may be considered as a substitute for the (Fock) vacuum for all non-Ω_0 (i.e., hard) photon modes:

$$a_j(s)|\mathbf{p}_1,e_1,\ldots,\mathbf{p}_{n+m},e_{n+m}) = 0$$
$$\forall j=1,2,\ \forall s>0. \tag{4.20}$$

On the other hand, due to the sufficiently bad $\mathbf{k}=0$ behavior of the boson transformation parameters f^μ,

$$a_\mu(\mathbf{k}) \to a_\mu(\mathbf{k}) + f^\mu_{nm}(\cdots|\mathbf{k}) \tag{4.21}$$

the hard photon (but soft) vacua (4.7) generate the rich family of unitarily inequivalent representations of the photon field (CCR) algebra, despite the fact that one exploits only the $\mathbf{k}\in\Omega_0$ modes for the construction of generating vectors. It is rather clear that the concentration on the infrared aspect of the radiation field made people[6,17–20,30,31] not even notice that theory does not seem to forbid such boson transformation parameters for the radiation field, which (1) behave like $f^\mu_{n,m}(\cdots|\mathbf{k})$ of (4.7) for $\mathbf{k}\in\Omega_0$ i.e., $|\mathbf{k}|\ll 1$, but (2) behave as badly at $|\mathbf{k}|\to\infty$, as, say, the sine-Gordon parameters (1.8). Property (1) would allow for the standard infrared construction, while property (2), upon appropriately varying the boson transformation parameters with respect to their (bad) $|\mathbf{k}|\to\infty$ behavior, allows in principle for the construction of the set of unitarily inequivalent families of infrared representations: Within each family the unitarily inequivalent infrared representations would still persist.

Below we shall not enter into the problem of determining the boson transformation parameters which in addition to (1) and (2) would exhibit a consistency with the equations of motion for the radiation field [as $f^\mu_{n,m}$ of (4.8) do]. Instead we shall assume the appropriate parameters to be granted, and analyze some consequences of this ansatz.

D. Let us notice that $f^\mu_{n,m}(\cdots|\mathbf{k})$ may be written as a sum of contributions following from single charged particles:

$$f^\mu_{n,m}(\cdots|\mathbf{k}) = \sum_{i=1}^{n}\frac{p_i^\mu}{p_i\cdot k}\varphi(\mathbf{k},p_i) - \sum_{j=1}^{m}\frac{q_i^\mu}{q_i\cdot k}\varphi(\mathbf{k},q_i) = \sum_{i=1}^{n+m} f^\mu(e_{ij},\mathbf{r}_i|\mathbf{k}), \tag{4.22}$$

$$e_i = \pm 1,\ \mathbf{r}_i = \mathbf{p}_i,\ i\leqslant n,\ \mathbf{r}_i = \mathbf{q}_i,\ i>n,$$
$$\varphi(\mathbf{k}, r_i) = 1\ \text{for}\ \mathbf{k}\in\Omega_0\ \forall i=1,2,\ldots,n+m.$$

Our idea is to introduce a modified boson transformation parameter in the place of this in (4.7)–(4.8):

$$f^\mu(e,\mathbf{r},\lambda,\lambda'|\mathbf{k}) = \chi_0(\mathbf{k})f^\mu(e,\mathbf{r}|\mathbf{k}) + [1-\chi_0(\mathbf{k})]g^\mu(e,\mathbf{r},\lambda,\lambda'|\mathbf{k}) \tag{4.23}$$

with $\chi_0(\mathbf{k})=1$ for $\mathbf{k}\in\Omega_0$, 0 otherwise. We demand further that the $|\mathbf{k}|\to\infty$ dependence be regulated by the choice of two real parameters λ, λ'. An analogy with the sine-Gordon considerations of Sec. 2 suggests that the modified coherent photon states

$$|\mathbf{p}_1, e_1, \lambda_1, \lambda'_1, ..., \mathbf{p}_N, e_N, \lambda_N, \lambda'_N)$$
$$= \exp\left\{ \frac{e}{(2\pi)^{3/2}} \int \sum_{s=1}^{2} \sum_{j=1}^{N} [f^s(\mathbf{p}_j, e_j, \lambda_j, \lambda'_j|\mathbf{k}) a_s^*(\mathbf{k}) - f^s(\mathbf{p}_j, e_j, \lambda_j \lambda'_j|\mathbf{k}) a_s(\mathbf{k})] \frac{d\mathbf{k}}{(2k_0)^{1/2}} \right\} |0)_B \quad (4.24)$$

should form the set of vectors, which are pairwise neither equivalent nor weakly equivalent, unless the respective parametric sets $\mathscr{P}$ and $\mathscr{P}'$ with $\mathscr{P} = \{\mathbf{p}_1, e_1, \lambda_1, \lambda'_1, ..., \mathbf{p}_N, e_N, \lambda_N, \lambda'_N\}$ do coincide: $\mathscr{P} \cap \mathscr{P}' = \mathscr{P} = \mathscr{P}'$. Then at a fixed choice of N, and $\{\lambda_i, \lambda'_i\}_{1<i<N}$ we remain on the level of standard infrared recipes. The infrared photon Hilbert spaces received from the generating vectors $|\mathbf{p}_1, e_1, ..., \mathbf{p}_N, e_N)_{\underline{\lambda}\underline{\lambda}'}$ by varying $\mathbf{p}_i$, e_i carry pairwise inequivalent representations of the photon field algebra. In addition, the infrared families arising for different choices of $\{\lambda_i, \lambda'_i\}_{1<i<N}$ are unitarily inequivalent, albeit for each fixed choice of $\{\lambda_i, \lambda'_i\}_{1<i<N}$ the same infrared physics is described.

At this point let us fix N and the other charged particles data $\{e_i, \mathbf{p}_i\}_{1<i<N}$. For each sequence $\{\lambda_i, \lambda'_i\}_{1<i<N}$ we deal with a separable Hilbert space IDPS $(|e, p, \underline{\lambda}, \underline{\lambda}')_N)$ [whose orthonormal basis system we distinguish by an additional index $n = 1, 2,...$; compare, (2.9)]. We shall adopt both the direct integral procedures of Sec. 2 [see (2.20)] and the spin-$\frac{1}{2}$ SU(2) operator construction of (3.10) to arrive at operators in the Hilbert space:

$$\mathscr{H}_f^N = \sum_{\{e\}} \int_{R^1}^{\oplus} \mathrm{IDPS}(|e, p, \underline{\lambda}, \underline{\lambda}')_N)\, d\mu(\underline{\lambda}, \underline{\lambda}'). \quad (4.25)$$

Let us introduce

$$A_j^{\alpha\pm} = \frac{1}{\sqrt{2}}(Q_j^{\alpha\pm} + iP_j^{\alpha\pm}$$
$$= \sum_{\mathrm{conf}\{e\}}^{(j)} \int_{R^1}^{\oplus} \cdots \int_{R^1}^{\oplus} d\mu(\underline{\lambda}, \underline{\lambda}')$$
$$\times \sum_n |n, ..., \mathbf{p}_j, \pm, \lambda_j^1, \lambda_j^2, \cdots)$$
$$\times \frac{1}{\sqrt{2}} \left(\lambda_j^\alpha + \frac{\partial}{\partial \lambda_j^\alpha}\right)(n, ..., \mathbf{p}_j, \pm \lambda_j^1, \lambda_j^2, \cdots | \quad (4.26)$$

with $(A_j^{\alpha\pm})^* = A_j^{*\alpha\pm}$, $\alpha = 1, 2, \lambda_j = \lambda_j^1$, $\lambda'_j = \lambda_j^2$, $j = 1, 2..., N$, the sum $\Sigma_{\mathrm{conf}\{e\}}^{(j)}$ having exactly the same meaning as the one in (3.10). The object $A_j^{*\alpha\pm}$ differs from $A_j^{\alpha\pm}$ in the replacement of $(1/\sqrt{2})(\lambda_j^\alpha + \partial/\partial\lambda_j^\alpha)$ by $1/\sqrt{2}(\lambda_j^\alpha - \partial/\partial\lambda_j^\alpha)$.

Due to the orthogonality of inequivalent coherent states constituting objects (4.26) we have

$$A_j^{\alpha\pm} A_j^{\alpha\mp} = 0 = A_j^{\alpha\pm} A_j^{*\alpha\mp},$$
$$[A_j^{\alpha+}, A_j^{*\alpha+}]_- + [A_j^{\alpha-}, A_j^{*\alpha-}]_- = 1 \quad (4.27)$$

so that the operators

$$B_j^{\alpha q} = A_j^{\alpha+} + i(-1)^{q-1} A_j^{\alpha-},$$
$$\alpha = 1, 2,\ q = 1,\ 2,\ j = 1, 2, ..., N \quad (4.28)$$

together with their Hermitian adjoints satisfy the (infinitesimal) CCR algebra commutation relations

$$[B_j^{\alpha q}, B_{j'}^{*\alpha' q'}]_- = \delta_{\alpha\alpha'} \delta_{qq'} \delta_{jj'},$$
$$[B_j^{\alpha q}, B_{j'}^{\alpha' q'}]_{-1} = 0 = [B_j^{*\alpha q}, B_{j'}^{*\alpha' q'}]_{-1}. \quad (4.29)$$

The operators (4.28) do carry the charge–spin–momentum labels of the Dirac particles (fermions), though acting explicitly in the (direct integral) Hilbert space of the (reducible) electromagnetic field algebra and satisfying the canonical commutation relations (the CCR). Notice that by construction, operators (4.28) belong to the commutant of the electromagnetic field algebra and the Fock-ness property for the representation (4.29) can be introduced analogously to this of (2.21).

Once the CCR algebra generators are given, one can exploit the study of isomorphisms between Hilbert spaces of symmetric and antisymmetric functions,[27] to construct the CAR algebra generators and then to represent them in the Hilbert space of the Bose system. For Dirac fermions, this construction has been accomplished in Refs. 32, 29, and 5 (see also Ref. 33, and the number of integral degrees of freedom $\{\alpha, q\}$ is preserved in this CCR $\to$ CAR = CAR(CCR) mapping.

Operators (4.28) completely suffice for the construction of the Fermi set $\{b_\alpha(s), b_\alpha^*(s), d_\alpha(s), b_\alpha^*(s)\}$ of (4.12), where, however, in contrast to the case of Ref. 29, the Fermi operators do belong to the commutant of the photon field algebra, and obviously (this time like in the case of Ref. 29) do not need any separate Fermi Hilbert space to have them represented.

Let us once more emphasize that the present construction differs essentially from the one given in Ref. 29 and Ref. 5, example 2, where a possible electromagnetic field content of the massive Dirac field was analyzed. Namely, in the latter case fermions were essentially arising in the photon field algebra, while in the present case they arise in the commutant of this algebra, thus not belonging to it (but acting in the photon field Hilbert space).

It seems to be a quite appealing idea that the asymptotic solution for the quantized coupled Dirac–Maxwell system given for example, by Zwanziger[19] (on the basis of the ansatz concerning the asymptotic limit of the renormalized Heisenberg electric current operator) can be given in terms of the objects we have introduced above. However, for this purpose, we find it unavoidable to have solved the existence problem for boson transformation parameters (4.23). By existence we mean a consistency with the classical equations of motion for the coupled Dirac–Maxwell system.

Let us recall[29] our statement of belief: *A necessary condition for the existence of any physically meaningful quantum field theory model is that the corresponding classical model exists and is soluble.*

For the classical Dirac–Maxwell system both the existence and (local in time) solubility were proved,[34,35] but in contrast to (1 + 1)-dimensional models the explicit solutions are painfully lacking.

Let us mention that, despite the latter problem, the relationship between the quantized Fermi–Dirac–Maxwell system and the corresponding classical (c-number) one has been investigated in the series of papers of the present author[32,29,5,36] on the quantization of spinor fields, see also Refs. 37, 38.

[1] J. R. Klauder, Phys. Rev. Lett. **20**, 763 (1972).
[2] P. Garbaczewski, J. Math. Phys. **22**, 574 (1981).
[3] P. Garbaczewski, J. Math. Phys. **22**, 1272 (1981).
[4] P. Garbaczewski, Ann. Phys. (N.Y.) **139**, 293 (1982).
[5] P. Garbaczewski, J. Math. Phys. **24**, 341 (1983).
[6] T. W. B. Kibble, J. Math. Phys. **9**, 315 (1968).
[7] L. Mercaldo, I. Rabuffo, and G. Vitiello, Nucl. Phys. B **188**, 193 (1981).
[8] I. M. Gel'fand and N. Ja. Vilenkin, *Generalized Functions* (Moscow, 1961) (in Russian), Vol. 4.
[9] M. A. Neumark, *Normed Rings* (Nauka, Moscow, 1968) (in Russian).
[10] A. E. Nussbaum, Duke Math. J. **31**, 33 (1964).
[11] P. Garbaczewski, J. Math. Phys. **24**, 651 (1983).
[12] R. Haag, Nuovo Cimento **25**, 287 (1962).
[13] H. Araki and E. J. Woods, J. Math. Phys. **4**, 637 (1963).
[14] J. R. Klauder, J. Math. Phys. **8**, 1666 (1965).
[15] J. Lopuszański and H. Reeh, J. Math. Phys. **7**, 148, 1821 (1966).
[16] R. T. Powers, Commun. Math. Phys. **4**, 3 (1967).
[17] P. P. Kulish and L. D. Faddeev, Teor. Mat. Fiz. **4**, 153 (1970).
[18] J. Fröhlich, Ann. Inst. H. Poincaré, **19**, 1 (1973).
[19] D. Zwanziger, Phys. Rev. **D 11**, 3481 (1975).
[20] J. Fröhlich, G. Morchio, and F. Strocchi, Ann. Phys. (N.Y.) **119**, 241 (1979).
[21] H. Matsumoto, N. J. Papastamatiou, G. Semenoff, and H. Umezawa, Phys. Rev. **D 23**, 1339 (1981).
[22] H. Matsumoto, N. J. Papastamatiou, G. Semenoff, and H. Umezawa, Phys. Rev. **D 24**, 406 (1981).
[23] J. G. Taylor, Ann. Phys. (N.Y.) **115**, 153 (1978).
[24] L. A. Tahtajan and L. D. Faddeev, Usp. Mat. Nauk **34**, 13 (1979).
[25] M. Karowski and H. J. Thun, Nucl. Phys. B **130**, 295 (1977).
[26] A.B. Zamolodchikov and A. B. Zamolodchikov, Ann. Phys. (N.Y.) **120**, 253 (1979).
[27] P. Garbaczewski, Commun. Math. Phys. **43**, 131 (1975).
[28] V. E. Korepin, L. D. Faddeev, Theor. Math. Phys. **25**, 1039 (1975).
[29] P. Garbaczewski, J. Math. Phys. **23**, 442 (1982).
[30] J. M. Jauch and F. Rohrlich, *Quantum Theory of Photons and Electrons* (Springer-Verlag, Berlin, 1976).
[31] V. Chung, Phys. Rev. **140**, B110 (1965).
[32] P. Garbaczewski, J. Math. Phys. **19**, 642 (1978).
[33] P. Garbaczewski, Phys. Rep. C. **36**, 65 (1978).
[34] L. Gross, Comm. Pure Appl. Math. **19**, 1 (1966).
[35] J. M. Chadam, J. Math. Phys. **13**, 597 (1972).
[36] P. Garbaczewski, "Quantization of spinor fields IV. Joint Bose-Fermi spectral problems," J. Math. Phys. (to appear).
[37] P. Garbaczewski, Nucl. Phys. B **218**, 321 (1983).
[38] P. Garbaczewski, Ann. Phys. (NY) **149** (2) (1983).

On-shell singularities of the off-shell Jost functions and Jost solutions for long-range potentials

Jan Dušek
Department of Physics, Institute of Mechanical and Electrical Engineering, Nejedlého sady 14, 306 14 Plzeň, Czechoslovakia

(Received 8 March 1983; accepted for publication 29 April 1983)

Using the known form of the on-shell singularities of the off-shell incoming and outgoing wave solutions, we investigate the on-shell singularities of the off-shell Jost solutions and off- and half-shell Jost functions for local central potentials vanishing at infinity as $cr^{-\alpha}$, where $\alpha\in(0,1]$. We show that all the singularities can be expressed via the renormalization factors of the incoming and outgoing wave solutions. The obtained results explain the singular term appearing for long-range potentials in the partial waves of the half-shell scattering amplitude in the neighborhood of the energy shell.

PACS numbers: 03.80. + r, 11.20. − e, 02.30.Hq

I. INTRODUCTION

For various quantities relative to long-range potentials, the on-shell values cannot be obtained as limits of the off-shell ones. These on-shell singularities are closely related to the renormalization factors that have to be introduced in the time-dependent scattering theory in order that the asymptotic condition might be generalized to long-range potentials.[1,2] This connection has been proved by Zorbas[3] in the Coulomb case, and the method[3] can be used to give a formal expression relating the time-independent renormalization factors to the time-dependent ones in the general case of potentials of Ref. 2. In Refs. 4 and 5 we have investigated explicit forms of singularities of the partial wave projections of the off- and half-shell Coulomb and Coulomb-like T matrix and of the off-shell Coulomb wave functions. We have shown that when passing from the partial waves of the off-shell T matrix and off-shell wave function to half- and on-shell values, respectively, the same singularities appear as in the unprojected quantities. However, the partial waves of the half-shell T matrix have been shown to contain, unlike the unprojected half-shell T matrix,[6] not only a singular factor but also a singular term. The presence of the singular term has been explained in Ref. 4 using the properties of the partial wave decomposition of the Coulomb scattering amplitude found by Taylor.[7]

The method of Refs. 4 and 7 neither explains the form of the singular term nor reveals the connection to the on-shell singularities of the off-shell Coulomb wave functions.[3] It appears that such a connection exists. It can be established by investigating the on-shell singularities of half-shell Coulomb–Jost functions.

In the present paper, we study the on-shell singularities of various quantities obtained in the generalized off-shell Jost formalism.[8–13] In Sec. II, we analyze the possibilities of defining various quantities and of extending the validity of several relations of the standard short-range Jost formalism in the case when a long-range potential is present. We find that off the energy shell, the short-range relations can be taken over directly. The short-range definitions of the on-shell Jost solutions and half- and on-shell Jost functions can no longer be used. We give a convenient way of defining these quantities. In Sec. III, we derive the form of the on-shell singularities of the off-shell Jost solutions and off- and half-shell Jost functions. Our considerations are based only on the relation between the off- and on-shell values of the incoming and outgoing wave solutions resulting from the theory of Refs. 1–3. All the on-shell singularities appearing in the Jost formalism in the long-range case are determined by the renormalization factors for the incoming and outgoing wave solutions.

Units such that $\hbar = 2m = 1$ are used throughout the paper.

II. SOLUTIONS OF THE HOMOGENEOUS AND INHOMOGENEOUS SCHRÖDINGER EQUATION FOR LONG-RANGE POTENTIALS

We consider the case of local central potentials $V(r)$ such that

$$V(r) = V_L(r) + V_S(r),$$

where the short-range part of the potential V_S obeys the standard conditions of the short-range theory[14]:

$$\int_0^a |V_S(r)| r\, dr < \infty, \quad \int_{a'}^{\infty} |V_S(r)| dr < \infty \tag{1}$$

for some $a, a' \epsilon (0,\infty)$, and V_L satisfies[15]

$$\left|\frac{d^j V_L(r)}{dr^j}\right| \leqslant C(1+r)^{-j-\alpha}, \quad j \leqslant M, \tag{2}$$

where $\alpha > \frac{1}{2}$ if $M=1$, $\alpha > \frac{1}{3}$ if $M=2$, and $\alpha > 0$ if $M=3$. For such potentials there exists[15] a C^∞ function $W(\mathbf{k},t)$ for $\mathbf{k}\epsilon\mathbb{R}^3\setminus\{0\}$; $t\epsilon\mathbb{R}$, so that

$$W(\mathbf{k},s+t) - W(\mathbf{k},t) \to sk^2$$

as $t\to\pm\infty$ for each fixed $\mathbf{k},s$ and that the generalized wave operators

$$\Omega_\pm = \operatorname*{s-lim}_{t\to\mp\infty} e^{iHt}\exp[-iW(-i\nabla,t)]$$

exist for $H = H_0 + V (H_0 = -\Delta)$.

For a central potential V_L, the function $W(\mathbf{k},t)$ can be found so that it depends only on $k^2 = H_0$. Denoting then

$$G(H_0,t) = W(H_0,t) - H_0 t,$$

we can write

$$\Omega_{\pm} = \operatorname*{s-lim}_{t\to\mp\infty} e^{iHt} \exp[-iG(H_0,t)] e^{-iH_0t}. \tag{3}$$

(Both functions $W(k^2,t)$ and $G(k^2,t)$ are real and can be chosen odd.)

Having defined the interaction, we can write the homogeneous radial Schrödinger equation

$$\left(\frac{d^2}{dr^2} - \frac{\lambda^2 - \frac{1}{4}}{r^2} - V(r) + k^2\right)\psi(\lambda,k,r) = 0, \tag{4}$$

where $\psi(\lambda,k,r)$ stands for an arbitrary on-shell solution and where, for half-integer values, the angular variable is related to the angular momentum quantum number by

$$\lambda = l + \tfrac{1}{2}.$$

λ can, in general, assume complex values.

All the off-shell solutions can be written as solutions of the inhomogeneous Schrödinger equation

$$\left(\frac{d^2}{dr^2} - \frac{\lambda^2 - \frac{1}{4}}{r^2} - V(r) + k^2\right)\psi(\lambda,k,p,r) = (k^2 - p^2)\psi_0(\lambda,p,r) \tag{5}$$

with properly chosen solutions $\psi_0(\lambda,p,r)$ of the free Schrödinger equation and with properly defined boundary conditions. In Ref. 13 we have shown that, in the short-range case, it is useful to define the boundary conditions also in terms of the free solutions. For the off-shell solutions of Eq. (5), this method can still be used. There are many ways to choose these free solutions. We shall use the choice of Ref. 16, substituting for $\psi_0(\lambda,p,r)$ one of the following Riccati–Bessel–Hankel functions:

$$\hat{j}(\lambda,pr) = \sqrt{\pi pr/2}\, J_\lambda(pr),$$

$$\hat{n}(\lambda,pr) = -\sqrt{\pi pr/2}\, N_\lambda(pr),$$

$$\hat{h}^{\pm}(\lambda,pr) = \pm i\sqrt{\pi pr/2}\, H_\lambda^{(1,2)}(pr),$$

where the functions $J_\lambda(z)$, $N_\lambda(z)$, and $H_\lambda^{(1,2)}(z)$ are the Bessel, Neumann, and Hankel functions,[17] respectively.

The solutions of Eq. (5) can be defined equivalently as solutions of integral equations. For $\operatorname{Re}\lambda > 0$, the regular solution is the solution of the integral equation

$$\varphi(\lambda,k,p,r) = \hat{j}(\lambda pr) + \int_0^r g_0(\lambda,k,r,r')V(r')\times\varphi(\lambda,k,p,r')dr', \tag{6}$$

where[16]

$$g_0(\lambda,k,r,r') = [\hat{j}(\lambda,kr)\hat{n}(\lambda,kr') - \hat{n}(\lambda,kr)\hat{j}(\lambda,kr')]/k = (1/2ik)[\hat{h}^{+}(\lambda,kr)\hat{h}^{-}(\lambda,kr') - \hat{h}^{-}(\lambda,kr)\hat{h}^{+}(\lambda,kr')]. \tag{7}$$

The integral on the rhs of Eq. (6) is regular even if $p = k$ for $\operatorname{Re}\lambda > 0$. It can therefore also be used to define the regular on-shell solution $\varphi(\lambda,k,r)$, and the regular on-shell solution is an ordinary limit of the off-shell one:

$$\varphi(\lambda,k,r) = \lim_{p\to k}\varphi(\lambda,k,p,r). \tag{8}$$

The existence of the limit (8) is due to the fact that the asymptotic part of the potential is absent in Eq. (6) for finite values of r. The situation is different for the Jost and wave solutions. The usual integral equations can be used to define the off-shell Jost solutions:

$$\mathscr{f}^{(\pm)}(\lambda,k,p,r) = h^{\pm}(\lambda,pr) - \int_r^\infty g_0(\lambda,k,r,r')V(r')\times\mathscr{f}^{(\pm)}(\lambda,k,p,r')dr' \quad (p\neq k). \tag{9}$$

If $p\to k$, Eq. (9) contains a singular integral and does not define the on-shell Jost solutions. We shall define the on-shell Jost solutions later in a different way. Similarly, the Lippmann–Schwinger equation can be used to define the off-shell wave solutions but involves a singular integral if considered on-shell. Let us denote by $\psi_+(\lambda,k,p,r)$ and $\psi_-(\lambda,k,p,r)$ the incoming and outgoing off-shell wave solutions, respectively. Then

$$\psi_{\pm}(\lambda,k,p,r) = \hat{j}(\lambda,pr) + \int_0^\infty G_{0\pm}(\lambda,k,r,r')V(r')\times\psi_{\pm}(\lambda,k,p,r')dr', \tag{10}$$

where $G_{0\pm}(\lambda,k,r,r')$ are given by[16]

$$G_{0\pm}(\lambda,k,r,r') = -\hat{j}(\lambda,kr_<)\hat{h}^{\pm}(\lambda,kr_>)/k, \tag{11}$$

with

$$r_< = \min\{r,r'\},\quad r_> = \max\{r,r'\}.$$

The integral on the rhs of Eq. (10) is nonsingular provided $p\neq k$, but Eq. (10) is not suitable for the investigation of the asymptotic behavior of the wave solutions. The reason is that the integral has no limit for $r\to\infty$, i.e., the off-shell wave solutions do not behave asymptotically like the free solutions. This is characteristic of the long-range interactions. We therefore have to require that the asymptotic behavior of the scattered wave of the incoming wave solution be described by the on-shell Jost solution $\mathscr{f}^{(+)}(\lambda,k,r)$:

$$\psi_+(\lambda,k,p,r) \sim \hat{j}(\lambda,pr) + pf(\lambda,k,p)\mathscr{f}^{(+)}(\lambda,k,r) \quad (r\to\infty). \tag{12}$$

Equation (5), with $\psi_0(\lambda,p,r) = \hat{j}(\lambda,pr)$ and with the boundary conditions (12) and

$$\lim_{r\to 0}\psi_+(\lambda,k,p,r) = 0,$$

can be solved by the standard Green's function techniques yielding

$$\psi_+(\lambda,k,p,r) = \hat{j}(\lambda,pr) + \int_0^\infty G_+(\lambda,k,r,r')V(r')\hat{j}(\lambda,pr')dr', \tag{13}$$

where

$$G_+(\lambda,k,r,r') = -[W^{(+)}(\lambda,k)]^{-1}\varphi(\lambda,k,r_<)\mathscr{f}^{(+)}(\lambda,k,r_>). \tag{14}$$

$W^{(\pm)}(\lambda,k)$ stands for the Wronskians

$$W^{(\pm)}(\lambda,k) = \mathscr{f}^{(\pm)}(\lambda,k,r)\frac{d}{dr}\varphi(\lambda,k,r) - \varphi(\lambda,k,r)\frac{d}{dr}\mathscr{f}^{(\pm)}(\lambda,k,r). \tag{15}$$

Inserting Eq. (14) into Eq. (13), taking $r\to\infty$, and comparing the obtained asymptotic expression with the rhs of relation (12), we obtain

$$f(\lambda,k,p) = -[pW^{(+)}(\lambda,k)]^{-1}\int_0^\infty \varphi(\lambda,k,r')V(r')\hat{j}(\lambda,pr')dr'. \tag{16}$$

Due to the fact that [see Eq. (9)]

$$\mathscr{f}^{(\pm)}(\lambda,k,p,r) \sim \hat{h}^{\pm}(\lambda,pr) \text{ for } r\to\infty$$

and that

$$\hat{j}(\lambda,pr) = [\hat{h}^{+}(\lambda,pr) - \hat{h}^{-}(\lambda,pr)]/2i,$$

the asymptotic relation (12) implies

$$\psi_+(\lambda,k,p,r) = [\mathscr{f}^{(+)}(\lambda,k,p,r) - \mathscr{f}^{(-)}(\lambda,k,p,r)]/2i + pf(\lambda,k,p)\mathscr{f}^{(+)}(\lambda,k,r). \tag{17}$$

Equations (15)–(17) can be used to define the on-shell Jost solution $\mathscr{f}^{(+)}(\lambda,k,r)$ up to an arbitrary constant factor.

The case of the outgoing wave solution need not be considered separately as by Eqs. (10) and (11),

$$\psi_-(\lambda,k,p,r) = \psi_+^*(\lambda,k,p,r). \tag{18}$$

Equations (15)–(17) then also imply

$$\mathscr{f}^{(-)}(\lambda,k,r) = \mathscr{f}^{(+)*}(\lambda,k,r). \tag{19}$$

(* means complex conjugation.)

The half-shell Jost functions can be defined by the expressions of the short-range theory[13]

$$F^{(\pm)}(\lambda,k,q) - 1 = k^{-1}\int_0^\infty \mathscr{f}^{(\pm)}(\lambda,k,r)V(r)\hat{j}(\lambda,q,r)dr \tag{20}$$

and

$$F^{(\pm)}(\lambda,q,k) - 1 = k^{-1}\int_0^\infty \hat{h}^{\pm}(\lambda,qr)V(r)\varphi(\lambda,k,r)dr. \tag{21}$$

[Sometimes slightly different normalization is used. For example,[8,11,12] the rhs of Eq. (20) is multiplied by $(k/q)^{\lambda+1/2}$ to obtain expressions regular for $k\to 0$.] Using the Jost functions (20), we can express the solution $\varphi(\lambda,k,p,r)$ by means of the Jost solutions as follows[12,13]:

$$\begin{aligned}\varphi(\lambda,k,p,r) = (1/2i)[&\mathscr{f}^{(+)}(\lambda,k,p,r)\\ &+ (F^{(-)}(\lambda,k,p) - 1)\mathscr{f}^{(+)}(\lambda,k,r)\\ &- \mathscr{f}^{(-)}(\lambda,k,p,r)\\ &- (F^{(+)}(\lambda,k,p) - 1)\mathscr{f}^{(-)}(\lambda,k,r)].\end{aligned} \tag{22}$$

Similarly, the Jost functions (21) express the behavior of the Jost solutions in the neighborhood of $r = 0$. Assuming all the time $\operatorname{Re}\lambda > 0$, Eqs. (7) and (9) yield (see also Ref. 13)

$$\begin{aligned}\mathscr{f}^{(\pm)}(\lambda,k,p,r) \sim (&F^{(\pm)}(\lambda,p,k) - 1)\\ &\times \hat{n}(\lambda,kr) + \hat{n}(\lambda,pr) + O(r^{\lambda+1/2})\end{aligned}$$

$$\text{for } r\to 0. \tag{23}$$

It is immediately seen that for $q\to k$, both integrals (20) and (21) diverge. A special definition is thus needed for the on-shell Jost functions. In analogy with Ref. 14, we shall define them as follows:

$$F^{(\pm)}(\lambda,k) = W^{(\pm)}(\lambda,k)/k. \tag{24}$$

Using definition (24), we can write

$$\begin{aligned}\varphi(\lambda,k,r) = [&F^{(-)}(\lambda,k)\mathscr{f}^{(+)}(\lambda,k,r)\\ &- F^{(+)}(\lambda,k)\mathscr{f}^{(-)}(\lambda,k,r)]/2i\end{aligned} \tag{25}$$

and

$$\mathscr{f}^{\pm}(\lambda,k,r) \sim F^{(\pm)}(\lambda,k)\hat{n}(\lambda,kr) + O(r^{\lambda+1/2})$$

$$(r\to 0). \tag{26}$$

Of course, the fully off-shell Jost functions[12,13] can be defined by the relations of the short-range theory[13]:

$$\begin{aligned}F^{(\pm)}(\lambda,p,q;k) - 1 &= k^{-1}\int_0^\infty \hat{h}^{\pm}(\lambda,pr)V(r)\varphi(\lambda,k,q,r)dr\\ &= k^{-1}\int_0^\infty \hat{j}(\lambda,qr)V(r)\mathscr{f}^{(\pm)}(\lambda,k,p,r)dr.\end{aligned} \tag{27}$$

Equation (27) implies that the limit $\lim_{p\to k} F^{(\pm)}(\lambda,p,q;k)$ does not exist but that

$$F^{(\pm)}(\lambda,p,k) = F^{(\pm)}(\lambda,p,k;k) \tag{28}$$

if $p\neq k$.

We have examined the definition of all necessary quantities involved in the Jost formalism for long-range potentials. We have not defined the on-shell incoming and outgoing wave solutions and the normalization of the on-shell Jost solutions. This will be done in the next section.

III. ON-SHELL SINGULARITIES OF THE JOST FUNCTIONS AND OF THE JOST SOLUTIONS

In this section, we derive the form of the on-shell singularities of the Jost functions and Jost solutions. We shall take the relation between the off- and on-shell incoming and outgoing solutions as the starting point. The procedure of Ref. 3, establishing the connection between time-dependent and time-independent renormalization factors, can, in fact, be applied directly to the general class of potentials of the type (1),(2). The time-dependent renormalization factor on the rhs of Eq. (3) yields

$$R_+(k^2+i\epsilon,k) = \int_0^\infty e^{-z}\exp[-iG(k^2+i\epsilon,z/\epsilon)]dz, \tag{29}$$

where the time-independent renormalization factor $R_+(k^2+i\epsilon,k)$ gives the on-shell wave function as the limit

$$\psi_+(\lambda,k,r) = \lim_{\epsilon\to 0} R_+^{-1}(k^2+i\epsilon,k)\psi_+(\lambda,\sqrt{k^2+i\epsilon},k,r).$$

[The function of the lhs is the incoming wave solution of Eq. (4) and the second factor on the rhs is the incoming off-shell solution (10) with complex energy. As the renormalization factor does not depend on angular variables, the same relation holds for the full wave function as well as its partial waves.]

As our off-shell wave functions are with real energy, we need renormalization factors $R_+(k,p)$ for k,p real, and $p\neq k$. It is not our purpose to discuss their existence in detail. We only observe that the theorem[15] as formulated in Ref. 2 gives no result in this respect. However, if the function $G(k^2,t)$ has the property that the real t axis can be rotated around the origin of the complex t plane so that the integral

$$R_+(k,p)=\int_0^\infty e^{-z}\exp\left[-iG\left(k^2,\frac{z}{i(p^2-k^2-i0)}\right)\right]dz \quad (30)$$

is defined, the sought time-independent renormalization factor, yielding the on-shell incoming wave function as the limit

$$\psi_+(\lambda,k,r)=\lim_{p\to k}R_+^{-1}(k,p)\psi_+(\lambda,k,p,r), \quad (31)$$

is obtained from factor (29) by analytic continuation. For the Coulomb potential and a large class of long-range potentials with $\alpha<\frac{1}{2}$ for which time-dependent renormalization factors are given in Ref. 18, integral (30) exists.

The case of outgoing wave solutions need not be considered separately. Taking into acount that the function $G(k^2,t)$ is real for real values of the arguments and odd as a function of t, and assuming Eq. (30) holds, we can determine the renormalization factor for the outgoing wave solution as

$$R_-(k,p)=R_+^*(k,p). \quad (32)$$

The on-shell outgoing wave solution, defined by Eq. (31) with the subscripts + replaced by −, is due to Eqs. (10) and (32), the complex conjugate of the incoming one:

$$\psi_-(\lambda,k,r)=\psi_+^*(\lambda,k,r). \quad (33)$$

In the Coulomb case ($V_L=V_C=c/r$) the renormalization factor (30) has the form[3,19]

$$R_{c+}(k,p)=\Gamma(1-i\gamma)[i(p^2-k^2-i0)/4k^2]^{i\gamma}, \quad (34)$$

where $\gamma=c/2k$.

The on-shell incoming wave solution is again zero for $r\to 0$ which implies that

$$\psi_+(\lambda,k,r)=A(\lambda,k)\varphi(\lambda,k,r). \quad (35)$$

Moreover, $\psi_+(\lambda,k,r)$ is a linear combination of the on-shell Jost functions $f^{(+)}(\lambda,k,r)$ and $f^{(-)}(\lambda,k,r)$, which could be defined only up to an undetermined factor in the previous section. We shall define this factor by requiring

$$\psi_+(\lambda,k,r)=(1/2i)[f^{(+)}(\lambda,k,r)-f^{(-)}(\lambda,k,r)] + kf(k)f^{(+)}(\lambda,k,r). \quad (36)$$

Equation (36) defines the on-shell scattering amplitude $f(k)$ as well.

If we compare Eqs. (25) and (35) to Eq. (36), we get

$$A(\lambda,k)=[F^{(+)}(\lambda,k)]^{-1}. \quad (37)$$

By inserting Eq. (14) into Eq. (13), taking $r\to 0$, and using Eqs. (20) and (24), we obtain

$$\psi_+(\lambda,k,p,r)=\varphi(\lambda,k,p,r) - ((F^{(+)}(\lambda,k,p)-1)/F^{(+)}(\lambda,k))\varphi(\lambda,k,r). \quad (38)$$

Equations (8), (31), (35), (37), and (38) now yield

$$\lim_{p\to k}[F^{(\pm)}(\lambda,k,p)+R_\pm(k,p)]=F^{(\pm)}(\lambda,k)+1. \quad (39)$$

To include the index − when writing Eq. (39), we have used the relation

$$F^{(-)}(\lambda,\ldots)=F^{(+)*}(\lambda,\ldots), \quad (40)$$

which holds due to Eqs. (9), (19), (24), and (27) for off-, half-, and on-shell Jost functions defined in Sec. II. Equation (39) shows that $F^{(\pm)}(\lambda,k,p)$ contain singular terms.

If we insert Eqs. (24) and (25) into Eq. (16), and use Eq. (20), we get

$$f(\lambda,k,p)=-(1/2ipF^{(+)}(\lambda,k)) \times[F^{(-)}(\lambda,k)(F^{(+)}(\lambda,k,p)-1) - F^{(+)}(\lambda,k)(F^{(-)}(\lambda,k,p)-1)]. \quad (41)$$

Equation (39) then yields [see also Eqs. (25) and (36)]

$$\lim_{p\to k}R_+^{-1}(k,p)[2ipf(\lambda,k,p)+R_-(k,p)] =F^{(-)}(\lambda,k)/F^{(+)}(\lambda,k) =2ikf(k)+1. \quad (42)$$

This relation has already been discussed in Ref. 4 in the Coulomb case. Equation (42) shows that a singular term $[-R_-(k,p)]$, independent on λ, is to be left out in $2ipf(\lambda,k,p)$ in order that the renormalization by the factor $R_+^{-1}(k,p)$ might be performed.

The on-shell singularities of the off-shell Jost solutions can be obtained simply with the help of Eqs. (22) and (39):

$$\lim_{p\to k}R_\mp^{-1}(k,p)f^{(\pm)}(\lambda,k,p,r)=f^{(\pm)}(\lambda,k,r). \quad (43)$$

Equation (43) and relation (23) directly yield the singularity of the half-shell Jost function $F^{(\pm)}(\lambda,p,k)$:

$$\lim_{p\to k}R_\mp^{-1}(k,p)F^{(\pm)}(\lambda,p,k)=F^{(\pm)}(\lambda,k). \quad (44)$$

To obtain the singularity of the fully off-shell Jost function (27) arising if the variable p approaches the energy shell, it is sufficient to consider the integral representations (20) and (27). The singularity of integral (27) is determined manifestly by the singularity (43) of the Jost solution. Thus

$$\lim_{p\to k}R_\mp^{-1}(k,p)[F^{(\pm)}(\lambda,p,q;k)-1]=F^{(\pm)}(\lambda,k,q)-1. \quad (45)$$

This completes the description of all the possible ways that the on-shell Jost functions $F^{(\pm)}(\lambda,k)$ can be calculated from the fully off-shell Jost functions (27). We have also obtained the relation between the off- and on-shell Jost solutions. It is interesting to note that the on-shell Jost solutions defined by Eqs. (13)–(17) and (36) have the same Wronskian as in the short-range case:

$$f^{(-)}(\lambda,k,r)\frac{d}{dr}f^{(+)}(\lambda,k,r) - f^{(+)}(\lambda,k,r)\frac{d}{dr}f^{(-)}(\lambda,k,r)=2ik.$$

This equality is obtained by calculating the Wronskian of $\psi_+(\lambda,k,r)$ and $f^{(+)}(\lambda,k,r)$ once using Eq. (36) and once using Eqs. (15), (24), (35), and (37) and by comparing the results. It indicates that, in analogy with the Coulomb case, the asymptotic behavior of the general long-range on-shell Jost solutions differs from the asymptotic behavior of the short-range Jost solutions only by an oscillating phase factor.

TABLE I. On-shell behavior of off-shell solutions and half- and off-shell Jost functions for long-range potentials.

Function	Transition off-shell→on-shell
$\psi_\pm(\lambda,k,p,r)$	$\psi_\pm(\lambda,k,r)=\lim_{p\to k} R_\pm^{-1}(k,p)\psi_\pm(\lambda,k,p,r)$
$\varphi(\lambda,k,p,r)$	$\varphi(\lambda,k,r)=\lim_{p\to k}\varphi(\lambda,k,p,r)$
$f^{(\pm)}(\lambda,k,p,r)$	$f^{(\pm)}(\lambda,k,r)=\lim_{p\to k} R_\mp^{-1}(k,p) f^{(\pm)}(\lambda,k,p,r)$
$F^{(\pm)}(\lambda,p,q;k)$	$p\neq k$: $F^{(\pm)}(\lambda,p,k)=\lim_{q\to k}F^{(\pm)}(\lambda,p,q;k)$ $q\neq k$: $F^{(\pm)}(\lambda,k,q)-1=\lim_{p\to k}R_\mp^{-1}(k,p)$ $\times[F^{(\pm)}(\lambda,p,q;k)-1]$
$F^{(\pm)}(\lambda,p,k)$	$F^{(\pm)}(\lambda,k)=\lim_{p\to k}R_\mp^{-1}(k,p)F^{(\pm)}(\lambda,p,k)$
$F^{(\pm)}(\lambda,k,p)$	$F^{(\pm)}(\lambda,k)=\lim_{p\to k}[F^{(\pm)}(\lambda,k,p)+R_\pm(k,p)-1]$

IV. AN EXAMPLE FROM THE THEORY OF COULOMB SCATTERING

We have not found enough examples in the literature to illustrate all the obtained relations. One simple example can be given in the case when the asymptotic part of the interaction is dominated by the Coulomb potential. As we have seen, the investigated singularities are independent of the short-range part of the interaction and, therefore, we can limit ourselves to the scattering by the pure Coulomb potential

$$V(r)=V_C(r)=2k\,\gamma/r. \tag{46}$$

In this case, an analytic formula for the half-shell Jost function $F^{(+)}(l+\frac{1}{2},p,k)$ $(l=0,1,2,\ldots)$ can be found in Ref. 20. The normalization of the half-shell Jost functions of Ref. 20 is such that [cf. Eq. (4) of Ref. 20 and Eqs. (21), (35), and (37)]

$$f_{C,l}(k,p)-1=(p/k)^l\,[F_C^{(+)}(l+\tfrac{1}{2},p,k)-1],$$

where $f_{C,l}(k,p)$ stands for the half-shell Jost function of Ref. 20 and $F_C^{(+)}(l+\frac{1}{2},p,k)$ is the half-shell Jost function (21) corresponding to the Coulomb potential (46). Note that $zi^{-l}h_l^{(+)}(z)=\hat{h}^{+}(l+\frac{1}{2},z)$, where $h_l^{(+)}(z)$ is the spherical Hankel function of Ref. 20. The factor $(p/k)^l$, equal to unity if $p\to k$, is, however, unimportant and the Jost function $f_{C,l}(k,p)$ has to satisfy Eq. (44) with [see Eqs. (32) and (34)]

$$R_{C-}(k,p)=\Gamma(1+i\gamma)[-i(p^2-k^2+i0)/4k^2]^{-i\gamma}. \tag{47}$$

Indeed, this can easily be verified. The analytic expression obtained in Ref. 20 [Eq. (10)] is

$$f_{C,l}(k,p)=1+\binom{l+i\gamma}{l}^{-1}\binom{l-i\gamma}{l}^{-1}\Big[-\mathscr{A}_l\Big(\frac{p^2}{k^2};\gamma^2\Big)+\Big(\frac{p}{k}\Big)^l\Big(\frac{p+k}{p-k+i0}\Big)^{i\gamma}P_l^{(-i\gamma,i\gamma)}\Big(\frac{p^2+k^2}{2pk}\Big)\Big], \tag{48}$$

where $P_l^{(\alpha/\beta)}$ denotes the Jacobi polynomials[17] and $\mathscr{A}_l(x,y)$ is a polynomial with real coefficients of lth degree in both variables such that [Eq. (9) of Ref. 20]

$$\mathscr{A}_l(1;\gamma^2)=\binom{l+i\gamma}{l}\binom{l-i\gamma}{l}. \tag{49}$$

Inserting Eq. (49) into Eq. (48) taken for $p\to k$ and using

$$P_l^{(-i\gamma,i\gamma)}(1)=\binom{l-i\gamma}{l},$$

we get

$$f_{C,l}(k,p)\sim\Big(\frac{2k}{p-k+i0}\Big)^{i\gamma}\binom{l+i\gamma}{l}^{-1},\quad p\to k.$$

If we renormalize this relation according to Eq. (44) using the renormalization factor (47) we obtain

$$\lim_{p\to k}R_{C-}^{-1}(k,p)f_{C,l}(k,p)=e^{\pi/2}\frac{\Gamma(l+1)}{\Gamma(l+1+i\gamma)}\equiv f_{C,l}^{(k)},$$

which is the correct Coulomb on-shell Jost function.

V. CONCLUSIONS

We have found the on-shell singularities of the off-shell Jost solutions and off- and half-shell Jost functions for central long-range potentials. We have shown that all the singularities are expressed by the renormalization factor for the incoming wave function and by its complex conjugate. The results are assembled in Table I.

For Coulomb interaction, the found expressions reproduce the already known results. In particular, the singularities of the half-shell Jost functions $F^{(\pm)}(\lambda,k,p)$ show that the long-range half-shell scattering amplitude always contains a singular term independent of the angular momentum.

The found relations can serve to determine the on-shell singularities of other quantities used in the scattering theory, such as the standing wave solution and the matrix elements of the reaction operator.[12]

[1] J. D. Dollard, J. Math. Phys. **5**, 729 (1964).
[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1979), Vol. III, Chap. XI, Sec. 9.
[3] J. Zorbas, Nuovo Cimento Lett. **10**, 121 (1974).
[4] J. Dušek, Czech. J. Phys. B **31**, 941 (1981).
[5] J. Dušek, Czech. J. Phys. B **32**, 1325 (1982).
[6] H. van Haeringen, J. Math. Phys. **17**, 995 (1976).
[7] J. R. Taylor, Nuovo Cimento B **23**, 313 (1974).
[8] M. G. Fuda and J. S. Whiting, Phys. Rev. C **8**, 1225 (1973).
[9] M. G. Fuda, Phys. Rev. C **14**, 37 (1976).
[10] J. Y. Pasquier and R. Pasquier, Ann. Phys. **111**, 269 (1978).
[11] L. Trlifaj, Czech. J. Phys. B **32**, 344 (1982).
[12] L. Trlifaj, Czech. J. Phys. B **31**, 969 (1981).
[13] J. Dušek, Czech. J. Phys. B **33**, 593 (1983).
[14] V. de Alfaro and T. Regge, *Potential Scattering* (North Holland, Amsterdam 1965).
[15] Reference 2, Theorem XI. 74.
[16] J. R. Taylor, *Scattering Theory* (Wiley, New York, 1972).
[17] H. Bateman and A. Erdélyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II.
[18] E. B. Manoukian and E. Prugovečki, Can. J. Phys. **49**, 102 (1971).
[19] J. Zorbas, J. Math. Phys. **19**, 177 (1978).
[20] H. van Haeringen, J. Math. Phys. **19**, 1379 (1978).

Time delay and dilation properties in scattering theory

Heide Narnhofer
Institut für Theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria

(Received 5 October 1982; accepted for publication 13 May 1983)

We show the identity of the Eisenbud–Wigner time delay and that of Lavine for potentials $V = V_1 + V_2$, $V_1 \sim 1/|x|^{1+\epsilon}$, $x\nabla V_1 \sim 1/|x|^{1+\epsilon}$, $V_2 \sim 1/|x|^{2+\epsilon}$ for $x \to \infty$. Finally we show the additivity of time delay for the same class of potentials.

PACS numbers: 03.80. + r

I. INTRODUCTION

In a recent paper,[1] we discussed the connection between the Eisenbud–Wigner time delay, the time delay of Lavine, and expectation values of the dilation operator. Since the ideas behind these connections follow completely classical considerations, all relations find their counterpart in classical scattering theory.[2] In classical theory, the basic property is the convergence of $x(t) - p(t)t$, which corresponds to the fact that $x\nabla V$ and V have to be integrable in time or $x\nabla V$ and V have to decrease faster than $1/|x|^{1+\epsilon}$.

In Ref. 1, domain problems are ignored. Thus the results could not be accepted to be rigorous. In the mean time, other results on the connection of the different definitions of time delay were obtained.[3–5] But the assumptions on the potential had to be stronger than for the classical problem. As is shown in Ref. 3, the Eisenbud–Wigner time delay $S^{-1}(\lambda)[dS(\lambda)/d\lambda]$ exists[6] if either V and $x\nabla V$ decrease faster than $1/|x|^{1+\epsilon}$ or if V decreases faster than $1/|x|^{2+\epsilon}$, whereas for the other definitions, stronger decrease properties on V are necessary. It seems worthwhile to improve the results so that equality holds in all cases when the Eisenbud–Wigner time delay is defined. Thus we learn that scattering is essentially to be understood in classical terms. Evolution of expectation values of operators behave essentially in the same way. Of course in the proof, we have to take into account the noncommutativity, but as a matter of fact, it only improves the limit behavior because, due to the uncertainty principle, local fluctuations are less important in quantum than in classical theory and, therefore, conditions on $x\nabla V$ can be replaced by conditions on decrease properties of V.

Finally, we want to mention the definition of time delay given by Jauch *et al.*,[7,8] namely,

$$\tau = \lim_{R\to\infty} \tau_R = \lim \int_{-\infty}^{+\infty} dt\,(\|P_R e^{-iHt} \times \Omega_- \phi\|^2 - \|P_R e^{-iH_0 t}\phi\|^2),$$

with P_R the projection operator on the ball with radius R. As shown in Ref. 3 it coincides with the Eisenbud–Wigner time delay if V decreases at infinity as fast as $1/|x|^{4+\epsilon}$. For radial symmetric potentials, the result can be improved up to $1/|x|^{2+\epsilon}$ (Ref. 5) due to our better knowledge of the generalized eigenfunctions. We have not been able to show that equality also holds in the above cases and, in fact, we believe that really the $1/|x|^{2+\epsilon}$ could be the relevant border line since diffusion effects might become important.

II. THE EISENBUD–WIGNER TIME DELAY AND THE DILATION OPERATOR

The Eisenbud–Wigner time delay is defined by the decomposition in the spectral representation of H_0 to be[6]

$$T = iS(\lambda)^* \frac{dS(\lambda)}{d\lambda}.$$

Evidently it commutes with H_0. It is well defined if $V = V_1 + V_2$, where V_1 and V_2 are H_0-relatively bounded and V_1 decreases as $1/|x|^{2+\epsilon}$ and V_2 and $x\nabla V_2$ as $1/|x|^{1+\epsilon}$ for $|x|\to\infty$. This is shown in Ref. 3 for V_1 or V_2 but the generalization for the sum is rather obvious.

Using the representation of the dilation operator in the spectral representation of H_0, it is also shown there that

$$TH_0 = -\tfrac{1}{2}S^*[D,S].$$

In Ref. 1 we did not define the time delay as an operator acting on the incoming wave function, but according to the definition in classical scattering theory as form τ defined on the actual wave function at $t = 0$, namely, with $\phi\in\mathscr{H}_{ac}(H)$

$$\langle\phi|H\tau|\phi\rangle = \lim_{t\to\infty} \langle\phi|e^{iHt}e^{-iH_0t}De^{iH_0t}e^{-iHt} - e^{-iHt}e^{iH_0t}De^{-iH_0t}e^{iHt}|\phi\rangle.$$

Under the above assumptions on V, the limit exists for the dense class of $\phi = g(H)(1/(D+i))\psi$ with $g\in C_0^1((0,\infty)/\sigma_p(H))$ as bounded form and thus corresponds to the operator

$$H\tau = \Omega_- TH_0\Omega_-^*$$

according to the following results.

Lemma 1: Let V and $x\nabla V$ be relatively bounded with respect to H. Let $\psi = g(H)(1/(D+i))\phi$ with $g\in C_0^1$. Then $e^{iHt}\psi\in\mathscr{D}(D)\ \forall\ t$.

Similar results can be obtained for D^k with the assumption that $g\in C_0^k$ and $[D,V]^l$ is relatively bounded for all $l\leqslant k$.[9] $[D,V]^l = [D[D,V]^{(l-1)}]$.

Lemma 2:

$$e^{iHt}g(H_0)(1/(D+i))\phi\in\mathscr{D}(D).$$

This is shown in Ref. 10. For completeness we state the following.

Lemma 3: Let V decrease faster than $1/|x|^{1+\epsilon}$ for $|x|\to\infty$. Then

$$\langle\phi|(1/(D+i))g(H_0)e^{iHt}De^{-iHt} \times g(H_0)(1/(D+i))|\phi\rangle < \infty \quad \forall \quad t,$$

 0022-2488/84/040987-05$02.50

$$\langle\phi\,|(1/(D+i))g(H_0)e^{iHt}e^{iH_0t}De^{-iHt}e^{-iH_0t}$$
$$\times g(H)(1/(D+i))|\phi\rangle<\infty\quad\forall\quad t.$$

Thus the weaker assumption on V only allows statements for the form.

Proof: The expression equals, e.g.,

$$\langle\phi\,|(1/(D+i))g(H_0)Dg(H_0)(1/(D+i))|\phi\rangle$$
$$-\int_0^t dt'\langle\phi\,(1/(D+i))g(H_0)e^{iH(t-t')}$$
$$(2H_0-i[D,V])e^{-H(t-t')}g(H_0)(1/(D+i))|\phi\rangle.$$

Since p and $x\cdot V$ are relatively bounded, the expression is bounded. We have not been able to prove that $e^{iHt}e^{-iH_0t}De^{iH_0t}e^{-iHt}$ is bounded uniformly in t as a form on an appropriate domain under this weak assumption on V. Thus we have to return to the stronger conditions on V.

Lemma 4: Let V_1 and $x\nabla V_1$ decrease faster than $1/|x|^{1+\epsilon}$ for $|x|\to\infty$, and V_2 faster than $1/|x|^{2+\epsilon}$. Then,

$$D\Omega\,(1/(D+i))$$

is a bounded operator.

Proof: We have to estimate $[D,\Omega]$. We use stationary formalism for the wave operator Ω which is well defined due to the properties of the resolvent as bounded operator between weighted $\mathscr{L}^2$ spaces and the equality[11]

$$\delta(H_0-E)=\frac{1}{2\pi i}\left(\frac{1}{H-E-i0}-\frac{1}{H-E+i0}\right).$$

$$i\int_0^\infty dE\left[D,\frac{1}{H-E-i0}V\delta(H_0-E)\right]$$
$$=\int_0^\infty dE\frac{d}{d\alpha}\left(\frac{1}{e^{2\alpha}H_0+V(\alpha)-E-i0}\right.$$
$$\left.\left.\times V(\alpha)\delta(e^{2\alpha}H_0-E)\right)\right|_{\alpha=1}$$
$$=\int_0^\infty dEe^{2\alpha}\left(\frac{d}{d\alpha}\frac{e^{-2\alpha}}{H_0+e^{-2\alpha}V(\alpha)-E-i0}\right.$$
$$\left.\left.\times V(\alpha)e^{-2\alpha}\delta(H_0-E)\right)\right|_{\alpha=1}$$
$$=\int_0^\infty dE\left(\frac{1}{H_0+V-E-i0}(2V-i[D,V])\right.$$
$$\times\frac{1}{H_0+V-E-i0}V-\frac{1}{H_0+V-E-i0}V$$
$$\left.\times(4V-i[D,V])\right)\delta(H_0-E).$$

We have formally used that $V(\alpha)$ is well defined and differentiable, but the last equality can be shown without this assumption. Now either $x\nabla V_1$ or $x\cdot V_2$ decrease faster than $1/|x|^{1+\epsilon}$ and therefore control the resolvent, respectively, the δ-function, by standard arguments.[11,12]

Lemma 5:

$$\lim_{t\to\pm\infty}e^{iHt}e^{-iH_0t}De^{iH_0t}e^{-iHt}P_{ac}=\Omega_\pm D\Omega_\pm^*,$$

Proof: By the same arguments as before, $D\Omega^*(1/(D+i))$ is a bounded operator. The limit is defined for $\Omega_\pm g(H_0)Dg(H_0)\Omega_\pm^*$ on a domain where the operator is essentially self-adjoint and, finally, we can use the fact that $\lim_{g\to1}g(H_0)Dg(H_0)=D$ in the strong resolvent sense.

III. THE TIME DELAY OF LAVINE

The time delay is defined by Lavine[13] to be

$$\langle\phi\,|H\tilde{\tau}|\phi\rangle=\int_{-\infty}^{+\infty}dt\,\langle\phi\,|e^{iHt}(2V+x\nabla V)e^{-iHt}|\phi\rangle.$$

Its connection with dilation can be seen by

$$e^{iHt}e^{-iH_0t}De^{iH_0t}e^{-iHt}$$
$$=e^{iHt}(D-2tH_0)e^{-iHt}$$
$$=\int_0^t dt'\,e^{iH(t-t')}(2H-2V-x\nabla V)e^{iH(t-t')}$$
$$+D-2te^{iHt}H_0e^{-iHt}$$
$$=D+2te^{iHt}Ve^{-iHt}+\int_0^t e^{iHt'}$$
$$\times(2V-i[D,V])e^{-iHt'}dt'.$$

Thus we have to show that $\lim_{t\to\infty}t\,||Ve^{-iHt}\phi\,||=0$, and therefore to consider the decrease properties of $||Ve^{-iHt}\phi\,||$.

Lemma 6: Let $V(x)$ be a dilation analytic operator.[11] Then with $g\in C_0^\infty$,

$$||(1/(1+x^2))^{\mu/2}e^{-iHt}g(H)(1/(D^2+1)^{\mu/2})||\leqslant c_\mu|t|^{-\mu+\epsilon}.$$

Remark: The above result corresponds essentially to the fact that $x/t-p\to0$ in the strong resolvent sense, though this convergence does not rigorously imply the above behavior. To obtain a uniform estimate in t we have to ensure that we start with a wave function at $t=0$ that is sufficiently concentrated so that, for large t, the scattering process is essentially finished. In Refs. 9 and 13, similar results can be found, except that the restriction on the wave function was different ($\in P_+(D)\mathscr{H}$) so that the result was not uniform for positive and negative t. We will just copy their techniques. In Ref. 3, a somewhat stronger result for H_0 was obtained, especially without the ϵ, but it is harder to generalize it for dilation analytic potentials.

Proof: We use the fact that

$$\frac{1}{(D^2+1)^{m/2}}e^{-iHt}g(H)\frac{1}{(D^2+1)^{m/2}}$$
$$=\frac{n!}{2\pi i(it)^n}\int_{-\infty}^{+\infty}dE\,\bar{e}^{iEt}\frac{1}{(D^2+1)^{m/2}}$$
$$\times\frac{1}{(H-E\pm i0)^{n+1}}g(H)\frac{1}{(D^2+1)^{m/2}}\quad\text{for }t\gtrless0.$$

Now we consider the analytic operator valued function

$$F(\alpha)=\frac{1}{(D^2+1)^{m/2}}\frac{1}{(H(\alpha)-E-i0)^{n+1}}\frac{1}{(D^2+1)^{m/2}},$$

and obtain

$$\left|\left|\frac{dF(\alpha_0+i\alpha)}{d\alpha}\right|\right|$$
$$=2\left|\left|\frac{D}{(D^2+1)^{m/2}}\frac{1}{(H(\alpha)-E-i0)^{n+1}}\frac{1}{(D^2+1)^{m/2}}\right|\right|.$$

Now we proceed as in Refs. 12 and 13, and conclude from their results that

$$\left|\left|\frac{1}{(D^2+1)^{m/2}}\frac{1}{(H-E-i0)^{m'}}\frac{1}{(D^2+1)^{m/2}}\right|\right|\leqslant c(m,m')$$

for $m'<m$, m' m integer and, therefore,

$$\left|\left|\frac{1}{(D^2+1)^{m/2}}e^{-iHt}g(H)\frac{1}{(D^2+1)^{m/2}}\right|\right| \leqslant c(m,\epsilon)\left(\frac{1}{t}\right)^{m-1-\epsilon}.$$

As in Refs. 9 and 14, we can interpolate with respect to m, where for Re $m=0$, we have the bound 1 and for Re $(m+k)$ the bound $(1/t)^{m+k-1-\epsilon}$. Therefore with $m=0\times k/(m+k)+(m/(m+k))(m+k)$,

$$G(m)||\leqslant||G(m+k)||^{m/(m+k)} = c(m+k)(1/t)^{m(m+k-1-\epsilon)/(m+k)},$$

and for large enough k, we obtain

$$\left|\left|\frac{1}{(D^2+1)^{m/2}}e^{-iHt}g(H)\frac{1}{(D^2+1)^{m/2}}\right|\right|\leqslant\bar{c}(m,\epsilon)\left(\frac{1}{t}\right)^{m-\epsilon}$$

or, since p is dominated by H and therefore $(1/(1+x^2)^{m/2})$ $g(H)(D^2+1)^{m/2}$ bounded (see Ref. 9),

$$\left|\left|\frac{1}{(1+x^2)^{m/2}}e^{-iHt}g(H)\frac{1}{(D^2+1)^{m/2}}\right|\right|\leqslant\hat{c}(m,\epsilon)\left(\frac{1}{t}\right)^{m-\epsilon}.$$

Thus for dilation analytic potentials that decrease as $1/|x|^{1+\epsilon}$,

$$\lim_{t\to\infty} t\,||Ve^{-iHt}g(H)\frac{1}{(D^2+1)^{1/2+\epsilon}}||=0$$

and, therefore, the different definitions of time delay coincide. To find the corresponding estimate for other potentials we use

$$e^{iHt}\Omega=\Omega e^{iH_0t}$$

and, thus, we need the following lemma.

Lemma 7: Let V and $x\nabla V$ decrease faster than $1/|x|^{1+\epsilon}$, or V decrease faster than $1/|x|^{2+\epsilon}$. Then for $\delta<\epsilon$,

$$|D|^{1+\delta}\Omega g(H_0)1/(D^2+1)^{(1+\delta)/2},$$
$$|D|^{1+\delta}\Omega^*g(H)1/(D^2+1)^{(1+\delta)/2}$$

are bounded operators.

Remark: To start with, the expressions exist as densely defined quadratic forms. Since they turn out to be bounded on their domains, they correspond to bounded operators.

Proof: We concentrate on the first operator, the calculation works as well for the other one. Using the methods of the proof of Lemma 4, the expression differs only by bounded operators from

$$[D^\delta,[D,\Omega]]g(H_0)\frac{1}{(D^2+1)^{(1+\delta)/2}}$$
$$=-i\left[D^\delta,\int\frac{1}{H-E-i0}\left\{(2V-i[D,V])\frac{1}{H-E-i0}V -4V+i[D,V]\right\}\delta(H_0-E)\right]g(H_0)\frac{1}{(D^2+1)^{(1+\delta)/2}},$$

where $D^\delta:=\operatorname{sgn}D|D|^\delta$.

Now we consider the operator-valued function analytic in γ for $0<\operatorname{Re}\gamma<1$:

$$-i\Big[D^\gamma\int dE\frac{1}{H-E-i0}\Big\{(2V^{(1+\gamma)/(1+\delta)} -(i[D,V])^{(1+\gamma)/(1+\delta)})\frac{1}{H-E-i0}V^{(1+\gamma)/(1+\delta)} -4V^{(1+\gamma)/(1+\delta)}+(i[D,V])^{(1+\gamma)/1+\delta}\Big\}\delta(H_0-E)\Big]$$

if V and $i[D,V]$ decrease faster than $1/|x|^{1+\epsilon}$ (again $W^\gamma:=\operatorname{sgn}W|W|^\gamma$), or

$$-i\Big[D^\gamma,\int dE\frac{1}{H-E-i0}\{(2V^{(1+\gamma)/(1+\delta)} -i[D,V^{(1+\gamma)/1+\delta)}])\frac{1}{H-E-i0}V^{(1+\gamma)/(1+\delta)} -4V^{(1+\gamma)/(1+\delta)}+i[D,V^{(1+\gamma)/(1+\delta)}]\}\delta(H_0-E)\Big]$$

if V decreases faster than $1/|x|^{2+\epsilon}$.

In both cases we have an expression

$$[D^\gamma,A(\gamma)],$$

where $A(\gamma)$ are bounded operators. We now have to investigate the boundaries of the strip, i.e., Re $\gamma=0$ and Re $\gamma=1$. In the first case, the commutator $[\operatorname{sgn}D|D|^{i\operatorname{Im}\gamma},A(\gamma)]$ is bounded since it involves only bounded operators (we take $\delta<\epsilon$). On the other side for Re $\gamma=1$ we have, besides the bounded operator, $|D|^{i\operatorname{Im}\gamma}$ terms of the type

$$\int dE\frac{1}{H-E-i0}\times[D,2V^{2/(1+\delta)}-(i[D,V])^{2/(1+\delta)}]\delta(H_0-E)$$

or

$$\int dE\frac{1}{H-E-i0}[D,[D,V^{2/(1+\delta)}]]\delta(H_0-E).$$

They are of the form

$$\int dE\frac{1}{H-E-i0}pW\delta(H_0-E)$$

or, in the worst case,

$$\int dE\frac{1}{H-E-i0}p^2W\delta(H_0-E),$$

where W decreases faster than $1/|x|^{1+\epsilon-\delta}$ and therefore, by standard arguments, the operator is well defined and bounded (Ref. 11; XIII, 8).

For the commutator with the remaining terms we study

$$\frac{d}{d\alpha}\int dE\frac{1}{e^{2\alpha}H_0+V-E-i0}V\delta(e^{2\alpha}H_0-E),$$

and this is bounded as we have seen already in Lemma 4. Thus by interpolation, $[D^\gamma,A(\gamma)]$ is bounded for all γ in the strip and, therefore, especially $[D^\delta,A(\delta)]$ which coincides with $[D^\delta[D,\Omega]]$.

We finally obtain

$$\left\|\frac{1}{(D^2+1)^{(1+\delta)/2}}e^{-iHt}g(H)\phi\right\|$$
$$\leqslant\left\|\frac{1}{(D^2+1)^{(1+\delta)/2}}\Omega g(H_0)^{1/2}(D^2+1)^{(1+\delta)/2}\right\|$$
$$\times\left\|\frac{1}{(D^2+1)^{(1+\delta)/2}}g^{1/2}(H_0)\right.$$
$$\left.\times e^{-iH_0t}\Omega^*\phi\right\|\leqslant ct^{-1-\delta}.$$

By the same argument, we have that $1/(x^2+1)^{m/2} g(H)(D^2+1)^{m/2}$ is bounded[9] and thus for all ϕ in H, there exists a $\bar{c}$ such that

$$\left\|\frac{1}{(1+x^2)^{(1+\epsilon)/2}}e^{-iHt}g(H)\frac{1}{(D^2+1)^{(1+\delta)/2}}\phi\right\|\leqslant\bar{c}t^{-1-\delta}.$$

Therefore we can state our final theorem.

Theorem: Assume that $V=V_1+V_2$, V_1 and $x\nabla V_1$ relatively bounded and decreasing faster than $1/|x|^{1+\epsilon}$, V_2 relatively bounded and decreasing faster than $1/|x|^{2+\epsilon}$. Then,

(1) $H_0S^{-1}[D,S]$ exists,

(2) $\lim e^{iHt}e^{-iH_0t}De^{iH_0t}e^{-iHt}P_{ac}=D_\pm$ exists in the strong resolvent sense,

(3) $\Omega_-H_0S^{-1}[D,S]\Omega_-^*=D_+-D_-=\int dt\, e^{iHt}(2V+x\nabla V)e^{-iHt}=H\tau.$

So far, the result seems to be optimal. Generalization for the Coulomb potential with modified Moeller operators are obvious.[5] Our result improves those of Ref. 5 insofar as we are not restricted to the spherical symmetric potential.

IV. ADDITIVITY OF TIME DELAY AND PHASE SHIFT

The additivity of the phase shift for different potentials is a well-known phenomenon if one works with Jost functions.[16] If the potentials are not spherically symmetric, the phase shift has to be regarded as an operator acting in different spaces (constant energy for different Hamiltonians) and therefore additivity does not make sense *a priori*. Additivity of time delay as expectation values with certain wave functions can be handled more easily: We observe, with $S_{ik}(\lambda)$ the scattering matrix between the Hamiltonian H_i and H_k in the spectral representation of H_k,

$$\langle\phi_{0\text{in}}|i\lambda S_{20}^{-1}(\lambda)\frac{d}{d\lambda}S_{20}(\lambda)|\phi_{0\text{in}}\rangle$$
$$=\langle|\Omega_{20}^+D\Omega_{20}^{+*}-\Omega_{20}^-D\Omega_{20}^{-*}|\phi\rangle$$
$$=\langle\phi|\Omega_{21}^+\Omega_{10}^+D\Omega_{10}^{+*}\Omega_{21}^{+*}$$
$$-\Omega_{21}^+\Omega_{10}^-D\Omega_{10}^{-*}\Omega_{21}^{+*}|\phi\rangle$$
$$+\langle\phi|\Omega_{21}^+\Omega_{10}^-D\Omega_{10}^{-*}\Omega_{21}^{+*}$$
$$-\Omega_{21}^-\Omega_{10}^-D\Omega_{10}^{-*}\Omega_{21}^{-*}|\phi\rangle$$
$$=\langle\phi_{0,\text{in}}|D-S_{10}^{-1}DS_{10}|\phi_{0,\text{in}}\rangle$$
$$+\langle\phi_{\text{in}}^1|\Omega_{10}^-D\Omega_{10}^{-*}$$
$$-S_{21}^{-1}\Omega_{10}^-D\Omega_{10}^{-*}S_{21}|\phi_{1,\text{in}}\rangle.$$

Now S_{21} is commuting with H_1 with the decomposition $S_{21}(\lambda)$ in the spectral representation of H_1. Therefore $\Omega_{10}^{-*}S_{21}\Omega_{10}^-$ has the same decomposition in the spectral representation of H_0. We denote the corresponding operator with $\hat{S}_{21}(\lambda)$:

$$\langle\phi_{0,\text{in}}|-i\lambda S_{10}^{-1}(\lambda)\frac{d}{d\lambda}S_{10}(\lambda)|\phi_{0,\text{in}}\rangle$$
$$+\langle\phi_{0,\text{in}}|-i\lambda S_{21}^{-1}(\lambda)\frac{d}{d\lambda}\hat{S}_{21}(\lambda)|\phi_{0,\text{in}}\rangle$$
$$=\langle\phi_{0,\text{in}}|-i\lambda S_{10}^{-1}(\lambda)\frac{d}{d\lambda}S_{10}(\lambda)|\phi_{0,\text{in}}\rangle$$
$$+\langle\phi_{1,\text{in}}|-i\lambda S_{21}^{-1}(\lambda)\frac{d}{d\lambda}S_{21}(\lambda)|\phi_{1,\text{in}}\rangle.$$

If we assume that the potentials are spherically symmetric, this corresponds to

$$\frac{d\delta_{20}(l,k)}{dk}=\frac{d\delta_{21}(l,k)}{dk}+\frac{d\delta_{10}(l,k)}{dk}.$$

The additivity of the phase shift follows if we can show that the time delay converges faster than $E^{-1-\delta}$ and is thus integrable at infinity. But this follows from estimating the time delay given by Lavine's formula (Theorem 3). It is known that

$$\left\|\frac{1}{1+|x|^{(1+\epsilon)/2}}\frac{1}{H_0-E-i0}\frac{1}{1+|x|^{(1+\epsilon)/2}}\right\|$$
$$<\frac{1}{E^\delta}\quad\text{for }\frac{1}{2}>\delta>0$$

in one dimension,[17]

$$\left\||x|^{-(1+\epsilon)/2}\frac{1}{H_0-E-i0}|x|^{-(1+\epsilon)/2}\right\|<c\frac{1}{E^{1-\epsilon}}$$

in other dimensions[18] and scaling arguments. Therefore $1/(1+|V|^{1/2}(1/(H_0-E-i0))V^{1/2})$ is uniformly bounded for these E and therefore

$$\left\|V^{1/2}\frac{1}{H-E-i0}|V|^{1/2}\right\|<c\frac{1}{E^\delta}.$$

It follows that $|\tau|<c(1/E^{1+\delta})$.

$$\delta(l,k)=-\int_k^\infty dk'\,\tau(l,k)=+i\int_k^\infty S_l^{-1}(k)\frac{d}{dk}S_l(k)dk,$$

with $\lim_{k\to\infty}S_l(k)=1$.

ACKNOWLEDGMENT

I want to thank Professor W. Thirring for critical comments on the manuscript.

[1] H. Narnhofer, Phys. Rev. D 22, 2387 (1980).
[2] H. Narnhofer and W. Thirring, Phys. Rev. A 23, 1688 (1980).
[3] A. Jensen, Commun. Math. Phys. 82, 435 (1981).
[4] A. Jensen, "On Lavine's Formula for Time Delay," Preprint CPT/CNRS Marseille, 1982.
[5] D. Bollé, F. Gesztesy, and H. Grosse, "Time delay for long range interactions," J. Math. Phys. 24, 1529 (1983).
[6] E. Wigner, Phys. Rev. 98, 145 (1955).
[7] J. M. Jauch, B. Misra, and K. B. Sinha, Helv. Phys. Acta 45, 398 (1972).
[8] Ph. Martin, Acta Phys. Austriaca. Suppl. 23, 157 (1981).
[9] P. A. Perry, Commun. Math. Phys. 81, 243 (1981).
[10] E. Mourre, Commun. Math. Phys. 78, 391 (1981).
[11] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV* (Academic, New York, 1978).
[12] R. Lavine, Commun. Math. Phys. 20, 301 (1971).

[13]R. Lavine, "Commutators and Local Decay," in *Scattering Theory in Mathematical Physics*, edited by A. Lavita and J. R. Marchard (Reidel, Holland, 1974), p. 141.
[14]E. Mourre, Commun. Math. Phys. **68**, 91 (1979).
[15]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II* (Academic, New York, 1975).
[16]F. Gesztesty, W. Plessas, and B. Thaller, J. Phys. A **13**, 2659 (1980).
[17]S. Agmon, Ann. Scuola Norm. Sup. Pisa, Ser. IV **2**, 151 (1975).
[18]W. Thirring, *Lehrbuch der mathematischem Physik III* (Springer, Wien, 1979).

The total effective mass of the Kerr–Newman metric

Jeffrey M. Cohen
Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104
F. de Felice
Dipartimento di Fisica "G. Galilei," Universitá di Padova, Padova, Italy and International School for Advanced Studies, Trieste, Italy

(Received 29 June 1983; accepted for publication 17 October 1983)

We find the total effective gravitational mass which acts on a particle moving in the Kerr–Newman metric. We also identify the contribution to that mass by the electric field energy which is stored in any given portion of the space surrounding the metric source.

PACS numbers: 04.20.Cv, 04.20.Me

I. INTRODUCTION

The repulsive nature of timelike singularities is a general relativistic effect which has no Newtonian analog. The physical understanding of this effect, which becomes particularly evident when one studies the classical motion, resides in the correct identification of the total effective gravitational mass which acts on the particle paths.

This goal was attained in the Reissner–Nordström metric,[1,2] where it was found that a test particle, a coordinate distance R apart from the metric source of mass M and charge Q, experiences the gravitational field of a varying mass, i.e.,

$$M_T = M - Q^2/R. \tag{1}$$

As R decreases, M_T also decreases because the electric-field energy inside a sphere of radius R decreases.

At radii less than $R = Q^2/M$, the particle experiences a negative gravitating mass which will act repulsively. In this paper we generalize (Sec. II) Eq. (1) to the Kerr–Newman metric; we also identify in Sec. III the amount of the electric-field energy which is stored in any given source-free portion of the space-time and which contributes gravitationally to the total effective mass.

II. THE TOTAL EFFECTIVE MASS

In a space-time endowed with symmetries, the conservation law of general relativity, i.e.,

$$T^{j}_{i;j} = 0 \tag{2}$$

yields conserved quantities. If ξ^i is a Killing vector, then Komar's conserved quantity can be written in the coordinate-free form[3-4]

$$16\pi I = \int_{\partial\Sigma} {}^*d\xi, \tag{3}$$

where ${}^*d\xi$ means the dual to the two-form $d\xi$ and ξ is the Killing one-form. In (3) the integration is extended over a spacelike surface $\partial\Sigma$ of the background metric. The use of Eq. (3) leads to correct answers; in the Kerr metric, for example, if ξ is the axisymmetric Killing one-form, Eq. (3) yields the angular momentum of the metric source.[3] In our case we shall consider ξ as the stationary Killing one-form over a Kerr–Newmann background metric.

In Boyer and Lindquist coordinates, this reads

$$ds^2 = -\left(\frac{\Delta\Sigma}{A}\right)dt^2 + \left(\frac{A\sin^2\theta}{\Sigma}\right)[d\phi + \Omega dt]^2 + \left(\frac{\Sigma}{\Delta}\right)dr^2 + \Sigma d\theta^2, \tag{4}$$

where

$$\begin{aligned} \Omega &= (a/A)(Q^2 - 2Mr) \equiv (a/A)\beta, \\ \Sigma &= r^2 + a^2\cos^2\theta, \\ \Delta &= r^2 + a^2 + Q^2 - 2Mr, \\ A &= (r^2+a^2)\Sigma - a^2(Q^2 - 2Mr)\sin^2\theta. \end{aligned} \tag{5}$$

Let us now introduce the orthonormal frame of one-forms:

$$\begin{aligned} \underset{\sim}{\theta}_{\hat 0} &= -(\Delta\Sigma/A)^{1/2}\,dt, \\ \underset{\sim}{\theta}_{\hat 1} &= (\Sigma/\Delta)^{1/2}\,dr, \\ \underset{\sim}{\theta}_{\hat 2} &= \Sigma^{1/2}\,d\theta, \\ \underset{\sim}{\theta}_{\hat 3} &= (A\sin^2\theta/\Sigma)^{1/2}\,(d\phi + \Omega\,dt); \end{aligned} \tag{6}$$

in terms of these, the time–Killing one-form $\xi = g_{i0}dx^i$ reads

$$\xi = \left(\frac{\Delta\Sigma}{A}\right)^{1/2}\underset{\sim}{\theta}_{\hat 0} + \frac{a\sin\theta}{(\Sigma A)^{1/2}}(Q^2 - 2Mr)\underset{\sim}{\theta}_{\hat 3}. \tag{7}$$

Differentiation of (7) yields, after some algebra

$$d\xi = \tilde f \underset{\sim}{\theta}_{\hat 1}\Lambda\underset{\sim}{\theta}_{\hat 0} + \tilde g\underset{\sim}{\theta}_{\hat 2}\Lambda\underset{\sim}{\theta}_{\hat 0} + \tilde h\underset{\sim}{\theta}_{\hat 1}\Lambda\underset{\sim}{\theta}_{\hat 3} + \tilde K\underset{\sim}{\theta}_{\hat 2}\Lambda\underset{\sim}{\theta}_{\hat 3}; \tag{8}$$

here we have

$$\tilde f = -(2A^{1/2}/\Sigma^3)(M\Sigma + \beta r)(1 + a\Omega\sin^2\theta), \tag{9}$$

$$\tilde g = a\Omega(A^3/\Sigma^6\Delta)^{1/2}\sin 2\theta\left[1 + \left(\frac{r^2+a^2}{a}\right)\Omega\right], \tag{10}$$

$$\tilde h = -(\Delta/A)^{1/2}(2a/\Sigma^2)(M\Sigma + \beta r)\sin\theta, \tag{11}$$

$$\tilde K = \frac{2A^{1/2}\Omega\cos\theta}{\Sigma^2}(r^2 + a^2). \tag{12}$$

The dual of (8) reads

$${}^*d\xi = -\tilde f\underset{\sim}{\theta}_{\hat 2}\Lambda\underset{\sim}{\theta}_{\hat 3} + \tilde g\underset{\sim}{\theta}_{\hat 1}\Lambda\underset{\sim}{\theta}_{\hat 3} + \tilde h\underset{\sim}{\theta}_{\hat 2}\Lambda\underset{\sim}{\theta}_{\hat 0} - \tilde K\underset{\sim}{\theta}_{\hat 1}\Lambda\underset{\sim}{\theta}_{\hat 0}, \tag{13}$$

where we used the convention $(0, 1, 2, 3) = +1$. Finally expressing (13) in terms of the coordinates, we have

$$
\begin{aligned}
{}^{*}d\xi = {} & [\, -\tilde{f}A^{1/2}\Omega \sin\theta - \tilde{h}\,\Sigma(\Delta/A)^{1/2}]\, d\theta \wedge dt \\
& + [\tilde{g}(A/\Delta)^{1/2}\Omega \sin\theta + \tilde{K}\,\Sigma(1/A^{1/2})]\, dr \wedge dt \\
& - \tilde{f}A^{1/2} \sin\theta \, d\theta \wedge d\phi \\
& + \tilde{g}(A/\Delta)^{1/2} \sin\theta \, dr \wedge d\phi. \qquad (14)
\end{aligned}
$$

To perform the integration in (3), let us choose a surface $r = \text{const}$ such that the time coordinate on it satisfies

$$dt = -\frac{g_{0\phi}}{g_{00}}\, d\phi. \qquad (15)$$

In this way, infinitesimally close points are simultaneous events; this is not true for points far apart because (15) is not integrable.

After one-loop integration over ϕ, we end up at an event which is different from that we started from; *in order to have an integration over a surface of simultaneous events, we need to subtract the contribution to the integral given by the time difference between the initial and final events.* That difference amounts exactly to the one-loop integral over ϕ,[5] so the integral reduces to

$$\int_{\partial\Sigma} {}^{*}d\xi = -\iint \tilde{f}A^{1/2} \sin\theta \, d\theta \, d\phi. \qquad (16)$$

From (9) and (5), Eq. (16) yields

$$4I = 2M - \frac{Q^2}{r} - \frac{Q^2(r^2+a^2)}{ar^2} \arctan\frac{a}{r}. \qquad (17)$$

Here we identify the quantity $2I$ as the total effective gravitational mass which a particle experiences while approaching the Kerr–Newman source.

When $Q = 0$, we have $2I = M$, that is the total mass of the metric source; when $a = 0$ we have

$$2I = M - Q^2/r, \qquad (18)$$

which is the well-known result in the Reissner–Nordström metric.[1,2]

III. THE GRAVITATIONAL MASS OF THE ELECTRIC FIELD

The negative contribution to the total effective gravitational mass by the source charge, explains why at a sufficiently close distance from the metric source a particle experiences a repulsive field. Worth mentioning is that the repulsive gravitational effects in the Kerr–Newman metric are not entirely due to the charge but also to the rotational properties of the metric. The negative contributions to the effective mass by the rotation, however, do not appear explicitly in (17) because that expression was obtained by an integration over a closed surface which surrounds the source and is entirely confined to the $r > 0$ sheet of the metric. The repulsive rotational effects, we believe, arise from the net asymmetry in the source properties extending from the $r > 0$ to $r < 0$ sheet of the metric.[6]

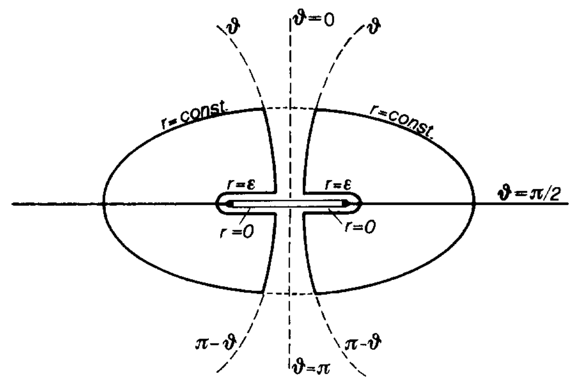


FIG. 1. The integration in (3) is performed over a closed toruslike surface which excludes the ring singularity.

Here we confine our attention to the space localization of the effective gravitational mass due to the electric field which permeates all the space. Let us consider a closed toruslike surface bounded by the surfaces $\theta = \text{const}$ and $\pi - \theta$, $r = \text{const}$ and $r = \epsilon \sim 0$ which excludes the metric source (i.e., the ring singularity) as is shown in Fig. 1. The integration in Eq. (3) will now be extended over this surface and will provide the total effective gravitating mass of the energy stored there. The integral in (3), with (17), decomposes as follows:

$$
\begin{aligned}
\int_{\partial\Sigma} {}^{*}d\xi = {} & -2\pi\Big[\int_{\theta}^{\pi-\theta} \tilde{f}A^{1/2}\sin\theta\, d\theta\Big]_{r=\text{const}} - 2\pi\Big[\int_{\pi-\theta}^{\theta} \tilde{f}A^{1/2}\sin\theta\, d\theta\Big]_{r=\epsilon} \\
& + 2\pi\Big[\int_{\epsilon}^{r} \tilde{g}\Big(\frac{A}{\Delta}\Big)^{1/2}\sin\theta\, dr\Big]_{\theta\neq\pi/2} + 2\pi\Big[\int_{r}^{\epsilon} \tilde{g}\Big(\frac{A}{\Delta}\Big)^{1/2}\sin\theta\, dr\Big]_{\pi-\theta}. \qquad (19)
\end{aligned}
$$

From (9) and (10) we readily, after some algebra and in the limit of $\epsilon \to 0$, have

$$
\begin{aligned}
4I'(r,\theta) = {} & \frac{Q^2}{r\cos\theta}\Big[1 + \frac{r}{a\cos\theta}\arctan\Big(\frac{r}{a\cos\theta}\Big)\sin^2\theta + \Big(\frac{r^2+a^2}{a^2}\Big)\Big(\frac{a\cos\theta}{r}\Big)\arctan\Big(\frac{a\cos\theta}{r}\Big)\Big]; \\
& r > 0,\ \theta \neq \pi/2. \qquad (20)
\end{aligned}
$$

As expected, it vanishes when $Q = 0$ and yields the charge term in the total effective mass in (17) when $\theta = 0$.

IV. CONCLUSION

The behavior of test particles in the Kerr–Newman space-time of being repelled by the gravity source under suitable conditions it due to the negative contribution to the total effective gravitational mass by the source rotation and charge. The effects of the latter is here derived explicitly in an expression [Eq. (17)] which generalizes a similar result obtained for the Reissner–Nordström space-time. We also derive [Eq. (20)] the gravitational mass of the electric field energy which is stored in any given portion of the space around the metric source.

ACKNOWLEDGMENT

One of us (F. de Felice) gratefully acknowledges the hospitality at the University of Pennsylvania.

Work partially supported by CNR Grant No. CT81-0052.02.

[1]J. M. Cohen and R. Gautreau, Phys. Rev. D **19**, 2273 (1979).
[2]V. de La Cruz and W. Israel, Nuovo Cimento **51**, 744 (1967).
[3]J. M. Cohen, J. Math. Phys. **9**, 905 (1968).
[4]A. Komar, Phys. Rev. **113**, 934 (1959).
[5]J. M. Cohen and H. E. Moses, Phys. Rev. Lett. **26**, 1641 (1977).
[6]W. Israel, Phys. Rev. D **15**, 935 (1977).

Shear-free perfect fluids with zero magnetic Weyl tensor

C. B. Collins
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, N2L 3G1 Canada

(Received 6 October 1983; accepted for publication 9 November 1983)

Rotating, shear-free general-relativistic perfect fluids are investigated. It is first shown that, if the fluid pressure, p, and energy density, μ, are related by a barotropic equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$, and if the magnetic part of the Weyl tensor (with respect to the fluid flow) vanishes, then the fluid's volume expansion is zero. The class of all such fluids is subsequently characterized. Further analysis of the solutions shows that, in general, the space-times may be regarded as being locally stationary and axisymmetric (they admit a two-dimensional Abelian isometry group with timelike orbits, which is in fact orthogonally transitive), although various specializations can occur, with the "most special" case being the well-known Gödel model, which is space-time homogeneous (it admits a five-dimensional isometry group acting multiply transitively on the space-time). All solutions are of Petrov type D. The fact that there are any solutions in the class at all means that a theorem appearing in the literature is invalid, and the existence of some special solutions in which the fluid's vorticity vector is orthogonal to the acceleration reveals the incompleteness of a previous study of a class of space-times, in which there are Killing vectors parallel to the fluid four-velocity and to the vorticity vector.

PACS numbers: 04.20.Cv, 02.40. + m

1. INTRODUCTION

Shear-free perfect fluids in general relativity have been the object of much study in recent years. In particular, it seems likely that when a shear-free fluid satisfies a barotropic equation of state $p = p(\mu)$, either the fluid's expansion scalar or the fluid's vorticity must vanish. This has been shown indeed to be the case in a wide variety of special circumstances (see, e.g., the references cited by White and Collins[1]), although as far as I am aware there is no proof available that pertains to the general case. The object of the present article is to show that this result holds in yet another special case, thus providing further evidence that it might be more universally valid.

The additional requirement that will be imposed relates to the "magnetic" part[2] H_{ab} of the Weyl tensor, defined by

$$H_{ab} := \tfrac{1}{2}\eta_{ac}{}^{gh}C_{ghbd}u^c u^d,$$

where $\mathbf{u}$ is the unit fluid flow vector. I shall be investigating shear-free fluids that obey an equation of state $p = p(\mu)$, and for which the tensor H_{ab} is zero. After proving that if the shear (σ) is zero, either the expansion scalar (θ) or the vorticity (ω) must vanish, I shall then consider the various cases that naturally arise, viz., $\omega \equiv 0 \neq \theta$, $\omega \equiv \theta \equiv 0$ and $\theta \equiv 0 \neq \omega$. If $\omega \equiv 0$ in any shear-free fluid, then it follows immediately[2] that $H_{ab} \equiv 0$; thus in the two cases where $\omega \equiv 0$, the requirement that H_{ab} vanish represents no further restriction, and so the situation is as analyzed in a recent work.[3] On the other hand, the case $\theta \equiv 0 \neq \omega$ with $H_{ab} \equiv 0$ does involve the examination of a number of constraints that arise. In the present article, I shall examine the compatibility of these constraints, and I shall show that a self-consistent system of equation results, thereby determining the existence of a class of rotating shear-free perfect fluids in which $H_{ab} \equiv 0$, the most familiar member of this class being the Gödel[4] solution (cf. Ref. 2). This result contradicts the theorem of Glass,[5] who claimed that any shear-free general-relativistic perfect fluid (irrespective of an equation of state) is irrotational if and only if $H_{ab} \equiv 0$.

The plan of this article is as follows. In Sec. 2 the theorem is proved which shows that for a shear-free general-relativistic perfect fluid that obeys an equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$, and for which $H_{ab} \equiv 0$, either the expansion scalar or the vorticity must vanish. In Sec. 3, the class of all rotating solutions is examined. Particular attention is paid to the admissible isometry groups and to the allowed Petrov types, in the manner of the investigation by Collins and White.[3] Throughout, I assume a familiarity with the orthonormal tetrad formalism (see, e.g., the article by MacCallum,[6] whose conventions are closely followed). I shall choose a tetrad so that $\mathbf{e}_0$ is aligned along the fluid flow vector, $\mathbf{u}$, and with $\boldsymbol{\omega}$, being nonzero, pointing along $\mathbf{e}_1$, so that $\boldsymbol{\omega} = (\omega,0,0)$. Further, it follows from Proposition 2.1 of Ref. 1 that in such a tetrad, $\Omega_2 = \Omega_3 = 0$, and that, without loss of generality, $\omega + \Omega_1 = 0$. Thus $\boldsymbol{\omega} + \boldsymbol{\Omega} = \mathbf{0}$ in the chosen frame. One small notational departure occurs: I shall relabel those commutation functions with spatial indices in the manner given by White and Collins,[1] viz.,

$$\gamma^{\alpha}{}_{\beta\lambda} = \epsilon_{\beta\lambda\nu}n^{\nu\alpha} + \delta^{\alpha}{}_{\lambda}a_{\beta} - \delta^{\alpha}{}_{\beta}a_{\lambda}, n_{\alpha\beta} = n_{(\alpha\beta)},$$

where

$$n_{\alpha\beta} = \begin{bmatrix} n & \tfrac{1}{2}(d_3 + A_3) & -\tfrac{1}{2}(d_2 + A_2) \\ \tfrac{1}{2}(d_3 + A_3) & \hat{\sigma}_{23} + \hat{\Omega} & -\hat{\sigma}_{22} \\ -\tfrac{1}{2}(d_2 + A_2) & -\hat{\sigma}_{22} & -(\hat{\sigma}_{23} - \hat{\Omega}) \end{bmatrix}$$

and $a_\alpha = (-\tfrac{1}{2}\hat{\theta}, \tfrac{1}{2}(d_2 - A_2), \tfrac{1}{2}(d_3 - A_3))$. At times, it will also be convenient to refer to the quantities $\hat{\theta}_{22} := \hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}$, $\hat{\theta}_{33} := -\hat{\sigma}_{22} + \tfrac{1}{2}\hat{\theta}$ and $\hat{\theta}_{23} := \hat{\sigma}_{23}$. As explained in Ref. 1, the new quantities d_A, n, $\hat{\theta}$, $\hat{\sigma}_{AB}$, and $\hat{\theta}_{AB}$ ($A,B = 2,3$) are closely related to the "kinematic" quantities associated with the $\mathbf{e}_1$ congruence, and by using them as basic variables, the geometrical features of the problem are brought to the fore. The Jacobi identities, Einstein field equations, Bianchi identities, and commutation relations are written out in the Appendix, for the present tetrad, and in the revised notation. The tetrad components of H_{ab}, specialized to the present situation, are also given. These may be obtained from Ref. 2 or from the explicit tetrad expressions in Ref. 7.

2. THE RESULT $H_{ab} \equiv 0$, $\sigma \equiv 0 \Rightarrow \omega\theta \equiv 0$

In the case where the fluid flows along geodesics, it automatically follows (even without an equation of state) that either the vorticity vanishes or the pressure p is constant, and that, if the shear is zero, so also must be either the expansion scalar or the vorticity.[1,8] This result holds irrespective of whether or not H_{ab} vanishes. In this section, it will therefore be assumed that the fluid's acceleration $\dot{\mathbf{u}}$ is nonzero. In this case, the equation of state $p = p(\mu)$ must clearly satisfy $p'(\mu) \neq 0$ by virtue of the Bianchi identities (A26). It will also be assumed that the equation of state satisfies the condition $\mu + p \neq 0$, which is physically highly plausible.

Suppose that the vorticity $\boldsymbol{\omega}$ is nonzero. I shall show that the expansion scalar θ, is then necessarily zero. This can be readily demonstrated by reference to the full Bianchi identities,[2] expressed in terms of E_{ab}, the "electric" part of the Weyl tensor, together with H_{ab}. Thus it follows from Eq. (4.21c) of Ref. 2, viz.,

$$h^t{}_a H^{as}{}_{;d} h_s{}^d + \eta^{tbpq} u_b \sigma_p{}^d E_{qd} - 3E^t{}_s \omega^s = (\mu + p)\omega^t,$$

that if $\sigma_{ab} \equiv H_{ab} \equiv 0$, then $\boldsymbol{\omega}$ is an eigenvector of E_{ab}, with eigenvalue $-\frac{1}{3}(\mu + p)$. If Eq. (4.21d) of Ref. 2, viz.,

$$\begin{aligned} &h_a{}^m h_c{}^t \dot{E}^{ac} + h_a{}^{(m} \eta^{t)rsd} u_r H^a{}_{s;d} \\ &\quad - 2H_q{}^{(t} \eta^{m)bpq} u_b \dot{u}_p + h^{mt} \sigma^{ab} E_{ab} \\ &\quad + \theta E^{mt} - 3E_s{}^{(m} \sigma^{t)s} - E_s{}^{(m} \omega^{t)s} = -\tfrac{1}{2}(\mu + p)\sigma^{tm}, \end{aligned}$$

is specialized to the present case, in which $\sigma_{ab} \equiv H_{ab} \equiv 0$, and to the present tetrad, in which $\omega^\alpha = \omega \delta^\alpha_1$, it follows by considering the (11) component that

$$\partial_0 E_{11} + \theta E_{11} = 0, \tag{2.1}$$

where use is made of the fact that $E_{12} = E_{13} = 0$ in the chosen frame. Now if $\omega \neq 0$, it follows that $E_{11} = -\frac{1}{3}(\mu + p)$, and so (2.1) together with the Bianchi identity (A25) implies that $\partial_0 p = 0$. However, since $p'(\mu) \neq 0$, this requires that $\partial_0 \mu = 0$, in which case, again by (A25), it follows that $\theta \equiv 0$. This therefore proves the following.

Theorem: Consider a shear-free perfect fluid in general relativity, with an equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$. Suppose that the magnetic part of the Weyl tensor, $H_{ab} := \frac{1}{2}\eta_{ac}{}^{gh} C_{ghbd} u^c u^d$ is zero. Then either the fluid's expansion scalar or the fluid's vorticity must vanish.

3. ROTATING SOLUTIONS WITH $H_{ab} \equiv 0$

In this section, the existence and properties of rotating, shear-free perfect fluid solutions of Einstein's field equations having $H_{ab} \equiv 0$ will be established. Since $\boldsymbol{\omega} \neq \mathbf{0}$, the theorem of Sec. 2 requires that the expansion scalar θ must vanish. This means that $\partial_0 \gamma = 0$ for all commutation functions γ, and that $\partial_0 \mu = 0$ by (A25). It also implies that $\partial_0 p = 0$, since if $\dot{\mathbf{u}} \neq \mathbf{0}$, then $p'(\mu) \neq 0$ and the result follows immediately, whereas if $\dot{\mathbf{u}} \equiv \mathbf{0}$, then as observed in Sec. 2, p is identically constant.

The first step is to show that the vortex lines are geodesics. Since $\theta \equiv 0$, this amounts to proving that $d_2 = d_3 = 0$. By the (0α) field Eqs. (A17) and (A18),

$$\partial_2 \omega = \omega(d_2 - 2\dot{u}_2) \text{ and } \partial_3 \omega = \omega(d_3 - 2\dot{u}_3), \tag{3.1}$$

whereas from the conditions $H_{ab} \equiv 0$ [see Eqs. (A39) and (A40)],

$$\partial_2 \omega = -\omega(d_2 + 2\dot{u}_2) \text{ and } \partial_3 \omega = -\omega(d_3 + 2\dot{u}_3). \tag{3.2}$$

Upon comparing (3.1) and (3.2) and applying the fact that $\omega \neq 0$, it follows immediately that

$$d_2 = d_3 = 0. \tag{3.3}$$

It will be useful to note that $n = 0$, as is apparent from the (01) field Eq. (A16). This means that the vortex lines are hypersurface orthogonal (cf. the analogous results in Ref. 1). It will also be useful to observe that frequently equations can be checked against each other, and that some calculation can be avoided, by invoking a rotation $\mathbf{e}_2 \to \mathbf{e}_3$, $\mathbf{e}_3 \to -\mathbf{e}_2$, in the manner of White and Collins.[1] Thus Eqs. (3.1) are interrelated, as are (3.2).

I shall now investigate the equations for the propagation of the quantities $\dot{u}_\alpha$. If the $[\mathbf{e}_1, \mathbf{e}_2]$ commutator (A30) is applied to ω, and Eqs. (3.1), (A1), (A37), and (A38) employed, there results

$$-2\,\partial_1 \dot{u}_2 + \partial_2 \dot{u}_1 - 2\dot{u}_1 \dot{u}_2 + 2\dot{u}_3 \hat{\Omega} = 0, \tag{3.4a}$$

and similarly

$$-2\,\partial_1 \dot{u}_3 + \partial_3 \dot{u}_1 - 2\dot{u}_1 \dot{u}_3 - 2\dot{u}_2 \hat{\Omega} = 0, \tag{3.4b}$$

which follows either by applying the $[\mathbf{e}_3, \mathbf{e}_1]$ commutator (A32) to ω, or by invoking the rotation $\mathbf{e}_2 \to \mathbf{e}_3$, $\mathbf{e}_3 \to -\mathbf{e}_2$, as described above. Now since $E_{12} = 0$, we have[2,8]

$$\partial_1 \dot{u}_2 + \partial_2 \dot{u}_1 + \dot{u}_1 \dot{u}_2 - \dot{u}_3 \hat{\Omega} = 0, \tag{3.5a}$$

and similarly

$$\partial_1 \dot{u}_3 + \partial_3 \dot{u}_1 + \dot{u}_1 \dot{u}_3 + \dot{u}_2 \hat{\Omega} = 0. \tag{3.5b}$$

For compatibility of Eqs. (3.4) and (3.5), we therefore obtain

$$\partial_2 \dot{u}_1 = \partial_3 \dot{u}_1 = 0, \tag{3.6}$$

and Eqs. (3.4) and (3.5) are modified accordingly.

It is now convenient to specialize the tetrad further. Since there is still the freedom to rotate the tetrad about the $\mathbf{e}_1$ direction at any one point on each fluid flow line, $\dot{u}_3$ may be set zero initially, and then propagation of $\dot{u}_3$ along $\mathbf{e}_0$ requires that $\dot{u}_3 = 0$ everywhere, by Eq. (A33). The tetrad is now completely specified, unless of course $\dot{u}_2$ is also zero, in which case the fluid's vorticity and acceleration are parallel (cf. Refs. 1 and 3).

In such a frame, Eqs. (3.5b) and (3.6) imply that

$$\dot{u}_2 \hat{\Omega} = 0, \tag{3.7}$$

and the discussion now divides into two cases, depending on whether or nor $\dot{u}_2$ vanishes.

Case 1: $\dot{u}_2 \neq 0$. By (3.7), $\hat{\Omega} = 0$. We also have

$$\partial_3 \dot{u}_2 = A_3 = 0, \tag{3.8}$$

since by (3.3), (A24), and (A38), $\partial_3 \dot{u}_2 = \dot{u}_2 A_3$, yet from (A35), $\partial_3 \dot{u}_2 = -\dot{u}_2 A_3$, and using the fact that $\dot{u}_2 \neq 0$, (3.8) results.

Eqs. (3.3), (A37), and (A38) simplify the (11) field Eq. (A19), to yield

$$\partial_1 \dot{u}_1 = -\dot{u}_1{}^2 - \tfrac{1}{6}(\mu - p) - \tfrac{1}{3}\Lambda,$$

which is compatible with (3.6) if and only if

$$p'(\mu) = 1, \tag{3.9}$$

by virtue of the commutation relation (A30) and the assumption that $\dot{u}_2(\mu + p) \neq 0$.

The difference between Eqs. (A20) and (A21) gives, upon using (3.3), (A37), and (A38),

$$\partial_2 \dot{u}_2 = -\dot{u}_2{}^2 + \dot{u}_2 A_2. \tag{3.10}$$

By Eq. (3.8), it follows that $\partial_3 \dot{u}_2 = 0$, and moreover, applying the $[\mathbf{e}_2,\mathbf{e}_3]$ commutator (A31) to $\dot{u}_2$, $\partial_3\partial_2\dot{u}_2 = 0$, where again use is made of (3.8). Now differentiating (3.10) along $\mathbf{e}_3$ and recalling that $\dot{u}_2 \neq 0$, it is seen that $\partial_3 A_2 = 0$. These results, together with (3.2), (3.3), (3.6), and (3.8), show that $\partial_3 \gamma = 0$ for all commutation functions γ. In addition, $\partial_3 p = \partial_3 \mu = 0$ by (3.9) and (A26). The remaining equations can be simplified to provide propagation equations in the $\mathbf{e}_1$ and $\mathbf{e}_2$ directions for the commutation functions ($\dot{u}_1$, $\dot{u}_2$, ω, and A_2) and for the fluid quantities μ and p. These are

$$\begin{aligned}
&\partial_1 \omega = -\omega \dot{u}_1, && \partial_1 \dot{u}_1 = -\dot{u}_1{}^2 - \tfrac{1}{6}(\mu - p) - \tfrac{1}{3}\Lambda,\\
&\partial_2 \omega = -2\omega \dot{u}_2, && \partial_2 \dot{u}_1 = 0,\\
&\partial_1 \dot{u}_2 = -\dot{u}_1 \dot{u}_2, && \partial_1 A_2 = -\dot{u}_1 A_2,\\
&\partial_2 \dot{u}_2 = -\dot{u}_2{}^2 + \dot{u}_2 A_2, && \partial_2 A_2 = 4\omega^2 + \dot{u}_2 A_2 - (\mu + p) - A_2{}^2,\\
&\partial_1 p = -(\mu + p)\dot{u}_1,
\end{aligned} \tag{3.11}$$

and

$$\partial_2 p = -(\mu + p)\dot{u}_2,$$

together with the algebraic constraints

$$\omega^2 + \dot{u}_1{}^2 + \dot{u}_2 A_2 = \tfrac{1}{3}(\mu + 2p - \Lambda)$$

and (3.12)

$$p'(\mu) = 1.$$

It may be readily verified that Eqs. (3.11) are consistent, by virtue of the $[\mathbf{e}_1,\mathbf{e}_2]$ commutator (A30), and that Eq. (3.12) is propagated along $\mathbf{e}_1$ and $\mathbf{e}_2$ without generating further constraints.

It has therefore been demonstrated that a class of rotating shear-free perfect solutions exists, in which $H_{ab} \equiv 0$. Some properties of this class will now be studied.

It is clear that the solutions possess a G_2 isometry group, whose orbits are timelike two-surfaces orthogonal to both $\dot{\mathbf{u}}$ and $\boldsymbol{\omega}$. By the commutators (A29) and (A30), it is apparent that this group is Abelian and orthogonally transitive. In general, this is the maximal isometry group. The only exceptional case that can arise occurs when $\dot{u}_1 = 0$ and $p = \mu + 2\Lambda$. This is evident from Eqs. (3.11), since if $\mathbf{k}$ is a Killing vector, $\mathbf{k}(p) = \mathbf{k}(\omega) = 0$, which requires that either $\mathbf{k} = \mathbf{0}$, or that $\dot{u}_1 = 0$, and then $k_2 = 0$. The requirement that $\dot{u}_1 = 0$ be preserved along $\mathbf{e}_1$ is obtained from (3.11), and yields $p = \mu + 2\Lambda$, which is consistent with (3.9). The commutation relations (A27), (A30), and (A32) show that $\mathbf{e}_1$ now commutes with all basis vectors, and since, for a geometrically invariantly defined frame, Killing's equations are equivalent to the equations $[\mathbf{k},\mathbf{e}_a] = \mathbf{0}$ for a Killing vector $\mathbf{k}$ (cf. Ref. 8), it follows that $\mathbf{e}_1$ is a third Killing vector. In this case, there is a G_3 isometry group of Bianchi Type I acting transitively on timelike hypersurfaces orthogonal to the fluid's acceleration vector. Since in this special case there is a Killing vector parallel to the vorticity vector, the solutions fall into the class of models investigated by Krasinski.[9,10] However, while the present solutions share some common features with Krasinski's acceleration models, such as the facts that the acceleration vector is orthogonal to the vorticity vector, and that that is an Abelian isometry group acting transitively on timelike hypersurfaces orthogonal to the acceleration vector, it is important to recognize that they are not recoverable from Krasinski's class, and that therefore, contrary to his claims, Krasinski does not determine all "flow-stationary" (i.e., rigidly rotating) "vortex-homogeneous" solutions for a perfect fluid satisfying an equation of state. This can be demonstrated by using the relationship $p = \mu + 2\Lambda$ in Eqs. (1.10), (1.12), and (5.13) of Ref. 9, from which it follows, in Krasinski's notation, that $\rho H_{,2} = \rho_{,2} H$, a possibility that Krasinski dismisses after his Eq. (5.1). Thus, accepting the form of Krasinski's solution, and imposing the equation of state appropriate to the present special solution, results in a contradiction. In fact, the reason for this is that Krasinski considers only a very special subset of solutions, which, in the present notation, may be described by the requirement that ω^2 be proportional to $\mu + p$. The fact that the present solutions are incompatible with those of Krasinski simply means that in Krasinski's solutions $H_{ab} \neq 0$.

The solutions in Case I are all of Petrov type D, as follows from a consideration of the eigenvalues of E_{ab}. The chosen frame is an eigenframe of E_{ab}, and $E_{22} = E_{33}$, by (3.11). The possibility that all eigenvalues are equal is ruled out by recalling that E_{ab} is trace-free, and that then $\mu + p = 0$. This is in agreement with the result quoted by Ellis[2] and attributed to Trümper, that the only conformally flat perfect fluid solutions in which there is an equation of state are the spatially homogeneous Friedmann–Robertson–Walker models (in which $\boldsymbol{\omega} = \mathbf{0}$). Wainwright[11] has provided a classification scheme for Petrov type D perfect fluids, in terms of which the solutions in the present case are of Class IC, and the fluid's acceleration vector does not lie in the two-space defined by the principal null directions of the Weyl tensor.

Case 2: $\dot{u}_2 = 0$. In this case, the fluid acceleration and vorticity are parallel (and the acceleration is possibly zero). Thus the result of Collins and White[3] can be employed. Since Eqs. (3.3), (A37), and (A38) show that $\hat{\sigma}_{AB} = 0$ and that $d_A = 0$, the solutions are locally rotationally symmetric,[8,12] by Proposition 2.4 of Ref. 3. Now the tetrad is not fixed uniquely; also $\partial_2 \gamma = \partial_3 \gamma = 0$ for all commutation functions γ, save possibly Ω, A_2 and A_3, and $\partial_2 p = \partial_2 \mu = 0$. The solutions in general belong to type IIIAGii of Collins and White,[3] but in the special case when $\dot{\mathbf{u}} \equiv \mathbf{0}$, they belong to type IIIGGii. It remains to impose the requirement that $H_{ab} \equiv 0$. Of Eqs. (A37)–(A40), the only nontrivial restriction is

$$\dot{u}_1 = \tfrac{1}{2}\hat{\theta}, \tag{3.13}$$

and this must be preserved by the propagation equations

$$\partial_1 \dot u_1 = -\dot u_1{}^2 - 2\omega^2 - \dot u_1 \hat\theta + \tfrac{1}{2}(\mu + 3p - 2\Lambda) \qquad (3.14)$$

and

$$\partial_1 \hat\theta = -\tfrac{1}{2}\hat\theta^2 + 2\omega^2 + \dot u_1 \hat\theta - (\mu + p), \qquad (3.15)$$

which are obtained from Eqs. (A15) and (A19). Suppose first that $\dot u_1 \neq 0$. Then preservation of (3.13) gives rise to the constraint

$$3(\omega^2 + \dot u_1{}^2) = \mu + 2p - \Lambda, \qquad (3.16)$$

which itself must be propagated, to yield

$$p'(\mu) = 1, \qquad (3.17)$$

where use has been made of (3.15), (A1), (A26), and (A37). With Eqs. (3.13), (3.16), and (3.17) in place, it is clear that a class of solutions exists with the properties claimed.

Now suppose that $\dot u_1 = 0$. Equations (3.14) and (3.15) provide algebraic relationships equivalent to

$$\mu = \omega^2 - \Lambda \text{ and } p = \omega^2 + \Lambda.$$

By (A1), ω is constant, and hence so also are μ and p. This is the Gödel solution, generalized to include pressure.

The solutions in which $\dot u_1 \neq 0$ are of Class Id in the terminology of Stewart and Ellis,[12] and admit a G_4 isometry group acting multiply transitively on timelike hypersurfaces orthogonal to $\boldsymbol\omega$ (and to $\dot{\mathbf u}$). The solutions in which $\dot u_1 = 0$ admit a G_5 isometry group acting multiply transitively on space-time, and belong to Case Ia of Stewart and Ellis,[12] and to Case Ia of Ellis[8] if the fluid is interpreted as dust ($p = 0$).

Since all of the models in Case 2 are locally rotationally symmetric, they belong to Petrov type D. The specialization to type O is forbidden, because in order to be conformally flat, the solutions would have to be Friedmann–Robertson–Walker,[2] and hence irrotational. The solutions are of Class Ia in the classification scheme of Wainwright,[11] and the fluid's acceleration vector lies in the two-space spanned by the principal null directions.

4. DISCUSSION

In this article, a characterization has been given of all rotating, shear-free perfect fluid solutions of Einstein's field equations in which $H_{ab} := \tfrac{1}{2}\eta_{ac}{}^{gh} C_{ghbd} u^c u^d$ is zero, and in which there is an equation of state $p = p(\mu)$ such that $\mu + p \neq 0$. The results contradict those of Glass,[5] who claimed that there were no such solutions. It appears that the source of Glass' error lies in the conclusion that his Eq. (11) is valid for solutions of Petrov type I and D.

As we have seen in Sec. 3, in the most general solutions, $\dot{\mathbf u}$ and $\boldsymbol\omega$ are not parallel, and there is a G_2 isometry group whose orbits are timelike two-surfaces orthogonal to both $\dot{\mathbf u}$ and $\boldsymbol\omega$. It is of interest to note that while in certain circumstances ($\dot{\mathbf u}$ orthogonal to $\boldsymbol\omega$; $p = \mu + 2\Lambda$), there is a third Killing vector $\mathbf e_1$, in *all* cases (irrespective of the relative orientation of $\dot{\mathbf u}$ and $\boldsymbol\omega$), there is a conformal Killing vector parallel to $\mathbf e_1$. For suppose that M is a nontrivial solution of the equations $\partial_1 M = M\dot u_1$, $\partial_a M = 0$ ($a = 0,2,3$). This set of equations is consistent, by virtue of (3.6) and the commutation relations (A27)–(A32). Then $\mathbf k := M\mathbf e_1$ is a conformal Killing vector, since $\mathscr{L}_{ij} = 2k_{(i;j)} = 2\phi g_{ij}$, where $\phi = M\dot u_1$. The vector $\mathbf k$ is a homothetic vector if and only if $\phi = M\dot u_1$ is constant, which is satisfied if and only if $\partial_1 \dot u_1 + \dot u_1{}^2 = 0$, i.e., if and only if $p = \mu + 2\Lambda$, as follows from Eqs. (3.11), (3.14), and (3.16). The vector $\mathbf k$ is a Killing vector if and only if $\phi = 0$, i.e., if and only if $\dot u_1 = 0$ (and then of course $p = \mu + 2\Lambda$). Note that these results are in agreement with those of McIntosh,[13] who showed that if there is a proper homothetic vector orthogonal to the flow of a perfect fluid, then necessarily $p = \mu$ [adjusted for the cosmological term, this condition would be the same as the present equation, viz., $p = \mu + 2\Lambda$; note also that, in McIntosh's notation, $F_{ab} := k_{[a;b]} = 0$, and so $J^a := (1/4\pi)F^{ab}{}_{;b} = 0$]. In the special solutions where $\dot{\mathbf u}$ and $\boldsymbol\omega$ are parallel, there is higher symmetry, since the space-times are locally rotationally symmetric; if $\dot{\mathbf u}$ is nonzero, there is a G_4 isometry group acting multiply transitively on the timelike hypersurfaces orthogonal to $\boldsymbol\omega$, while if $\dot{\mathbf u}$ vanishes, there is a G_5 isometry group acting multiply transitively on space-time. A summary of the possible symmetries of the various solutions is provided in Table I. All solutions in the entire class are of Petrov type D.

We have remarked that in the general solutions the isometry group is Abelian and orthogonally transistive. As observed by Collins and White[3], such space-times may be regarded *locally* as stationary axisymmetric solutions of Einstein's field equations. However, while we can select a spacelike Killing vector field and identify points along the trajectories (cf. Ref. 14), thus making the space-time axisymmetric, in general there will not be a rotation axis on which the Killing vector field vanishes, and, even if there is, the space-time will not necessarily be regular on the axis (cf. the conical singularities constructed from Minkowski space-time by Ellis and Schmidt[15]). Provided that the equation of state is analytic, the solutions presently being considered are

TABLE I. The allowed symmetries of the class of solutions considered in this article. Here M is a nontrivial solution of the equations $\partial_1 M = M\dot u_1$, $\partial_a M = 0$ ($a = 0, 2, 3$). The most general space-times are given in the first row. Specialization occurs in two directions: directly downwards indicates a transition to the corresponding case when $\dot u_2 = 0$ (making $\dot{\mathbf u}$ and $\boldsymbol\omega$ parallel), whereas diagonally downwards indicates a specialization of the conformal Killing vector $M\mathbf e_1$, either to a homothetic vector (column 2) or to a Killing vector (column 3). The entry "G_r on T_n" means that the solutions admit an r-parameter isometry group whose orbits are timelike and of dimension n (and in the final row, "T_4" is the entire space-time; this corresponds to the Gödel solution). When $r > n$, the solutions are locally rotationally symmetric.

$M\mathbf e_1$ is a conformal Killing vector	$M\mathbf e_1$ is a homothetic vector	$M\mathbf e_1$ is a Killing vector
G_2 on T_2		
	G_2 on T_2 $\partial_1\dot u_1 + \dot u_1{}^2 = 0 \Leftrightarrow$ $p = \mu + 2\Lambda$	
		G_3 on T_3 $\dot u_1 = 0; p = \mu + 2\Lambda$
G_4 on T_3		
	G_4 on T_3 $\partial_1\dot u_1 + \dot u_1{}^2 = 0 \Leftrightarrow$ $p = \mu + 2\Lambda$	
		G_5 on T_4 $\dot u_1 = 0; p = \mu + 2\Lambda$

analytic.[16] Hence, having determined locally that there is an orthogonally transitive Abelian G_2 isometry group with timelike orbits, such a space-time must (provided that the equation of state is analytic) be extended analytically, so that in principle one could determine uniquely whether the initial region is locally part of a space-time that contains a regular rotation axis, and so is stationary and axisymmetric in the conventional sense. In the case where $\dot{u}_2 \neq 0$ and $\dot{u}_1 = 0$, the solutions admit an Abelian G_3 isometry group, and so may be regarded locally as stationary *cylindrically* symmetric space-times, although again not necessarily in the conventional sense.

As already observed, in any *nonrotating* shear-free perfect fluid, H_{ab} is *necessarily* zero. Thus the class of all shear-free perfect fluid solutions of Einstein's field equations, in which there is an equation of state $p = p(\mu)$ satisfying $\mu + p \neq 0$, is given by (a) the rotating solutions that have been discussed herein, and (b) the nonrotating solutions, details of which can be found in Ref. 3.

Note added in proof. Dr. A. Krasinski has kindly informed me that his results[9,10] considered herein were completed by him in a subsequent article.

ACKNOWLEDGMENT

This work was partially supported by an Individual Operating grant (#A3978) from the Natural Sciences and Engineering Research Council of Canada. I am grateful to Helen Warren for her careful typing of the manuscript.

APPENDIX

A. Jacobi identities

$$\partial_1\omega = \omega(\dot{u}_1 - \hat{\theta}), \tag{A1}$$

$$\partial_1 n + \partial_2 d_3 - \partial_3 d_2 = \tfrac{2}{3}\theta\omega - n\hat{\theta} + d_2A_3 - d_3A_2, \tag{A2}$$

$$\partial_1A_3 + \partial_2\hat{\theta}_{23} - \partial_3\hat{\theta}_{22} + \partial_2\hat{\Omega} + d_3\hat{\theta}_{22} + A_3\hat{\theta}_{33} - (d_2 - A_2)\hat{\theta}_{23} - (d_2 - A_2)\hat{\Omega} = 0, \tag{A3}$$

$$\partial_1A_2 + \partial_3\hat{\theta}_{23} - \partial_2\hat{\theta}_{33} - \partial_3\hat{\Omega} + d_2\hat{\theta}_{33} + A_2\hat{\theta}_{22} - (d_3 - A_3)\hat{\theta}_{23} + (d_3 - A_3)\hat{\Omega} = 0, \tag{A4}$$

$$\partial_0\omega = (p' - \tfrac{2}{3})\omega\theta, \tag{A5}$$

$$\partial_0\hat{\theta} = \tfrac{2}{3}\partial_1\theta + \tfrac{1}{3}\theta(2\dot{u}_1 - \hat{\theta}), \tag{A6}$$

$$\partial_0(d_2 - A_2) + \tfrac{2}{3}\partial_2\theta + \tfrac{1}{3}\theta(2\dot{u}_2 + d_2 - A_2), \tag{A7}$$

$$\partial_0(d_3 - A_3) + \tfrac{2}{3}\partial_3\theta + \tfrac{1}{3}\theta(2\dot{u}_3 + d_3 - A_3), \tag{A8}$$

$$\partial_0 n + \tfrac{1}{3}n\theta = 0, \tag{A9}$$

$$\partial_0(d_2 + A_2) + \tfrac{1}{3}(d_2 + A_2)\theta = 0, \tag{A10}$$

$$\partial_0(d_3 + A_3) + \tfrac{1}{3}(d_3 + A_3)\theta = 0, \tag{A11}$$

$$\partial_0\hat{\Omega} + \tfrac{1}{3}\hat{\Omega}\theta = 0, \tag{A12}$$

$$\partial_0\hat{\theta}_{23} + \tfrac{1}{3}\hat{\theta}_{23}\theta = 0, \tag{A13}$$

$$\partial_0(\hat{\theta}_{22} - \hat{\theta}_{33}) + \tfrac{1}{3}(\hat{\theta}_{22} - \hat{\theta}_{33})\theta = 0, \tag{A14}$$

where a prime (′) denotes differentiation with respect to μ.

B. Field equations

$$(00)\quad \partial_0\theta + \tfrac{1}{3}\theta^2 - 2\omega^2 - \partial_1\dot{u}_1 - \partial_2\dot{u}_2 - \partial_3\dot{u}_3 - (\dot{u}_1^2 + \dot{u}_2^2 + \dot{u}_3^2) - \dot{u}_1\hat{\theta} + \dot{u}_2(d_2 - A_2) + \dot{u}_3(d_3 - A_3) + \tfrac{1}{2}(\mu + 3p - 2\Lambda) = 0, \tag{A15}$$

$$(01)\quad \tfrac{2}{3}\partial_1\theta - n\omega = 0, \tag{A16}$$

$$(02)\quad \tfrac{2}{3}\partial_2\theta + \partial_3\omega + \omega(-d_3 + 2\dot{u}_3) = 0, \tag{A17}$$

$$(03)\quad \tfrac{2}{3}\partial_3\theta - \partial_2\omega + \omega(d_2 - 2\dot{u}_2) = 0, \tag{A18}$$

$$(11)\quad \partial_1\hat{\theta} - \partial_2 d_2 - \partial_3 d_3 + \hat{\theta}_{22}^2 + \hat{\theta}_{33}^2 + 2\hat{\theta}_{23}^2 + d_2^2 + d_3^2 - d_2A_2 - d_3A_3 - \tfrac{1}{2}n^2 = \tfrac{1}{3}\partial_0\theta - \partial_1\dot{u}_1 - \dot{u}_1^2 - \tfrac{1}{2}(\mu - p) - \Lambda + \tfrac{1}{3}\theta^2 + \dot{u}_2 d_2 + \dot{u}_3 d_3, \tag{A19}$$

$$(22)\quad \partial_1\hat{\theta}_{22} - \partial_2 d_2 + \partial_2A_2 + \partial_3A_3 + \hat{\theta}_{22}^2 + \hat{\theta}_{22}\hat{\theta}_{33} - d_3A_3 + A_2^2 + A_3^2 + d_2^2 - 2\hat{\theta}_{23}\hat{\Omega} + \tfrac{1}{2}n^2 - n(\hat{\Omega} - \hat{\theta}_{23}) = -\dot{u}_1\hat{\theta}_{22} - \tfrac{1}{2}(\mu - p) - \Lambda + \tfrac{1}{3}\theta^2 - \partial_2\dot{u}_2 - \dot{u}_2^2 + \tfrac{1}{3}\partial_0\theta + 2\omega^2 - \dot{u}_3A_3, \tag{A20}$$

$$(33)\quad \partial_1\hat{\theta}_{33} - \partial_3 d_3 + \partial_2A_2 + \partial_3A_3 + \hat{\theta}_{33}^2 + \hat{\theta}_{22}\hat{\theta}_{33} - d_2A_2 + A_2^2 + A_3^2 + d_3^2 + 2\hat{\theta}_{23}\hat{\Omega} + \tfrac{1}{2}n^2 - n(\hat{\Omega} + \hat{\theta}_{23}) = -\dot{u}_1\hat{\theta}_{33} - \tfrac{1}{2}(\mu - p) - \Lambda + \tfrac{1}{3}\theta^2 - \partial_3\dot{u}_3 - \dot{u}_3^2 + \tfrac{1}{3}\partial_0\theta + 2\omega^2 - \dot{u}_2A_2, \tag{A21}$$

$$(12)\quad \partial_1A_2 + \partial_2\hat{\theta}_{33} - \partial_3\hat{\theta}_{23} - \partial_3\hat{\Omega} - d_3\hat{\theta}_{23} - 3A_3\hat{\theta}_{23} + (d_3 - A_3)\hat{\Omega} + d_2\hat{\theta}_{33} + 2A_2\hat{\theta}_{33} - A_2\hat{\theta}_{22} + \partial_3 n - 2nd_3 = -\partial_2\dot{u}_1 - d_2\dot{u}_1 - \partial_1\dot{u}_2 - 2\dot{u}_1\dot{u}_2 + \dot{u}_2\hat{\theta}_{22} - n\dot{u}_3 + \dot{u}_3(\hat{\Omega} + \hat{\theta}_{23}), \tag{A22}$$

$$(13)\quad \partial_1A_3 + \partial_3\hat{\theta}_{22} - \partial_2\hat{\theta}_{23} + \partial_2\hat{\Omega} - d_2\hat{\theta}_{23} - 3A_2\hat{\theta}_{23} - (d_2 - A_2)\hat{\Omega} + d_3\hat{\theta}_{22} + 2A_3\hat{\theta}_{22} - A_3\hat{\theta}_{33} - \partial_2 n + 2nd_2 = -\partial_3\dot{u}_1 - d_3\dot{u}_1 - \partial_1\dot{u}_3 - 2\dot{u}_1\dot{u}_3 + \dot{u}_3\hat{\theta}_{33} + n\dot{u}_2 - \dot{u}_2(\hat{\Omega} - \hat{\theta}_{23}), \tag{A23}$$

$$(23)\quad 2\partial_1\hat{\theta}_{23} - \partial_2 d_3 - \partial_3 d_2 + 2\hat{\theta}\hat{\theta}_{23} + 2\hat{\Omega}(\hat{\theta}_{22} - \hat{\theta}_{33}) + 2d_2d_3 + d_2A_3 + d_3A_2 - n(\hat{\theta}_{22} - \hat{\theta}_{33}) = -2\dot{u}_1\hat{\theta}_{23} - \partial_2\dot{u}_3 - \partial_3\dot{u}_2 - 2\dot{u}_2\dot{u}_3 + \dot{u}_2A_3 + \dot{u}_3A_2. \tag{A24}$$

C. Contracted Bianchi identities

$$\partial_0\mu + (\mu + p)\theta = 0, \tag{A25}$$

$$\partial_\alpha p + (\mu + p)\dot{u}_\alpha = 0. \tag{A26}$$

D. Commutation relations

$$[\mathbf{e}_0,\mathbf{e}_1] = \dot{u}_1\mathbf{e}_0 - \tfrac{1}{3}\theta\mathbf{e}_1, \tag{A27}$$

$$[\mathbf{e}_0,\mathbf{e}_2] = \dot{u}_2\mathbf{e}_0 - \tfrac{1}{3}\theta\mathbf{e}_2, \tag{A28}$$

$$[\mathbf{e}_0,\mathbf{e}_3] = \dot{u}_3\mathbf{e}_0 - \tfrac{1}{3}\theta\mathbf{e}_3, \tag{A29}$$

$$[\mathbf{e}_1,\mathbf{e}_2] = -d_2\mathbf{e}_1 - \hat{\theta}_{22}\mathbf{e}_2 + (\hat{\Omega} - \hat{\theta}_{23})\mathbf{e}_3, \tag{A30}$$

$$[\mathbf{e}_2,\mathbf{e}_3] = -2\omega\mathbf{e}_0 + n\mathbf{e}_1 + A_3\mathbf{e}_2 - A_2\mathbf{e}_3, \tag{A31}$$

$$[\mathbf{e}_3,\mathbf{e}_1] = d_3\mathbf{e}_1 + (\hat{\Omega} + \hat{\theta}_{23})\mathbf{e}_2 + \hat{\theta}_{33}\mathbf{e}_3, \tag{A32}$$

E. Commutation relations applied to F

$$\partial_0\dot{u}_\alpha - p'\partial_\alpha\theta + \left[(p''/p')(\mu + p) - p' + \tfrac{1}{3}\right]\dot{u}_\alpha\theta = 0, \tag{A33}$$

$$\partial_1\dot{u}_2 - \partial_2\dot{u}_1 = -d_2\dot{u}_1 - \hat{\theta}_{22}\dot{u}_2 + (\hat{\Omega} - \hat{\theta}_{23})\dot{u}_3, \tag{A34}$$

$$\partial_2\dot{u}_3 - \partial_3\dot{u}_2 = -2\omega p'\theta + n\dot{u}_1 + A_3\dot{u}_2 - A_2\dot{u}_3, \tag{A35}$$

$$\partial_3\dot{u}_1 - \partial_1\dot{u}_3 = d_3\dot{u}_1 + (\hat{\Omega} + \hat{\theta}_{23})\dot{u}_2 + \hat{\theta}_{33}\dot{u}_3, \tag{A36}$$

where a prime (′) denotes differentiation with respect to μ.

F. The components of H_{ab}

$$H_{11} = \omega(2\dot{u}_1 - \hat{\theta});$$

$$H_{22} = -\omega\hat{\theta}_{33} - \partial_1\omega; \; H_{33} = -\omega\hat{\theta}_{22} - \partial_1\omega;$$

$$H_{12} = \omega\dot{u}_2 + \tfrac{1}{2}\partial_2\omega + \tfrac{1}{2}d_2\omega; \; H_{13} = \omega\dot{u}_3 + \tfrac{1}{2}\partial_3\omega + \tfrac{1}{2}d_3\omega;$$

and

$$H_{23} = \omega\hat{\theta}_{23}.$$

Hence (assuming that $\omega \neq 0$), $H_{ab} = 0$ if and only if

$$\dot{u}_1 = \hat{\theta}_{22} = \hat{\theta}_{33} (= \tfrac{1}{2}\hat{\theta}), \tag{A37}$$

$$\hat{\theta}_{23} = 0, \tag{A38}$$

$$\partial_2\omega = -\omega(d_2 + 2\dot{u}_2), \tag{A39}$$

and

$$\partial_3\omega = -\omega(d_3 + 2\dot{u}_3). \tag{A40}$$

[1] A. J. White and C. B. Collins, J. Math. Phys. **25**, 332 (1984).

[2] G. F. R. Ellis, "Relativity and Cosmology," in *General Relativity and Cosmology*, Proceedings of the International School of Physics "Enrico Fermi" Course XLVII, 1969, edited by R. K. Sachs (Academic, London, 1971), p. 104.

[3] C. B. Collins and A. J. White, "A Class of Shear-free Perfect Fluids in General Relativity, II," J. Math. Phys. (to be published).

[4] K. Gödel, Rev. Mod. Phys. **21**, 447 (1949).

[5] E. N. Glass, J. Math. Phys. **16**, 2361 (1975).

[6] M. A. H. MacCallum, "Cosmological Models from a Geometric Point of View," in Cargèse Lectures in Physics, Vol. 6, Lectures at the International Summer School of Physics, Cargèse, Corsica, 1971, edited by E. Schatzman (Gordon and Breach, New York, 1973), p. 61.

[7] M. A. H. MacCallum, "A Class of Homogeneous Cosmological Models," Ph.D. thesis, Cambridge University, 1969.

[8] G. F. R. Ellis, J. Math. Phys. **8**, 1171 (1967).

[9] A. Krasinski, Acta Phys. Pol. B **5**, 411 (1974).

[10] A. Krasinski, Acta Phys. Pol. B **6**, 223 (1975).

[11] J. Wainwright, Gen. Relativ. Gravit. **8**, 797 (1977).

[12] J. M. Stewart and G. F. R. Ellis, J. Math. Phys. **9**, 1072 (1968).

[13] C. B. G. McIntosh, Phys. Lett. A **50**, 429 (1975); Gen. Relativ. Gravit. **7**, 199 (1976).

[14] G. F. R. Ellis, Gen. Relativ. Gravit. **2**, 7 (1971).

[15] G. F. R. Ellis and B. G. Schmidt, Gen. Relativ. Gravit. **8**, 915 (1977).

[16] L. Lindblom, J. Math. Phys. **17**, 1810 (1976).

[17] A. Krasinski, Rep. Math. Phys. **14**, 225 (1978).

On bilinear and trilinear extensions to the form of the Einstein tensor

Paul McNamara and John Shepanski
Department of Theoretical Physics, University of New South Wales, Kensington 2033, New South Wales, Australia

(Received 23 November 1982; accepted for publication 17 June 1983)

As is well known, the Einstein tensor $\mathbf{G} \equiv \overset{[1]}{\mathbf{G}}$ is constructed from **Riemann** and the metric so that it is symmetric, divergence-free, linear in **Riemann**, second rank, and vanishes when space-time is flat. In this paper the condition of linearity in **Riemann** is extended to bilinearity and trilinearity in **Riemann** and the resulting tensors $\overset{[2]}{\mathbf{G}}$ and $\overset{[3]}{\mathbf{G}}$, respectively, are found by a direct calculation. These are then compared with parts of the general tensor expression derived by Lovelock (who started with a different set of assumptions) and shown to be equal to it to within an arbitrary multiplicative constant. From this comparison it follows that $\overset{[2]}{\mathbf{G}}$ and $\overset{[3]}{\mathbf{G}}$ both vanish in the case where the dimension of the space is 4 (as occurs in the usual space-time).

PACS numbers: 04.20.Cv

1. INTRODUCTION

In gravitation theory the Einstein field equations $\mathbf{G} = 8\pi\mathbf{T}$ (in geometrized units) are usually derived by constructing all tensors $\mathbf{G}$ having the same mathematical properties as the energy-momentum tensor $\mathbf{T}$, and built solely from the geometry of space-time.[1] Thus one seeks all $\mathbf{G}$ such that

(a) $\mathbf{G}$ vanishes when space-time is flat; (1.1)

(b) $\mathbf{G}$ is constructed from **Riemann** and the metric and nothing else; (1.2)

(c) $\mathbf{G}$ has automatically vanishing divergence, i.e.,

$$\nabla\cdot\mathbf{G} = 0; \tag{1.3}$$

(d) G is symmetric and of second rank,[2] i.e.,

$$G_{\alpha\beta} = G_{\beta\alpha}; \tag{1.4}$$

(e) $\mathbf{G}$ is linear in **Riemann**. (1.5)

To within a multiplicative constant one arrives at the well-known expression for $\mathbf{G}$, i.e.,

$$G_{\alpha\beta} \equiv \overset{[1]}{G}_{\alpha\beta} = R_{\alpha\beta} - \tfrac{1}{2}g_{\alpha\beta}R. \tag{1.6}$$

It will be shown by a simple calculation that by changing condition (1.5) to bilinearity in **Riemann** one gets, to within a multiplicative constant,

$$\begin{aligned}\overset{[2]}{G}{}^{\alpha}_{\beta} = {} & R R^{\alpha}_{\beta} + R^{\eta\alpha}_{\mu\nu} R^{\mu\nu}_{\eta\beta} - 2R^{\alpha}_{\mu} R^{\mu}_{\beta} - 2R^{\mu}_{\nu} R^{\nu\alpha}_{\mu\beta} \\ & - \tfrac{1}{4}R^{2}\delta^{\alpha}_{\beta} - \tfrac{1}{4}R^{\eta\gamma}_{\mu\nu} R^{\mu\nu}_{\eta\gamma}\delta^{\alpha}_{\beta} + R^{\mu}_{\nu} R^{\nu}_{\mu}\delta^{\alpha}_{\beta}.\end{aligned} \tag{1.7}$$

Similarly, for trilinearity in **Riemann**, one gets, again to within a multiplicative constant, the following rather longer expression:

$$\begin{aligned}\overset{[3]}{G}{}^{\alpha}_{\beta} = {} & R^{2}R^{\alpha}_{\beta} - 4R^{\delta}_{\gamma}R^{\gamma\alpha}_{\delta\beta} - 4RR^{\gamma}_{\beta}R^{\alpha}_{\gamma} + 8R^{\delta}_{\eta}R^{\eta\gamma}_{\delta\nu}R^{\nu\alpha}_{\gamma\beta} + 2R^{\mu\nu}_{\gamma\beta}R^{\lambda\delta}_{\mu\nu}R^{\gamma\alpha}_{\lambda\delta} + 8R^{\mu\gamma}_{\nu\beta}R^{\nu\lambda}_{\mu\delta}R^{\delta\alpha}_{\gamma\lambda} - 8R^{\delta}_{\eta}R^{\nu\alpha}_{\delta\gamma}R^{\eta\gamma}_{\nu\beta} \\ & - 4R^{\eta}_{\delta}R^{\delta\alpha}_{\nu\gamma}R^{\nu\gamma}_{\eta\beta} - 4R^{\mu}_{\nu}R^{\nu}_{\mu}R^{\alpha}_{\beta} + 8R^{\mu}_{\beta}R^{\nu}_{\mu}R^{\alpha}_{\nu} + 2RR^{\mu\nu}_{\gamma\beta}R^{\gamma\alpha}_{\mu\nu} - 4R^{\mu\nu}_{\gamma\lambda}R^{\gamma\eta}_{\mu\nu}R^{\lambda\alpha}_{\eta\beta} + 8R^{\mu}_{\nu}R^{\eta}_{\mu}R^{\nu\alpha}_{\eta\beta} \\ & + R^{\eta\gamma}_{\delta\lambda}R^{\delta\lambda}_{\eta\gamma}R^{\alpha}_{\beta} - \tfrac{1}{6}R^{3}\delta^{\alpha}_{\beta} + 2RR^{\gamma}_{\mu}R^{\mu}_{\gamma}\delta^{\alpha}_{\beta} - \tfrac{1}{3}R^{\mu\nu}_{\eta\gamma}R^{\lambda\sigma}_{\mu\nu}R^{\eta\gamma}_{\lambda\sigma}\delta^{\alpha}_{\beta} + \tfrac{4}{3}R^{\mu\nu}_{\eta\gamma}R^{\eta\lambda}_{\mu\sigma}R^{\gamma\sigma}_{\nu\lambda}\delta^{\alpha}_{\beta} - 4R^{\lambda}_{\nu}R^{\eta}_{\mu}R^{\nu\mu}_{\lambda\eta}\delta^{\alpha}_{\beta} \\ & - \tfrac{8}{3}R^{\gamma}_{\lambda}R^{\nu}_{\gamma}R^{\lambda}_{\nu}\delta^{\alpha}_{\beta} + 4R^{\nu}_{\gamma}R^{\lambda\eta}_{\mu\nu}R^{\gamma\mu}_{\lambda\eta}\delta^{\alpha}_{\beta} - \tfrac{1}{2}RR^{\eta\mu}_{\gamma\lambda}R^{\gamma\lambda}_{\eta\mu}\delta^{\alpha}_{\beta} - 4(R^{\eta}_{\beta}R^{\mu\nu}_{\lambda\eta}R^{\lambda\alpha}_{\mu\nu} + R^{\alpha}_{\eta}R^{\lambda\eta}_{\mu\nu}R^{\mu\nu}_{\lambda\beta}) \\ & - 8(R^{\mu}_{\nu}R^{\delta\nu}_{\mu\beta}R^{\alpha}_{\delta} + R^{\nu}_{\mu}R^{\mu\alpha}_{\delta\nu}R^{\delta}_{\beta}).\end{aligned} \tag{1.8}$$

It is now straightforward, though tedious, to show the relationships of $\overset{[2]}{G}{}^{\alpha}_{\beta}$ and $\overset{[3]}{G}{}^{\alpha}_{\beta}$ to parts of Lovelock's general tensor expression.[3] They are

$$\overset{[2]}{G}{}^{\alpha}_{\beta} = -\tfrac{1}{16}\delta^{\alpha\mu_1\mu_2\mu_3\mu_4}_{\beta\nu_1\nu_2\nu_3\nu_4} R^{\nu_1\nu_2}_{\mu_1\mu_2} R^{\nu_3\nu_4}_{\mu_3\mu_4}, \tag{1.9}$$

$$\overset{[3]}{G}{}^{\alpha}_{\beta} = -\tfrac{1}{64}\delta^{\alpha\mu_1\mu_2\mu_3\mu_4\mu_5\mu_6}_{\beta\nu_1\nu_2\nu_3\nu_4\nu_5\nu_6} R^{\nu_1\nu_2}_{\mu_1\mu_2} R^{\nu_3\nu_4}_{\mu_3\mu_4} R^{\nu_5\nu_6}_{\mu_5\mu_6}, \tag{1.10}$$

where

$$\delta^{\mu_1\mu_2\cdots\mu_N}_{\nu_1\nu_2\cdots\nu_N} \equiv \begin{vmatrix} \delta^{\mu_1}_{\nu_1} & \cdots & \delta^{\mu_1}_{\nu_N} \\ \vdots & & \vdots \\ \delta^{\mu_N}_{\nu_1} & \cdots & \delta^{\mu_N}_{\nu_N} \end{vmatrix}. \tag{1.11}$$

Note that to establish (1.10) by a direct calculation using (1.11) would involve the simplification and collection of $7! = 5{,}040$ terms, with some reduction in effort exploiting the symmetries of **Riemann**.

2. DERIVATION OF EQUATION (1.7)

The first step in the construction of (1.7) is to find all symmetric bilinear products of **Riemann**. To do this the following scheme is used. We introduce the notation in which

$$\overset{\mu \quad \lambda \quad \nu}{\bullet\!\!\longrightarrow\!\!\bullet}$$

will signify that there are μ contractions on the first **Riemann**, λ cross summations between first and second **Riemann**, and ν contractions on the second **Riemann**. Thus, e.g.,

$\overset{2 \quad 0 \quad 1}{\bullet\!\!\longrightarrow\!\!\bullet}$ will represent $R^{\mu\nu}_{\mu\nu} R^{\gamma\alpha}_{\gamma\beta}$

Now, it should be noted that, quite obviously, the total number of summations (which includes contractions) equals 3 if no δ^{α}_{β} term is present and that it equals 4 otherwise. So, we have the following list of all the possibilities [to save writing, in the following, the phrase "will represent" will be replaced by a single arrow, viz. $(\leftrightarrow)$]:

$$\begin{aligned}
&\overset{2\ 0\ 1}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R\ R^{\alpha}_{\beta},\\
&\overset{2\ 0\ 2}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R\ R\ \delta^{\alpha}_{\beta},\\
&\overset{1\ 2\ 1}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R^{\mu}_{\nu} R^{\nu}_{\mu} \delta^{\alpha}_{\beta},\\
&\overset{1\ 1\ 1}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R^{\alpha}_{\mu} R^{\mu}_{\beta},\\
&\overset{1\ 2\ 0}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R^{\mu}_{\nu} R^{\nu\alpha}_{\mu\beta},\\
&\overset{0\ 3\ 0}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R^{\eta\alpha}_{\mu\nu} R^{\mu\nu}_{\eta\beta},\\
&\overset{0\ 4\ 0}{\bullet\!\!\longrightarrow\!\!\bullet} \leftrightarrow R^{\eta\gamma}_{\mu\nu} R^{\mu\nu}_{\eta\gamma} \delta^{\alpha}_{\beta}.
\end{aligned} \tag{2.1}$$

The symmetry in α and β in the terms is easily shown when these are written in covariant notation, e.g.,

$$RR_{\alpha\beta} = RR_{\beta\alpha}, \quad \text{etc.}$$

Clearly, $\overset{[2]}{\mathbf{G}}$ must be a linear combination of the terms in (2.1), i.e.,

$$\begin{aligned}
\overset{[2]}{G}{}^{\alpha}_{\beta} = aRR^{\alpha}_{\beta} + bR^{2}\delta^{\alpha}_{\beta} + cR^{\mu}_{\nu}R^{\nu}_{\mu}\delta^{\alpha}_{\beta}&\\
+ dR^{\alpha}_{\mu}R^{\mu}_{\beta} + eR^{\mu}_{\nu}R^{\nu\alpha}_{\mu\beta}&\\
+ fR^{\eta\alpha}_{\mu\nu}R^{\mu\nu}_{\eta\beta} + gR^{\eta\gamma}_{\mu\nu}R^{\mu\nu}_{\eta\gamma}\delta^{\alpha}_{\beta}&.
\end{aligned} \tag{2.2}$$

Then, to find an expression for b, c, d, e, f, and g in terms of a in Eq. (2.2), the condition $\overset{[2]}{\mathbf{G}}{}^{\alpha}_{\beta;\alpha} = 0$ is imposed. This gives, on using the Bianchi identities and the symmetry of **Riemann**,

$$\begin{aligned}
\overset{[2]}{G}{}^{\alpha}_{\beta;\alpha}
&= (\tfrac{1}{2}a + 2b)RR_{;\beta} + (a + \tfrac{1}{2}d)R^{\alpha}_{\beta}R_{;\beta} + (d-e)R^{\alpha}_{\nu}R^{\nu}_{\beta;\alpha}\\
&\quad + (e + 2c)R^{\gamma}_{\alpha}R^{\alpha}_{\gamma;\beta} + (e + 2f)R^{\nu\alpha}_{\gamma\beta}R^{\gamma}_{\nu;\alpha}\\
&\quad + (f + 4g)R^{\eta\alpha}_{\mu\nu}R^{\mu\nu}_{\eta\beta;\alpha} = 0.
\end{aligned} \tag{2.3}$$

For this identity to hold the individual coefficients must vanish, i.e.,

$$\begin{aligned}
a + 4b &= 2a + d = 2d - 2e\\
&= e + 2c = e + 2f = f + 4g = 0.
\end{aligned} \tag{2.4}$$

This gives

$$a = -4b = c = -d/2 = -e/2 = f = -4g. \tag{2.5}$$

On substituting (2.5) into (2.2) and setting $a = 1$ (as an expression for $\overset{[2]}{\mathbf{G}}{}^{\alpha}_{\beta}$ is needed up to a multiplicative constant only), the required equation (1.7) is obtained.

3. DERIVATION OF EQUATION (1.8)

The construction of (1.8) is more complicated because symmetric trilinear products of **Riemann** are required. A similar scheme to that used in the previous section will be employed. We introduce the notation

$$\begin{array}{ccc} & \mu & \\ \nu & \triangle & \eta \leftrightarrow \\ \lambda & \delta & \gamma \end{array}$$

to mean there are μ contractions on the first **Riemann**, ν cross summations between the first and second **Riemann**, λ contractions on the second **Riemann**, δ cross summations between the second and third **Riemann**, γ contractions on the third **Riemann**, and η cross summations between the first and third **Riemann**. Thus, e.g.,

$$\begin{array}{ccc} & 2 & \\ 0 & \triangle & 0 \\ 1 & 2 & 1 \end{array} \leftrightarrow RR^{\delta}_{\gamma}R^{\gamma}_{\delta}\delta^{\alpha}_{\beta}.$$

To be sure that all possible combinations are found, it is sufficient to note that the sum of the numbers around the triangle must add up to either 5 or 6, with no individual number being greater than 4. So, writing out the set of such numbers will give all the possible combinations. They are, for the case where the numbers add up to 5,

$$\{1,1,1,1,1,0\} \leftrightarrow \begin{array}{ccc} & 1 & \\ 1 & \triangle & 0 \\ 1 & 1 & 1 \end{array} \ \text{or} \ \begin{array}{ccc} & 1 & \\ 1 & \triangle & 1, \\ 0 & 1 & 1 \end{array}$$

$$\{1,1,1,2,0,0\} \leftrightarrow \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0 \\ 1 & 1 & 1 \end{array} \ \text{or} \ \begin{array}{ccc} & 1 & \\ 2 & \triangle & 0 \\ 1 & 0 & 1 \end{array} \ \text{or} \ \begin{array}{ccc} & 1 & \\ 2 & \triangle & 0 \\ 0 & 1 & 1 \end{array} \ \text{or} \ \begin{array}{ccc} & 0 & \\ 2 & \triangle & 1, \\ 0 & 1 & 1 \end{array}$$

$$\{1,2,2,0,0,0\} \leftrightarrow \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0 \\ 2 & 0 & 1 \end{array} \ \text{or} \ \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0 \\ 0 & 2 & 1 \end{array} \ \text{or} \ \begin{array}{ccc} & 1 & \\ 2 & \triangle & 0 \\ 0 & 2 & 0 \end{array} \ \text{or} \ \begin{array}{ccc} & 0 & \\ 2 & \triangle & 2, \\ 0 & 1 & 0 \end{array}$$

$$\{1,1,3,0,0,0\} \leftrightarrow \begin{array}{ccc} & 1 & \\ 0 & \triangle & 1 \\ 0 & 3 & 0 \end{array} \text{ or } \begin{array}{ccc} & 0 & \\ 1 & \triangle & 3, \\ 0 & 1 & 0 \end{array}$$

$$\{2,3,0,0,0,0\} \leftrightarrow \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0, \\ 0 & 3 & 0 \end{array}$$

$$\{4,1,0,0,0,0\} \leftrightarrow \begin{array}{ccc} & 1 & \\ 0 & \triangle & 0. \\ 0 & 4 & 0 \end{array}$$

Similarly, for the case where the numbers add up to 6 we have

$$\{1,1,1,1,1,1\} \leftrightarrow \begin{array}{ccc} & 1 & \\ 1 & \triangle & 1, \\ 1 & 1 & 1 \end{array}$$

$$\{1,1,1,1,2,0\} \leftrightarrow \triangle \text{ No combination possible,}$$

$$\{1,1,2,2,0,0\} \leftrightarrow \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0 \\ 1 & 2 & 1 \end{array} \text{ or } \begin{array}{ccc} & 0 & \\ 2 & \triangle & 2, \\ 1 & 0 & 1 \end{array}$$

$$\{2,2,2,0,0,0\} \leftrightarrow \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0 \\ 2 & 0 & 2 \end{array} \text{ or } \begin{array}{ccc} & 0 & \\ 2 & \triangle & 2, \\ 0 & 2 & 0 \end{array}$$

$$\{1,1,1,3,0,0\} \leftrightarrow \begin{array}{ccc} & 1 & \\ 1 & \triangle & 1, \\ 0 & 3 & 0 \end{array}$$

$$\{3,3,0,0,0,0\} \leftrightarrow \triangle \text{ No combination possible,}$$

$$\{1,2,3,0,0,0\} \leftrightarrow \triangle \text{ No combination possible,}$$

$$\{2,4,0,0,0,0\} \leftrightarrow \begin{array}{ccc} & 2 & \\ 0 & \triangle & 0. \\ 0 & 4 & 0 \end{array}$$

It is then quite straightforward to write out the corresponding trilinear products of **Riemann**, except in the following five cases in each of which there are two distinct possible products corresponding to one diagram:

$$\text{(i) } \begin{array}{ccc} & 0 & \\ 2 & \triangle & 2 \\ 0 & 2 & 0 \end{array} \leftrightarrow \begin{cases} R^{\mu\nu}_{\ \ \eta\gamma} R^{\lambda\sigma}_{\ \ \mu\nu} R^{\eta\gamma}_{\ \ \lambda\sigma} \delta^{\alpha}_{\beta} & (3.1) \\ \text{or} & \\ R^{\mu\nu}_{\ \ \eta\gamma} R^{\eta\lambda}_{\ \ \mu\sigma} R^{\gamma\sigma}_{\ \ \nu\lambda} \delta^{\alpha}_{\beta}, & (3.2) \end{cases}$$

$$\text{(ii) } \begin{array}{ccc} & 1 & \\ 1 & \triangle & 1 \\ 0 & 2 & 0 \end{array} \leftrightarrow \begin{cases} R^{\delta}_{\ \eta} R^{\nu\alpha}_{\ \ \delta\gamma} R^{\eta\gamma}_{\ \ \nu\beta} & (3.3) \\ \text{or} & \\ R^{\eta}_{\ \delta} R^{\delta\alpha}_{\ \ \nu\gamma} R^{\nu\gamma}_{\ \ \eta\beta}, & (3.4) \end{cases}$$

$$\text{(iii) } \begin{array}{ccc} & 0 & \\ 2 & \triangle & 2 \\ 0 & 1 & 0 \end{array} \leftrightarrow \begin{cases} R^{\mu\nu}_{\ \ \gamma\beta} R^{\lambda\delta}_{\ \ \mu\nu} R^{\gamma\alpha}_{\ \ \lambda\delta} & (3.5) \\ \text{or} & \\ R^{\mu\gamma}_{\ \ \nu\beta} R^{\nu\lambda}_{\ \ \mu\delta} R^{\delta\alpha}_{\ \ \gamma\lambda}, & (3.6) \end{cases}$$

$$\text{(iv) } \begin{array}{ccc} & 1 & \\ 1 & \triangle & 0 \\ 0 & 3 & 0 \end{array} \leftrightarrow \{R^{\eta}_{\ \beta} R^{\mu\nu}_{\ \ \lambda\eta} R^{\lambda\alpha}_{\ \ \mu\nu} + R^{\alpha}_{\ \eta} R^{\lambda\eta}_{\ \ \mu\nu} R^{\mu\nu}_{\ \ \lambda\beta}, \tag{3.7}$$

$$\text{(v) } \begin{array}{ccc} & 1 & \\ 0 & \triangle & 2 \\ 1 & 1 & 0 \end{array} \leftrightarrow \{R^{\mu}_{\ \nu} R^{\delta\nu}_{\ \ \mu\beta} R^{\alpha}_{\ \delta} + R^{\nu}_{\ \mu} R^{\mu\alpha}_{\ \ \delta\nu} R^{\delta}_{\ \beta}. \tag{3.8}$$

In Eqs. (3.1)–(3.6), inclusive, there are two possible trilinear combinations for each triangle, and in (3.7) and (3.8) a symmetric combination of the trilinear products is formed.

It is now straightforward to show that all the trilinear combinations are symmetric and then to proceed as in the bilinear case. One, therefore, forms the tensor $\overset{[3]}{G}{}^{\alpha}_{\beta}$ by taking a linear combination of all the trilinear products and then evaluates the coefficients by using the condition

$$\overset{[3]}{G}{}^{\alpha}_{\ \beta;\alpha} = 0.$$

This then yields just the result (1.8) determined to within a multiplicative constant.

4. CONCLUSION

It has been shown above by a direct calculation that the natural generalizations of the Einstein tensor **G** to bilinearity and trilinearity in **Riemann** agree to within an arbitrary multiplicative constant with the second and third term, respectively, of the general tensor found by Lovelock (who used the assumption of linearity in $\partial^2 g\alpha\beta/\partial x^\nu \partial x^\nu$) and that it vanishes in the usual space-time (for which $n = 4$).

ACKNOWLEDGMENTS

The authors would like to thank M. Batchelor and W. Padden for assisting with and checking the calculations.

[1]From this point on, the restriction that space-time is four-dimensional is not necessary, and it will be assumed that space-time has dimension n unless stated otherwise.

[2]The following conventions have been adopted:

(a) $G^{\alpha}_{\beta} \equiv G^{\alpha}{}_{\beta}$, $R^{\mu\nu}_{\ \ \lambda\delta} \equiv R^{\mu\nu}{}_{\lambda\delta}$;

(b) $R^{\mu}_{\ \nu\alpha\beta} = \Gamma^{\mu}_{\ \nu\beta,\alpha} - \Gamma^{\mu}{}_{\nu\alpha,\beta} + \Gamma^{\mu}{}_{\sigma\alpha}\Gamma^{\sigma}{}_{\nu\beta} - \Gamma^{\mu}{}_{\sigma\beta}\Gamma^{\sigma}_{\ \nu\alpha}$;

(c) The Einstein summation convention will be used;

(d) $\Gamma^{\mu}_{\ \nu\beta} = \frac{1}{2} g^{\mu\gamma}(g_{\gamma\beta,\nu} + g_{\gamma\nu,\beta} - g_{\nu\beta,\gamma})$;

(e) A comma denotes partial differentiation, and a semicolon denotes covariant differentiation;

(f) Greek indices run from 1 to n, unless otherwise specified;

(g) $R_{\mu\nu} = R^{\alpha}{}_{\mu\alpha\nu}$;

(h) $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$.

[3]This general tensor has the form

$$\sum_{\gamma=1}^{\infty} a_\gamma \delta^{\mu\mu_1\cdots\mu_{2\gamma}}_{\nu\nu_1\cdots\nu_{2\gamma}} R^{\nu_1\nu_2}_{\ \ \mu_1\mu_2} \cdots R^{\nu_{2\gamma-1}\nu_{2\gamma}}_{\ \ \mu_{2\gamma-1}\mu_{2\gamma}} + a\delta^{\mu}_{\nu},$$

where a_γ and a are constants and $\delta^{\mu_1\cdots\mu_n}_{\nu_1\cdots\nu_n}$ is the generalized Kronecker delta. This expression can be found in the paper by D. Lovelock, J. Math. Phys. **13**, 874 (1972). This excellent paper became known to the authors only after the results contained in this paper had been derived as part of a general investigation of the mathematical properties of the Einstein tensor.

Positivity of the Bondi energy

O. Reula[a)]
Department of Physics and Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
K. P. Tod[b)]
Mathematical Institute, 24–29 St. Giles, Oxford OX1 3LB, England

(Received 30 August 1983; accepted for publication 7 October 1983)

We give a new proof that the Bondi momentum is time-like and future directed. This is done by extending a theorem showing the existence of solutions of Witten's equation on a complete, asymptotically flat surface, to the case where the surface is asymptotically hyperbolic and spans a cut $\hat{S}$ of $\mathscr{I}^+$. By choosing appropriate boundary conditions the positivity of the Bondi energy is also shown for the case where the surface is no longer free of inner boundaries but is internally bounded by the union of a finite number of marginally trapped surfaces. This argument also applies to the case of asymptotically flat surfaces.

PACS numbers: 04.20.Cv, 02.30. + g

I. INTRODUCTION

Recently, several authors[1–4] have considered the issue of extending Wittens proof of the the positivity of energy at spatial infinity[5] to a proof in the case of null infinity, i.e., to establish the positivity of the Bondi mass. The general technique is to use the same identity Witten uses for the spatial case,[5] but on a space-like surface Σ which is asymptotically null and which meets null infinity in a cut $\hat{S}$. The identity equates the Bondi energy at $\hat{S}$, expressed as a limit of 2-surface integrals on Σ, to a 3-surface integral over Σ which is manifestly positive provided solutions of the elliptic equation considered in Ref. 5 exist. An existence theorem for this equation on an asymptotically null hypersurface would therefore be sufficient to prove positivity of the Bondi energy.

In the absence of such an existence theorem, earlier authors have instead completed the proof of positivity by considering either a different elliptic equation[4] or a null rather than asymptotically null hypersurface for Σ,[2] and proving existence of solutions for those cases.

While the result itself is thus now established[2,4,6,7] it is still of interest to known whether precisely the same technique can be used in both the spatial and null infinity cases. This is what we do here. That is, we provide an existence theorem for the equation considered in Ref. 5 but on an asymptotically null surface (in fact, on a surface of asymptotically constant curvature). The proof of this theorem is closely related to an earlier existence proof for positivity at spatial infinity.[8] In fact, the proof in Ref. 8. contains an error, which we rectify here. We thus provide a unified treatment of the two cases: there is just one identity involved and one equation to be solved. The difference is in the boundary conditions and the details of the proof.

Witten's original argument and the subsequent extensions mentioned above require the existence of a space-like slice (i.e., a space-like surface without inner boundaries) of the space-time reaching infinity, and the validity of the dominant energy condition in such a slice. But we know there are examples of space-times having regions causally disconnected from infinity. It could happen that the space-like slice Σ encountered singularities in such a region or that the dominant energy condition was violated in such a region. Even so, one would still hope to be able to prove positivity of the energy at infinity by considering the parts of Σ outside such regions. Thus we will consider the question of inner boundaries on Σ, in Sec. IV. More specifically we shall consider the case when the inner boundary of Σ, S_0, is given by the union of marginally (future or past) trapped 2-surfaces. We shall find a boundary condition at S_0 which again leads to positivity of the Bondi energy. This boundary condition has also been found by Gibbons *et al.*[7]

II. POSITIVITY OF THE BONDI MASS AND EXISTENCE THEOREM

Consider an everywhere space-like, asymptotically null slice Σ (a surface without boundary)[9] of an asymptotically flat space-time (M, g_{ab}). Let t^a be its unit normal, S a topologically S^2 hypersurface of Σ, and $\Sigma(S)$ the interior of S. Then for any smooth spinor field α^A, the two-component version of Witten's identity[5,10] is[2,3]

$$\int_S F_{ab}[\alpha]dS^{ab} = \int_{\Sigma(S)} (8\pi T_{ab}k^a t^b - 2t^{AA'}D_b\alpha_A D^b\bar{\alpha}_{A'} - 4t^{AA'}D_{A'B}\alpha^B D_{AB'}\bar{\alpha}^{B'})d\Sigma, \qquad (2.1)$$

where

$$F_{ab} = i(\bar{\alpha}_{A'}\nabla_b\alpha_A - \bar{\alpha}_{B'}\nabla_a\alpha_B) + \text{complex conjugate},$$
$$k_a = \alpha_A\bar{\alpha}_{A'},$$
$$D_b = h_b{}^c\nabla_c = (g_b{}^c - t_b t^c)\nabla_c,$$

and T_{ab} is the energy momentum tensor of the matter contained in (M, g_{ab}).[11]

As remarked by several authors, if α^A is asymptotically constant, as defined in Ref. 12, and S recedes to infinity approaching a cut $\hat{S}$ of $\mathscr{I}^+$,[13] the left-hand side in (2.1) tends to the component of the Bondi momentum at $\hat{S}$ in the direction of the asymptotic translation defined by k_a.[1–4] On the right side, the first term is positive if the dominant energy condition is imposed, and the second is automatically positive

[a)]Supported in part by the Aillen S. Andrew Foundation and by the NSF.
[b)]S.E.R.C. Advanced Fellow.

 0022-2488/84/041004-05$02.50

since the metric of Σ is negative definite. The third term, however, is negative, so one equates it to zero. This requires,

$$D_{AA'}\alpha^A = 0. \tag{2.2}$$

Then positivity of the Bondi energy at $\hat{S}$ is ensured if one can find asymptotically constant solutions of (2.2) on a surface Σ spanning $\hat{S}$.

We will prove existence of these solutions for any surface satisfying the following condition:

Definition: An asymptotically null surface (Σ, h_{ab}) is said to be *asymptotically hyperbolic* (AH) if there exists a hyperbolic metric K_{ab} outside some compact set, and constant $C_1 > 0$ such that for any vector V^a,

$$C_1^{-1} K_{ab} V^a V^b \leqslant h_{ab} V^a V^b \leqslant C_1 K_{ab} V^a V^b, \tag{2.3}$$

where h_{ab} is the induced metric on Σ.[14]

Existence theorem: Let (Σ, h_{ab}) be an asymptotically hyperbolic surface embedded in (M, g_{ab}), an asymptotically flat space-time satisfying the dominant condition[15] on Σ, and $\mathring{\alpha}^A$ any asymptotically constant[12] spinor field. Then there exists a unique spinor field α^A, approaching $\mathring{\alpha}^A$ asymptotically, and satisfying

$$D_{AA'}\alpha^A = 0. \tag{2.2}$$

Proof: Following Ref. 8 we define the Hilbert space H as the completion of the space C_0^∞ (smooth and compactly supported) spinor fields σ^A on Σ in the norm

$$(\|\sigma^A\|_H)^2 = \int_\Sigma t^{AA'} D_{BA'}\sigma^B D_{AB'}\bar{\sigma}^{B'}\, d\Sigma. \tag{2.4}$$

We then have a key lemma.

Generalized Hardy lemma: There exists a constant $C > 0$ such that for any spinor field $\sigma^A \in C_0^\infty$,

$$(\|\sigma^A\|_H)^2 \geqslant C \int_\Sigma \frac{t^{AA'}\sigma_A \bar{\sigma}_{A'}}{r^2}\, d\Sigma. \tag{2.5}$$

We prove this in the next section.[16] From this inequality it follows that the elements of H are measurable fields and that, if σ_m^A is a C_0^∞ Cauchy sequence converging in H to σ^A, then $D_{AA'}\sigma_m^A$ weakly converges to the distributional derivative $D_{AA'}\sigma^A$ of σ^A. Thus the integral (2.4) makes sense for any element of H, not just the C_0^∞ ones, and gives the norm.

We now seek a solution of (2.2) of the form $\alpha^A = \mathring{\alpha}^A + \beta^A$ with $\beta^A \in H$. This is accomplished by the Riesz representation theorem[17] as follows. We define a linear functional $f(\cdot)$ on H by

$$f(\sigma^A) = -\int_\Sigma t^{AA'} D_{AB'}\bar{\sigma}^{B'} D_{BA'}\mathring{\alpha}^B\, d\Sigma. \tag{2.6}$$

Since $\mathring{\alpha}^A$ is an asymptotically constant spinor field, it follows from the results in Ref. 2 that

$$(\|D_{AA'}\mathring{\alpha}^A\|_{L^2})^2 = \int_\Sigma t^{AA'} D_{AB'}\bar{\mathring{\alpha}}^{B'} D_{BA'}\mathring{\alpha}^B\, d\Sigma < \infty. \tag{2.7}$$

So, using the Cauchy–Schwartz inequality for spinors[18] we have

$$|f(\sigma^A)| \leqslant \|\sigma^A\|_H \|D_{BA'}\mathring{\alpha}^B\|_{L^2}. \tag{2.8}$$

Thus $f(\ )$ is a bounded linear functional on H and by the Riesz representation theorem in H, there exists one and only one $\beta_A \in H$ with

$$f(\sigma^A) = \int_\Sigma t^{AA'} D_{AB'}\bar{\sigma}^{B'} D_{BA'}\beta^B\, d\Sigma. \tag{2.9}$$

Subtracting (2.6) from (2.9), restricting σ^A to be in C_0^∞, and integrating by parts leads to

$$\int_\Sigma t^{AA'} D_{A'}{}^B D_{BB'}\sigma^{B'}(\mathring{\alpha}_A + \beta_A)\, d\Sigma = 0$$

for all $\sigma^A \in C_0^\infty$. Thus we have shown that there exists a unique β^A such that $\alpha^A = \mathring{\alpha}^A + \beta^A$ satisfies the weak form of equation

$$D_A{}^{B'} D_{BB'}\alpha^B = 0. \tag{2.10}$$

Standard theorems on elliptic equations[19] show that this α^A is smooth and, therefore, satisfies this equation in its strong form, i.e., (2.10).

It remains to be proved that α^A also satisfies the first-order equation (2.2). Assume, for contradiction, that $\bar{\lambda}_{B'} := D_{BB'}\alpha^B \neq 0$; then from (2.10) we have a solution $\bar{\lambda}_{B'}$ of (2.2), which is in $L^2 \cap C^\infty$. But substituting it in Witten's identity, (2.1), leads immediately to a contradiction.[20]

Finally, one must check that the growth rate of the β^A obtained is compatible with the limit of the surface integral in (2.1), being the Bondi energy. But, since the solution is unique it must coincide with the asymptotic solution found in Ref. 3 for (2.2), which *is* compatible with the above limit.

III. PROOF OF THE GENERALIZED HARDY LEMMA[21]

Let (M, g_{ab}) be an asymptotically flat space-time, let Σ be an everywhere space-like, complete, asymptotically hyperbolic slice of (M, g_{ab}), with unit normal t^a. Then, there exists a constant $C > 0$ such that for any spinor field $\alpha^A \in C_0^\infty$ the following inequality holds:

$$(\|\alpha^A\|_H)^2 \geqslant C \int_\Sigma \frac{t^{AA'}\alpha_A \bar{\alpha}_{A'}}{r^2}\, d\Sigma.$$

Proof: We first prove the theorem for regions of Σ which, with respect to the metric K_{ab} are cones[22] (of arbitrary solid angle). Then we extend the proof to all of Σ, using the fact that it can be covered by a finite number of cones.

Besides (2.3), two other geometrical inequalities are needed:

(i) If V^a is any null vector, and n^a any unit time-like vector, then

$$\varphi V_a n^a \geqslant V_a t^a \geqslant \varphi^{-1} V_a n^a, \quad \text{where } \varphi = 2 n_a t^a.$$

(ii) If the above vector field n^a is chosen such that $K^{ab}\nabla_a r \nabla_b n_c = 0$, then there exists a constant $C' > 0$ such that $2 \leqslant \varphi \leqslant 2C'\sqrt{1+r^2} = C'\varphi_0$.

Let γ be any cone on Σ, n^a a unit time-like vector field on γ satisfying $K^{ab}\nabla_a r \nabla_b n_c = 0$, and α^A any C_0^∞ spinor field; then

$$
\begin{aligned}
(\|\alpha^A\|_H)^2 &\geqslant \frac{1}{2}\int_\gamma D_a\alpha_A D^a\bar\alpha_{A'} t^{AA'}\,d\Sigma \\
&\geqslant \frac{1}{2}\int_\gamma D_a\alpha_A D^a\bar\alpha_{A'} n^{AA'}\,\frac{d\Sigma}{\varphi} \\
&\geqslant C''\int_\gamma K^{ab} D_a\alpha_A D_b\bar\alpha_{A'} n^{AA'}\,\frac{d\Sigma_0}{\varphi_0} \\
&\geqslant C''\int_\gamma K^{ab}((-D_a r D_b r\alpha_A\bar\alpha_{A'} n^{AA'})/4r^2\varphi_0)d\Sigma_0 \\
&\quad + \frac{C''}{2}\int_0^\infty \frac{d}{dr}\left[\int_{S^2} r\alpha_A\bar\alpha_{A'} n^{AA'}\,d\Omega\right]dr \\
&\quad - C''\int_\gamma r^{-1}D_a r(D_b n^{AA'})\alpha_A\bar\alpha_{A'} K^{ab}\frac{d\Sigma_0}{\varphi_0},
\end{aligned}
$$

where in the first step we have used Witten's identity (2.1) and the dominant energy condition. In the subsequent steps all the three bounds given above, and the identity $r^{-1/2}D_a(\alpha_C r^{1/2}) = D_a\alpha_C + \frac{1}{2}\alpha_C r^{-1}D_a r$, have been used. The differential $d\Sigma_0$ is the surface element corresponding to K_{ab} $[d\Sigma_0 = (1+r^2)^{-1/2}r^2\,dr\,d\Omega]$. The last two terms vanish, one because it is an exact differential, the other because of the way n^a has been chosen. Thus, using once more the above inequalities we conclude

$$
\int_\gamma t^{AA'} D_{AB'}\bar\alpha^{B'} D_{BA'}\alpha^B\,d\Sigma \geqslant C\int_\gamma \frac{t^{AA'}\alpha_A\alpha_{A'}}{r^2}d\Sigma. \quad (3.1)
$$

To finish the proof we must extend the inequality (3.1) from the cone γ to all of Σ. Assume, for contradiction, that (3.1) with $\gamma = \Sigma$ fails, then there exists a sequence of C_0^∞ spinor fields $\underset{n}{\alpha}_A$ on Σ such that

$$
\int_\Sigma t^{AA'} D_{AB'}\underset{n}{\bar\alpha}^{B'} D_{BA'}\underset{n}{\alpha}^B\,d\Sigma = 1, \quad \int_\Sigma \frac{t^{AA'}\underset{n}{\alpha}_A\underset{n}{\bar\alpha}_{A'}}{r^2}\,d\Sigma = n.
$$

But since Σ can be covered by a finite number of cones, there must exist at least one cone for which the right side of (3.1), for $\underset{n}{\alpha}^A$, grows without limit as n goes to infinity, while the left side of (3.1) remains finite, leading to a contradiction.

It has been realized that the proof of the generalized Sobolev lemma in Ref. 8 is incorrect. Nevertheless the main theorem of the paper still holds true since the generalized Hardy lemma, which is still true when the surface Σ is asymptotically flat, serves the same purposes that were sought in the former lemma. To prove the generalized Hardy lemma on an asymptotically flat surface Σ, just notice that in this case the bounds in (2.3) are still valid if K_{ab} is taken to be a flat metric on Σ, and that with n^a defined as above, then $2 \leqslant \varphi \leqslant C'\varphi_0 = 2C'$ (which in fact requires $g_{ab,c} \sim 1/r^{1+\epsilon}, \epsilon > 0$).

IV. INNER BOUNDARIES

We now suppose that Σ has an interior boundary S_0 as well as the previously introduced exterior boundary $\hat S$ at infinity. Then (2.1) acquires an inner surface term, and for any solution α^A of Witten's equation, (2.2), we have

$$
\int_{\hat S} F_{ab}(\alpha)dS^{ab} = -\int_{S_0} F_{ab}(\alpha)dS^{ab} + \int_\Sigma [8\pi G T_{ab}t^a k^b - 2t^{AA'}D_b\alpha_A D^b\bar\alpha_{A'}]\,d\Sigma. \quad (4.1)
$$

Now, if α^A tends to a constant spinor field $\mathring{\alpha}^A$ at infinity we conclude

$$
P_a\mathring{k}^a \geqslant -\int_{S_0} F_{ab}(\alpha)dS^{ab}, \quad (4.2)
$$

where P_a is the Bondi momentum and $\mathring{k}^a$ is the asymptotic translation defined by $\mathring{\alpha}^A$. Thus, to prove positivity of mass in this case we require boundary conditions for the spinor field α^A at the inner boundary S_0 such that:

(i) Solutions of Witten's equation exist satisfying them.

(ii) The right-hand side of (4.2) is nonnegative.

These two conditions are in some sense related: Indeed, spinor fields in $C_0^\infty(\bar\Sigma)$ satisfying condition (ii) (nonnegativity) also satisfy inequality (2.5) (the generalized Hardy lemma), which is a key step in proving condition (i) (existence).

We now obtain such boundary conditions. The boundary integrals in (4.1) are conveniently expressed in the GHP formalism[23] for two preferred null directions, in this case the ones picked out by S_0. Introduce a spinor dyad (O_A, ι_A) in a neighborhood of S_0, such that $O_A\iota^A = 1$; $O^A\bar O^{A'}$ is tangent at points of S_0 to the outgoing null geodesics orthogonal to S_0 and $\iota^A\bar\iota^{A'}$ is tangent to the ingoing ones. Then, the integrand (over S_0) in (4.1) is

$$
F_{ab}(\alpha)t^{[a}\xi^{b]} = \rho\alpha^0\bar\alpha^{0'} - \rho'\alpha^1\bar\alpha^{1'} + \bar\alpha^{1'}\eth\alpha^0 - \bar\alpha^{0'}\bar\eth\alpha^1 \quad (4.3)
$$

where ξ^a is the unit normal to S_0 in Σ, α^0 and α^1 are such that $\alpha^A = \alpha^0 O^A + \alpha^1\iota^A$, ρ, ρ' are, respectively, the convergences of the out- and ingoing null directions orthogonal to S_0, and $\eth, \bar\eth$ are the differential operators[23] tangent to S_0.

Examining expression (4.3) for $F_{ab}(\alpha)$, it is clear that in general there is no boundary condition satisfying (i) and (ii) simultaneously. This is so because to have solutions of (2.2) one expects to have to give on S_0 only "half" of α^A, e.g., to give either the value of α^0 or of α^1 on S_0 but not both. The other half would then be determined by the solution inside Σ. But then there is no way of forcing the integrand (4.3) to be positive having only one of these. On the other hand, this must be so because otherwise one could prove positivity for almost anything!

We now restrict attention to the case in which S_0 is a marginally future trapped surface.[24] This means that $\rho = 0$, i.e., that the outgoing null geodesics orthogonal to S_0 are just failing to expand. In this case, a suitable boundary condition at S_0 is $\alpha^1|_{S_0} = 0$, a condition also found by Gibbons *et al.* in Ref. 7. It is easy to see that this boundary condition satisfies (ii); but it also satisfies (i), as follows from the following theorem.

Existence theorem (with trapped surfaces): Let (M, g_{ab}) be an asymptotically flat space-time and (Σ, h_{ab}) an embedded, asymptotically hyperbolic surface such that: (i) the dominant energy condition[15] is satisfied in a neighborhood of Σ, and (ii) the interior boundary of Σ, $S_0 = \partial\Sigma$, is a future trapped 2-surface. Let $\mathring{\alpha}^A$ be any asymptotically constant[12] spinor field on Σ which vanishes in a neighborhood of S_0.

Then, there exists a unique spinor field α^A, which satisfies-$D_{AA'}, \alpha^A = 0$ and $\alpha^1|_{S_0} = 0$, and which approaches $\mathring{\alpha}^A$ asymptotically.

The proof of this theorem is very similar to the one in Sec. II, so we merely sketch it. First, one proves existence of solutions of the second-order Eq. (2.10) with boundary conditions $\alpha^1|_{S_0} = 0, \bar{\lambda}^{1'}|_{S_0} \equiv D^{2'}{}_A \alpha^A|_{S_0} = 0$. To do this, one chooses $H = \{$completion of spinor fields σ^A in $C_0^\infty(\bar{\Sigma})$ with $\sigma^1, D^{1'}{}_A \sigma^A$ in $C_0^\infty(\Sigma)$; under the norm given by (2.4)$\}$. Proceeding as before one shows the existence of $\beta^A \in H$ such that $\alpha^A = \mathring{\alpha}^A + \beta^A$ satisfies (2.10). Since $\mathring{\alpha}^A$ was chosen to vanish in a neighborhood of S_0, then one checks[25] that β^A satisfies the required boundary conditions. Second, one used the fact that S_0 is a future trapped surface, and that $\bar{\lambda}^{1'}|_{S_0} = 0$ to cancel a boundary integral which in turn allows the application of the same argument as before to conclude that α^A must also satisfy the first-order equation (2.2).

The proof above can be immediately extended to the case of S_0 being a marginally past trapped surface [by choosing as new boundary conditions $\alpha^0|_{S_0} = 0, \bar{\lambda}^{0'}|_{S_0} = D^{0'}_A \alpha^A|_{S_0} = 0$], and even to the case of S_0 being the union of several marginally trapped, past or future, surfaces.

V. CONCLUSION

We have proven an existence theorem of solutions of Witten's equation (and thus the positivity of total mass) for the case in which the space-like slice Σ on which the equation is given is asymptotically null. This proof, together with that of Ref. 8 in the space-like case, provides a unified treatment for the two regimes.

The proof requires nothing about the topology of Σ and rather little about its geometry: it must only be asymptotically hyperbolic and singularity-free. (These conditions suffice to guarantee that, among other things, Σ can be covered by a finite number of suitable cones, as required in Sec. III.) Thus, for example, one can allow in Σ a finite number of handles—or even throats connecting different asymptotic regions—and still establish nonnegativity of total mass.

We have also proved (Sec. IV.) positivity of total mass for some black holes, using a generalization of the above theorem. In this generalization, we allow in Σ an inner boundary consisting of a union of marginally-trapped surfaces. Thus, the total mass of a black hole is positive if it shelters a marginally-trapped surface. Note that this result requires some (although limited) information about the interior of the black hole. This requirement seems to be rather unphysical, since the interior of a black hole cannot be seen from the asymptotic region, where the total mass is manifest. It is at least possible there could exist a black hole without any marginally-trapped surface. Can one find a more general proof which also applies to such a black hole? One might try to generalize the above theorem, this time taking as the inner boundary of Σ not a marginally trapped surface, but rather a cross section of the event horizon. Unfortunately it is unlikely that one could find a proof along these lines for the event horizon, as a global concept, does not have enough useful local properties.

Let (M, g_{ab}) be an asymptotically flat space-time satisfying the dominant energy condition. Let $\hat{S}$ be a cut of null infinity and let Σ be a space-like slice spanning it. Then we know, from the theorem above, that the Bondi mass associated with this cut is nonnegative. We further know, by the Bondi flux formula (or, alternatively, by deforming the slice Σ near infinity and again applying the theorem), that the Bondi mass associated with any cut to the past of Σ is also nonnegative. Can we reach the same conclusion for cuts to the future of $\hat{S}$? Singularities could develop in such a way that no slice could span some of these cuts, thus preventing the application of the above theorem. If, however, strong asymptotic predictability is assumed, then for any cut to the future of $\hat{S}$ there necessarily exists an everywhere regular and space-like slice spanning it. Now we can apply the theorem to such cuts. We thus establish nonnegativity of the Bondi mass for all times.

ACKNOWLEDGMENTS

It gives us pleasure to acknowledge discussions with A. Ashtekar, R. Geroch, and R. M. Wald in connection with Sec.III, and with G. Gibbons and N. J. Hitchin in connection with Sec. IV.

[1] J. Nester and W. Israel, Phys. Lett. A **85**, 259 (1981).
[2] M. Ludvigsen and J. Vickers, J. Phys. A **14**, L389 (1981); J. Phys. A **15**, L67–70 (1982); "A Simple Proof of the Positivity of the Bondi Mass," University of York, preprint, 1981.
[3] G. Horowitz and K. P. Tod, Commun. Math. Phys. **85**, 429 (1982).
[4] G. Horowitz and M. Perry, Phys. Rev. Lett. **48**, 371 (1982).
[5] E. Witten, Commun. Math. Phys. **80**, 381 (1981).
[6] R. Shoen and S. T. Yau, Phys. Rev. Lett. **48**, 369 (1982).
[7] G. W. Gibbons, S. W. Hawking, G. Horowitz, and M. J. Perry, Commun. Math. Phys. **88**, 295 (1983).
[8] O. Reula, J. Math. Phys. **23**, 810 (1982).
[9] This condition will be relaxed in Sec. IV.
[10] J. Nester, Phys. Lett. A **83**, 241 (1981).
[11] The energy momentum tensor T_{ab} appears by imposing Einstein's equation on (M, g_{ab}).
[12] B. D. Bramson, Proc. R. Soc. London Ser. A **341**, 451 (1975).
[13] The generalization to space-times with many ends is straightforward: S will consist of the union of a finite number of topologically S^2 surfaces.
[14] Clearly Σ will be asymptotically hyperbolic if it is chosen to be asymptotically of constant curvature. If we take the cut $\hat{S}$ to be the origin of u in a Bondi chart, and define Σ by an equation of the form $u = -r/2 + O(r^{-3})$; then Σ will be asymptotically hyperbolic.
[15] That is, the Einstein tensor field G_{ab} when contracted with any future-directed non-space-like vector field must result in a future-directed non-space-like vector field.
[16] Here r is the radial distance corresponding to the hyperbolic metric k_{ab} assumed to exist outside some compact set of Σ.
[17] See, for example, M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. 1: Functional Analysis* (Academic, New York, 1979).
[18] The Cauchy–Schwartz inequality for spinors is

$$|t^{AA'} \gamma_A \bar{\delta}_{A'}|^2 \leqslant (t^{AA'} \gamma_A \bar{\gamma}_{A'})(t^{BB'} \delta_B \bar{\delta}_{B'}),$$

where $t^{AA'}$ is any unit time-like vector.
[19] A. Friedman, *Partial Differential Equations* (Holt, Rinehart, and Winston, New York, 1969), p. 54.
[20] Subtituting $\bar{\lambda}_{B'}$ in Witten's identity and using the dominant energy condition we get

$$\int_{\Sigma(r)} (-D_a \lambda_C D^a \bar{\lambda}_{C'}) t^{CC'} d\Sigma \leqslant \int_{S(r)} |\lambda_C D_a \bar{\lambda}_{C'}| t^{CC'} dS^a.$$

Here, $S(r)$, with $r \in [r_0 \geqslant 0, \infty)$, denotes a typical element of a one-parameter family of nested surfaces (topologically S^2), and $\Sigma(r)$ denotes the region enclosed by $S(r)$. The surfaces and the parameters are to be chosen so that

$d\Sigma > dr\, dS$. Calling $g^2(r)$ the integral of $-D_a\lambda_C D^a \lambda_{C'} t^{CC'}$ over $S(r)$, and using the Hölder inequality in the right-hand side of the above expression we get

$$\int_{r_0}^{r} g^2(x)dx \leqslant C\, g(r), \quad \text{where } C^2 = \int_{S(\infty)} \lambda_C \bar{\lambda}_{C'} t^{CC'}\, ds^a .$$

But one can show there exists no positive function $g(r)$ defined everywhere on $[r_0 > 0, \infty)$ satisfying the above inequality.

[21]This name arises from a similar lemma for scalars on R of G. H. Hardy, see G. H. Hardy *et al.*, *Inequalities* (Cambridge University, Cambridge, England, 1934). Also see M. Reed and B. Simon, *Method of Modern Mathematical Physics* (Academic, New York, 1975), Vol. 2, p. 169.

[22]A cone $\{p,\gamma,\theta\}$ is the region of Σ given by $\{q \in \Sigma$, such that the distance (with respect to k_{ab}) between q and γ is less or equal to $\theta s\}$, where γ is a geodesic, with respect to k_{ab}, connecting the point p (the apex) and infinity, and s its affine parameter.

[23]R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).

[24]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University, Cambridge, England, 1973), p. 321.

[25]Since $\beta^A \in H$, β^1 is the limit, under the norm in H, of a sequence $\{\beta^1_i\}$ of functions in $C_0^\infty(\Sigma)$. To see that $\beta^1|_{S_0} = 0$ let φ be any smooth function with support in a bounded neighborhood Ω of S_0, and use the smoothness of β^A as a solution of (2.10) to obtain

$$\begin{aligned}\left|\int_{S_0} \varphi(\beta^1 - \beta^1_i)dS_0\right| &= \left|\int_\Sigma \xi^a D_a(\varphi(\beta^1 - \beta^1_i))d\Sigma\right| \\ &\leqslant \|\beta^1 - \beta^1_i\|_{L^2(\Omega)} \|D\varphi\|_{L^2(\Omega)} \\ &\quad + \|\beta^1 - \beta^1_i\|_H \|\varphi\|_{L^2(\Omega)} \\ &\leqslant C\|\beta^1 - \beta^1_i\|_H \underset{i\to\infty}{\longrightarrow} 0\end{aligned}$$

for some constant C. Since φ is arbitrary in S_0 and $\beta^1_i|_{S_0} = 0$ then $\beta^1|_{S_0} = 0$.

To show that $D^{1'}_B \alpha^B|_{S_0} = 0$ consider the equation

$$\int_\Sigma t^{AA'}[D_{A'}{}^B D_{BB'} \bar{\sigma}^{B'}]\alpha_A\, d\Sigma = 0$$

for all $\bar{\sigma}^{B'} \in C_0^\infty(\bar{\Sigma}) \cap H$, which is just the statement that α^A is a weak solution of (2.10). Integrating it twice by parts one obtains

$$\int_\Sigma t^{AA'}\bar{\sigma}_{A'} D_{AB'} D^{B'}{}_B \alpha^B\, d\Sigma + \int_{S_0}[\alpha^1 D_{B'}\bar{\sigma}^{B'} + \alpha^0 D^1{}_{B'}\bar{\sigma}^{B'} - \bar{\sigma}^{1'} D^{0'}_B \alpha^B - \bar{\sigma}^{0'} D^{1'}_B \alpha^B]\, dS_0 = 0.$$

The integral over Σ vanishes, because α^A is smooth and because it does so for σ^A restricted to $C_0^\infty(\Sigma)$. The first term on the boundary integral vanishes because $\alpha^1|_{S_0} = 0$, the second and third because $\sigma^A \in C_0^\infty(\bar{\Sigma}) \cap H$, but $\bar{\sigma}^{0'}$ can take arbitrary values at S_0, therefore, $D^{1'}_B \alpha^B$ must vanish on S_0.

Hamiltonian formulation of gravitating perfect fluids and the Newtonian limit[a)]

H. P. Künzle
Max-Planck-Institut für Physik und Astrophysik, Institut für Astrophysik, D-8046 Garching bei München, Federal Republic of Germany and Department of Mathematics, University of Alberta, Edmonton, Canada T6G 2G1[b)]

J. M. Nester
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Canada T6G 2J1

(Received 25 February 1983; accepted for publication 24 June 1983)

The canonical formalism is applied to self-gravitating perfect fluids with particular emphasis on recovering the correct nonrelativistic limit also in (quasi-) Hamiltonian form. We use essentially Lagrangian coordinates by considering the fluid defined by a map from space-time into a three-dimensional material manifold which is equipped with a volume element representing physically the matter (baryon number) density. By eliminating the coordinate freedom in this material space the usual matter conservation and (relativistic) Euler equations are recovered in a $(3 + 1)$-dimensional formalism which makes it very easy to compare them to their nonrelativistic limits. By splitting the 3-metric and its canonical momenta into a conformal part and the determinant we arrive at a system of evolution and constraint equations for the gravitational field that also has a well-defined Newtonian limit provided the geometric version of the Newtonian theory is also cast into an analogous $(3 + 1)$-dimensional form. Some of the evolution equations of the relativistic theory, however, become additional constraints in the limit which represents the freezing of the gravitational (or radiation) degrees of freedom. We then use this formalism to rederive the first-order post-Newtonian approximation and obtain the standard results in a flexible geometrical form since no gauge or coordinate conditions need be imposed in advance.

PACS numbers: 04.20.Fy, 04.40. + c, 02.40. + m

1. INTRODUCTION

A great deal of effort has gone into the study of the Hamiltonian (canonical, symplectic) formulation of general relativity, mainly in the form of the so-called ADM formalism.[1] (For two recent surveys with different emphasis see Refs. 2 and 3.) This approach starts with a Lagrangian for the gravitational and possibly other fields and ends up, at least in the form of Fischer and Marsden,[3] with a (somewhat implicitly defined) symplectic structure on the space of all true gravitational degrees of freedom which is some infinite-dimensional Banach manifold. Apart from the geometrical insight one gained into the structure of this classical field theory, which is meant to help in the understanding of its quantization, the formalism has proved useful for other applications like the comparison of different classical field theories and for theoretical and numerical problems in the construction of solutions from initial data.[3,2,4]

The main purpose of this paper is to use this well-developed and elegant geometrical theory to study once again the Newtonian limit of general relativity. This subject has a long history[5–8] and, at least in what concerns the slow-motion or post-Newtonian approximation to general relativity, is still not yet fully understood, in spite of being of considerable importance in astrophysics and cosmology. In most recent formulations[9–13] divergencies show up sooner or later that are not understood and removed to everybody's satisfaction. Of less practical importance, though an aesthetical defect, is the fact that the post-Newtonian formalism considers the Newtonian theory as a first-order perturbation to Minkowski space, while on the other hand, it is well known from the beginnings of relativity theory that classical mechanics and Lorentz invariance are not compatible. We will not, in this paper, consider the problem of divergencies in the higher-order post-Newtonian approximations but describe the limiting process in a consistent fully geometric form in which such questions can at least be rigorously formulated. The Newtonian solutions are here taken as freely given and are not a first approximation to flat space. As usual they determine uniquely the first post-Newtonian approximation.

For this purpose it is necessary to describe the relativistic and Newtonian theory in the same language which must, because of the greater scope of general relativity, necessarily be the one of differential geometry. We use the slightly generalized geometric Newtonian theory of Refs. 14 and 15 which, however, agrees exactly with the classical Newtonian theory if the usual asymptotic boundary conditions are imposed. As pointed out in Ref. 15 this theory has a few intriguing aspects if it is considered as a classical field theory in its own right. It can be regarded as a mildly nonlinear gauge theory somewhat resembling the Maxwell–Einstein theory. It is nonlinear enough to determine the equations of motion of test particles,[15] but it does not seem to admit a (four-dimensionally-covariant) Lagrangian formulation. Part of our motivation for this paper was also to better

a) Partially supported by the Natural Sciences and Engineering Research Council of Canada, Grant no. A8059.
b) Permanent address.

understand this degenerate kind of canonical structure. We do find that this may be somewhat easier in terms of the ADM canonical formalism.

Since the gravitational degrees of freedom are completely frozen in the Newtonian limit one cannot gain much insight into the structure of this limiting process by considering a pure gravitational field. On the other hand, the perfect fluid model is clearly one of the simplest theories that are of some interest and physically well understood also in the Newtonian case. (Electromagnetism has different Galilean limits which are all limited in scope[16,7]; for nonrelativistic quantum mechanics considered as a "Newtonian" classical field theory cf. Refs. 17 and 18.) We therefore couple the gravitational field to a barotropic perfect fluid source for which we first develop the canonical formalism.

In fact, there are a variety of Lagrangian and Hamiltonian formulations available in the literature for perfect fluids in general relativity.[19–25] Schutz[22] obtained a Hamiltonian theory for relativistic perfect fluids based on his relativistic generalization of Seliger and Whitham's[26] variational principle. It has been elaborated and reduced by Demaret and Moncrief[25] while Moncrief[24] also applied ADM techniques to Taub's[20] variational principle. These authors, as well as others (Refs. 27–29 more recently for the nonrelativistic case), start with Euler's equations of motion and introduce velocity potentials, some of which they can choose to be quantities that are physically observable at least in special circumstances. This approach may be useful, for example, for basing variational techniques in a concrete, possible numerical, problem directly on a minimal number of variables that can be measured in specific situations.

Since here our purpose is not so much a specific Hamiltonian model for a concrete application in astrophysics, but rather the study of the symplectic structure of the field theory and its degenerate Newtonian limit we prefer the approach that is based essentially on the description in Lagrangian coordinates. It also has been around for a long time (see, for example, a text like Ref. 30) and has recently been studied again in the form of a standard classical field theory by Kijowski and Tulczyjew.[31] The fields are regarded as cross sections of a trivial bundle over space-time M with as standard fiber a material or body manifold B which can be equipped with some geometrical structure incorporating intrinsic properties of the material. In this sense it is a much more general approach than those based on a specific set of field equations for which a variational formulation is constructed more or less *ad hoc*. (See, for example, Noll.[32])

In the case of hydrodynamics it is sufficient to have a volume element given in terms of which the *matter* density ρ and the (conserved) flow vector $\mathbf{j} = \rho \mathbf{v}$ on M are defined by the field $x \mapsto b(x) \in B$. If then (in the relativistic case) the (rest frame) mass-energy density μ is chosen as the Lagrangian with the fields $x \mapsto b(x)$ as independent variables it turns out that the Euler–Lagrange equations for the $b(x)$ correspond to the momentum balance equations that can also be obtained from covariance considerations. (Kijowski and Tulczyjew[33] have extended this approach to include thermodynamic variables.)

The canonical analysis[2] can be applied to this theory in a quite straightforward way except for some technical difficulties in solving for the canonical momenta in terms of velocities that can only be explicitly obtained for the simplest equations of state. The resulting Hamilton equations contain more variables than are physically needed, but if the coordinate freedom in the material space is eliminated the relativistic versions of the conservation of matter and momentum equations (Euler equations) are obtained in a form which makes it very easy to compare them with the classical Newtonian equations. We carry out this program in Sec. 2, carefully choosing the geometrical and physical quantities such that they have regular limits if $c \to \infty$.

In Sec. 3 the geometric Newtonian theory is briefly summarized and then cast into a $(3 + 1)$-dimensional form. Since no Lagrangian is known for this theory a strictly canonical form cannot be constructed either, but it is very easy to introduce quantities that are analogous to the canonical variables of the relativistic theory. Especially when the 3-metric of the time slices is split into a conformal part and its determinant (a technique well known to be effective also for the solution of the relativistic initial value problem[4]), it is possible to formulate the relativistic and the Newtonian equations in a way such that the latter are manifestly the limit for $c \to \infty$ of the former. This can be achieved even without fixing the lapse and shift function although the Newtonian theory only takes on its classical form when suitable gauge conditions are imposed. We make this explicit in Sec. 4, where we also attempt to interpret these results in the global symplectic framework of Fischer and Marsden.[3]

In Sec. 5 we derive the first post-Newtonian approximation systematically within our geometrical framework and recover all the standard features, but without somewhat vague appeals to the equivalence principle and without having to impose coordinate or gauge conditions in advance.

Concerning the general mathematical level of this paper we believe that the formalism as presented here is suitable for the most rigorous analysis of the occurring partial differential equations in the framework of asymptotically weighted Sobolev spaces (as reviewed, for example, in Ref. 34). We have not attempted here, however, to spell out all the exact differentiability and asymptotic falloff properties of the quantities involved. Instead we assume, as is standard practice in work on post-Newtonian approximations, that the function spaces are such that the Laplacian and other homogeneous elliptic operators with constant coefficients all have inverses.

When this work was essentially completed the authors became aware of a paper by Weinstein[35] where the notion of a Poisson manifold is applied to the study of some symplectic manifolds in mechanics that change type when certain parameters tend to limit values. It is likely that the situation encountered in Sec. 4 could also be better understood in terms of such (infinite-dimensional) Poisson structures.

2. CANONICAL FORMALISM FOR RELATIVISTIC PERFECT FLUIDS

The canonical Hamiltonian formalism for classical fields interacting with gravity has shown itself to be a practical way of investigating the nature of a given classical field theory. We set up this formalism for barotropic relativistic perfect fluids mainly following the review,[2] except for making the dependence on the velocity of light c explicit. We slightly depart from the usual definitions[2,3] occasionally in order to have regular Newtonian limits for all basic geometric and physical quantities.

(3 + 1)-dimensional formalism

Space-time is foliated by spacelike hypersurfaces which we take to be the level surfaces Σ_t of a function t. We will always use frame fields e_α (and their duals θ^α) adapted to these surfaces such that by the foliation condition $\theta^0 = N\,dt$ for some function N.[36] Then the space-time metric is

$$\gamma = -c^2\theta^0 \otimes \theta^0 + \delta_{\mathbf{ab}}\theta^{\mathbf{a}} \otimes \theta^{\mathbf{b}}, \tag{2.1}$$

and its inverse

$$\gamma^{-1} = -c^{-2}e_0 \otimes e_0 + \delta^{\mathbf{ab}}e_{\mathbf{a}} \otimes e_{\mathbf{b}}.^{37} \tag{2.2}$$

These adapted frames are used to decompose space-time covariant objects into their spatially covariant parts. In particular (2.1) induces a metric

$$\gamma = \gamma_{ab}\, dx^a \otimes dx^b \quad \text{and} \quad \gamma^{-1} = \gamma^{ab}\partial_a \otimes \partial_b \tag{2.3}$$

on the space slices. This spatial metric is the dynamical field of Einstein gravity. In the canonical Hamiltonian formalism one passes from a description in terms of these spatial fields and their "velocities" to one in terms of these fields and their conjugate momenta. The field equations then split up into two types: Those which are constraints on the initial values of the fields and momenta and those which give the time derivatives for these quantities. These derivatives are fixed by

$$\partial_t = Ne_0 + X \tag{2.4}$$

in terms of $X = X^{\mathbf{a}}e_{\mathbf{a}}$ (the "shift") along the spatial surface plus an amount N (the "lapse") along the normal leading to the "next" Σ_t surface.

We describe a barotropic perfect fluid by maps

$$\beta: M \mapsto B: x \to (b^A(x)) \tag{2.5}$$

from the space-time M into a three-dimensional *material space* (or body manifold) B. On B there is given a volume element $r = m\, d^3b$ which like the scalar density m on B has the physical dimension of mass. Then the pull back $\beta^*(r)$ of this volume element determines a conserved 4-vector (the *matter-flow vector*) $\mathbf{j}$, given in coordinates by

$$\mathbf{j}^\alpha = m\boldsymbol{\epsilon}^{\alpha\beta\gamma\delta}b^1_\beta\, b^2_\gamma\, b^3_\delta, \tag{2.6}$$

where $b^A_a = \partial_\alpha b^A$ and the tensor $\boldsymbol{\epsilon}^{\alpha\beta\gamma\delta}$ is given in terms of the Lorentz frame e_α by $\boldsymbol{\epsilon}^{\alpha\beta\gamma\delta} = e_0^{[\alpha}\, e_1^\beta\, e_2^\gamma\, e_3^{\delta]}$. Assume that $\mathbf{j}$ is a future pointing timelike vector and define the matter density ρ by

$$\mathbf{j} = \rho\mathbf{v}, \tag{2.7}$$

where

$$\gamma(\mathbf{v},\mathbf{v}) = -c^2. \tag{2.8}$$

Equation (2.8) and

$$\mathbf{v}^\alpha b^A_\alpha = 0 \tag{2.9}$$

characterize $\mathbf{v}$.

Hamiltonian form of the field equations

The field equations for the gravitational field and the fluid can be obtained from the Lagrangian density (with units of action)

$$\mathscr{L} = (k^{-1}c^4\mathbf{R} - \mu)\upsilon, \tag{2.10}$$

where $\upsilon = c^{-1}\sqrt{-\det(\gamma_{\alpha\beta})}$, $k = 16\pi G$, G is Newton's gravitational constant, $\mu = \rho(c^2 + \epsilon(\rho))$ is the mass-energy density per volume, ϵ the specific internal energy, and ρ is considered as a function of b^A and b^A_α.

To pass to the canonical Hamiltonian form $\mathscr{L}$ must be reexpressed in terms of three-dimensionally-covariant quantities and their time derivatives with respect to ∂_t. We have

$$\upsilon = N\upsilon, \quad \upsilon := \sqrt{\det(\gamma_{ab})} \tag{2.11}$$

and

$$\mathbf{R} = R + (c^{-1}K^{ab})(c^{-1}K_{ab}) - (c^{-1}K^r_r)^2 + \text{divergence terms}, \tag{2.12}$$

where the extrinsic curvature (second fundamental form) of Σ_t is

$$c^{-1}K_{ab} := -(2Nc)^{-1}(\dot\gamma_{ab} - £_X\gamma_{ab}) \text{ (Ref. 38)} \tag{2.13}$$

($\dot{} = \partial_t$ and $£_X$ denotes the Lie derivative in Σ_t). (See, e.g., Refs. 1 and 39, Chap. 21 or 2.)

It is convenient to split $\mathbf{v}$ in the form

$$\mathbf{v} = v^0(e_0 + v). \tag{2.14}$$

Then, by (2.6),

$$\mathbf{j}^0 = m \det(b^A_{\mathbf{c}}) = m\upsilon^{-1}\det(b^A_c) \tag{2.15}$$

depends only on the *configuration* of the material on a surface Σ_t, while according to (2.9) the velocities appear via

$$N^{-1}(\dot b^A - £_X\, b^A) = b^A_0 = -b^A_{\mathbf{c}}\, v^{\mathbf{c}} = -b^A_i\, v^i. \tag{2.16}$$

Therefore

$$\rho = \mathbf{j}^0(1 - |v|^2/c^2)^{1/2} =: \mathbf{j}^0 V. \tag{2.17}$$

Varying the Lagrangian (2.10) with respect to $\dot\gamma_{ab}$ and $\dot b^A$ leads to the canonical momenta

$$k^{-1}c^2\pi^{ab} := \frac{\partial L}{\partial\dot\gamma_{ab}} = k^{-1}c^2\upsilon(\gamma^{ab}K^r_r - K^{ab}) \tag{2.18}$$

and

$$P_A := \frac{\partial L}{\partial\dot b^A} = -c^{-2}\upsilon\rho\frac{d\mu}{d\rho}V^{-2}b^{-1i}_{\ A}v_i. \tag{2.19}$$

In order to find the Hamiltonian one must solve these equations for the velocities. This is easily done for $\dot\gamma_{ab}$ in terms of π^{ab} and leads to Eqs. (2.31) below. To solve (2.19) for $\dot b^A$ we first define

$$\Delta^{AB} := b^A_a\gamma^{ab}b^B_b. \tag{2.20}$$

Then, from (2.19) we have

$$|P| := (\Delta^{AB}P_A\, P_B)^{1/2} = c^{-2}\upsilon V^{-2}\rho\frac{d\mu}{d\rho}. \tag{2.21}$$

We will now assume that the in general transcendental equation (2.21) is solved for $|v|$ in terms of the scalar $\mathscr{v}^{-1}|P|$ by

$$|v| = (\mathscr{v}\mathbf{j}^0)^{-1} f(\mathbf{j}^0, |P|\mathscr{v}^{-1})|P|, \tag{2.22}$$

so that

$$\dot{b}^A - \pounds_X b^A = \mathscr{v}^{-1} N(\mathbf{j}^0)^{-1} f \Delta^{AB} P_B \quad (= -N b^A_i v^i). \tag{2.23}$$

For the simplest case (dust) we have $\mu = \rho c^2$ and we can solve (2.21) explicitly:

$$f = [1 + |P|^2 (c\mathscr{v}\mathbf{j}^0)^{-2}]^{-1/2} \quad (= V). \tag{2.24}$$

Now that we have the velocities in terms of the momenta we can recast the action in canonical Hamiltonian form

$$k^{-1}c^2 \int dt \int d^3x (\pi^{ab}\dot{\gamma}_{ab} + kc^{-2} P_A \dot{b}^A - c^2 N \mathscr{H} - X^a \mathscr{J}_a), \tag{2.25}$$

where the super-Hamiltonian is

$$\mathscr{H} = \mathscr{H}_{\mathrm{gr}} + \mathscr{H}_{\mathrm{fl}} \tag{2.26}$$

with

$$\mathscr{H}_{\mathrm{gr}} = -\mathscr{v} R + c^{-2}\mathscr{v}^{-1}(\pi^{ab}\pi_{ab} - \tfrac{1}{2}(\pi^r_r)^2) \tag{2.27}$$

and

$$\begin{aligned}\mathscr{H}_{\mathrm{fl}} &= \mathscr{H}_{\mathrm{fl}}(b^A, b^A_c, \gamma_{ab}, P_A)\\ &= kc^{-4}(\mathscr{v}\mu + \mathscr{v}^{-1} f |P|^2/\mathbf{j}^0)\\ &(= kc^{-4}\mathscr{v}\rho(c^2 + \epsilon + f^{-1}V^{-1}|v|^2))\end{aligned} \tag{2.28}$$

(which becomes in the case of dust

$$\mathscr{H}_{\mathrm{dust}} = k\mathbf{j}^0[1 + |P|^2/(c\mathscr{v}\mathbf{j}^0)]^{1/2}). \tag{2.29}$$

The supermomentum is

$$\mathscr{J}_a = \mathscr{J}^{\mathrm{gr}}_a + \mathscr{J}^{\mathrm{fl}}_a = -2\nabla_r \pi^r_a + kc^{-2} b^A_a P_A. \tag{2.30}$$

The Hamiltonian equations of evolution are now found to be

$$\dot{\gamma}_{ab} - \pounds_X \gamma_{ab} = \mathscr{v}^{-1} N (2\pi_{ab} - \pi^r_r \gamma_{ab}), \tag{2.31}$$

$$\dot{b}^A - \pounds_X b^A = \mathscr{v}^{-1} N f/\mathbf{j}^0 \Delta^{AB} P_B = -N b^A_i v^i, \tag{2.32}$$

$$\begin{aligned}c^{-2}(\dot{\pi}^{ab} - \pounds_X \pi^{ab}) &= -\mathscr{v}[N(R^{ab} - \tfrac{1}{2}R\gamma^{ab}) - \nabla^a\nabla^b N + \nabla^2 N \gamma^{ab}]\\ &\quad + c^{-2}\mathscr{v}^{-1} N [\pi^{ab}\pi^c_c - 2\pi^a_c \pi^{cb}\\ &\quad + \tfrac{1}{2}\gamma^{ab}(\pi^{rs}\pi_{rs} - \tfrac{1}{2}(\pi^r_r)^2)]\\ &\quad + \tfrac{1}{2}kc^{-4} N \mathscr{v} [p\gamma^{ab} + f^{-1}V^{-1}\rho v^a v^b],\end{aligned} \tag{2.33}$$

$$\begin{aligned}\dot{P}^A - \pounds_X P^A &= -N\frac{\partial \mathscr{H}_{\mathrm{fl}}}{\partial b^A} + \nabla_a\left(N \frac{\partial \mathscr{H}_{\mathrm{fl}}}{\partial b^A_a}\right)\\ &= c^2\mathscr{v}[-NVf^{-1}\rho m^{-1}\partial_A m + \nabla_i(NVf^{-1}\rho(\delta^i_j\\ &\quad + c^{-2}V^{-2}v^i v_j) b^{-1j}{}_A)].\end{aligned} \tag{2.34}$$

In these equations the velocity variables v^i have been introduced on the right-hand side because the expressions then look simpler. But the v^i, V, ρ, f must be regarded as functions of P_A via Eqs. (2.19) and (2.21). The variation of $\mathscr{H}_{\mathrm{fl}}$ amounts to a very long, but straightforward, calculation in which f must be correctly treated as an only implicitly defined function.

In addition to the evolution equations (2.31)–(2.34) one has the initial value constraints

$$\mathscr{H} = 0, \quad \mathscr{J}_a = 0 \tag{2.35}$$

corresponding to $2k^{-1}c^4\mathbf{G}^0_0 - \mathbf{T}^0_0 = 0$, and to $2k^{-1}c^4\mathbf{G}^0_{\mathbf{a}} - \mathbf{T}^0_{\mathbf{a}} = 0$, respectively. Equation (2.33), on the other hand, corresponds to the remaining Einstein equations

$$\mathbf{G}^{ab} = 8\pi G c^{-4}\mathbf{T}^{ab}. \tag{2.36}$$

Incidentally, the preservation of the constraints (2.35) corresponds to

$$c^4/(8\pi G)\boldsymbol{\nabla}_\lambda \mathbf{G}^\lambda_\alpha - \boldsymbol{\nabla}_\lambda \mathbf{T}^\lambda_\alpha = 0. \tag{2.37}$$

Coordinate freedom

Equations (2.32) and (2.34) together represent the equations of motion of the fluid (Euler's equations). They look different from the standard form mainly because we label the fluid elements by arbitrary coordinates in a material space rather than in terms of space-time coordinates and velocities. Because we have no constraints on our fluid variables we do not need the Dirac constraint formalism. The only constraints in our theory are the usual super-Hamiltonian and supermomentum constraints. We have, however, up to this point kept full flexibility in the choice of lapse and shift, space and material coordinates.

As far as the freedom in material coordinates is concerned it is easy to see that for given ρ and $\mathbf{v}$ on space-time the corresponding fields $T^*\beta: M \ni x \mapsto (b^A(x), \mathscr{v}^{-1}P_A(x)) \in T^*B$ are given up to a global diffeomorphism Φ of B. One can instead consider the matter field as a cross section of a trivial bundle $M \times B$ as Kijowski and Tulczyjew[31] do, but the Φ cannot be a local (i.e., x-dependent) gauge transformation.

If we eliminate this coordinate freedom we recover the relativistic version of the Euler equations (i.e., the equations $\boldsymbol{\nabla}_\lambda \mathbf{T}^\lambda_\alpha = 0$). First we rewrite (2.34) in the form

$$\begin{aligned}\dot{P}_A - \pounds_X P_A &= \mathscr{v} b^{-1i}{}_A [\nabla_j(NV^{-1}f^{-1}\rho v^j v_i)\\ &\quad - NV^{-1}f^{-1}\rho b^B_{ij} v^j b^{-1k}{}_B v_k\\ &\quad + \rho f^{-1}V\partial_i\Phi + N\partial_i p\\ &\quad - \tfrac{1}{2}\rho N V^{-1} f^{-1}\partial_i |v|^2],\end{aligned} \tag{2.38}$$

where we have defined

$$\Phi := c^2(N-1) \tag{2.39}$$

and

$$b^A_{ij} := \nabla_j b^A_i = \partial_j b^A_i - \Gamma^r_{ji} b^A_r \tag{2.40}$$

and where we used that (2.22) implies

$$f^{-1}V = c^{-2}\frac{d\mu}{d\rho} = 1 + c^{-2}\left(\epsilon + \frac{p}{\rho}\right) \tag{2.41}$$

(if we consider the isentropic case only where $p = \rho^2 d\epsilon/d\rho$), and that

$$m^{-1}\partial_A m = b^{-1i}{}_A \nabla_i(\rho V^{-1})V\rho^{-1} - b^{-1i}{}_A b^B_{ij} b^{-1j}{}_B. \tag{2.42}$$

The *conservation of matter equation* $\boldsymbol{\nabla}_\alpha \mathbf{j}^\alpha = 0$, becomes, in $(3+1)$-dimensional form,

$$\begin{aligned}&(V^{-1}\rho)^{\cdot} - \pounds_X(V^{-1}\rho) + \nabla_i(NV^{-1}\rho v^i)\\ &\quad - NV^{-1}\rho K^r_r - 2N^{-1}V^{-1}\rho\pounds_X N = 0.\end{aligned} \tag{2.43}$$

With its help and (2.19), Eq. (2.38) can finally be written in the form

$$\rho[(f^{-1}v_i)^{\cdot} - \pounds_X(f^{-1}v_i) + Nv^k\nabla_k(f^{-1}v_i) + f^{-1}\partial_i\Phi\,] + NV\partial_i p + 2\rho f^{-1}N^{-1}\pounds_X Nv_i = 0, \tag{2.44}$$

which are the *relativistic Euler equations.*

Gauge conditions

For specific applications one may wish to fix the lapse and shift function by suitable conditions. A good way to fix the lapse (especially for numerical integration) is the *maximal slicing condition* (see York[4]) which requires the trace of the second fundamental form, or equivalently of π^{ab}, to vanish for all time slices. The trace of (2.33) and (2.35) imply that this condition takes the form

$$\mathscr{T} := \pi^r_r = 0 \tag{2.45}$$

and

$$\mathscr{M} := \nabla^2\Phi - 4\pi GN\rho - \mathscr{v}^{-2}N\pi^{ij}\pi_{ij} - \tfrac{1}{4}kc^{-2}N(\rho\epsilon + 3p + 2\rho f^{-1}V^{-1}|v|^2) = 0. \tag{2.46}$$

A natural way to fix the shift vector in hydrodynamics is to use *comoving coordinates*, i.e., to require that $\dot b^A = 0$ (or equivalently that $v = 0$). This does simplify the equations, but the resulting system cannot normally be solved uniquely for X. A mathematically and for numerical calculations more satisfying method is York's[4] *minimal distortion condition*. It requires that the time dependence of the conformal metric $\bar\gamma^{ij}$ [see (2.48) below] or, more precisely, the quantity

$$\int (\dot{\bar\gamma}^{ij})^2\mathscr{v}^{-1/3}d^3x = \int (\dot\gamma_{ij} - \tfrac{1}{3}\gamma_{ij}\gamma^{rs}\dot\gamma_{rs})^2\mathscr{v}\,d^3x$$

regarded as a functional of X be minimized. York shows that, independent of the source terms, this leads to the elliptic equation

$$\nabla^r\nabla_r X^i + \tfrac{1}{3}\nabla^i\nabla_r X^r + R_{ir}X^r = -2\mathscr{v}^{-1}\nabla_r[N(\pi^r_i - \tfrac{1}{3}\pi^s_s\delta^r_i)] \tag{2.47}$$

for X which under suitable boundary conditions will have a unique solution.

New variables

We have been careful to write all equations in this section in a form which will facilitate taking the limit $c \to \infty$ since the factors c have already been placed in suitable positions. It will turn out, however, in the next section that not all of the evolution equations (2.33) will become constraints in the limit. Roughly speaking, the trace remains an evolution equation. It is therefore simpler to take the limit if we introduce new variables, splitting the metric into a conformal part and its determinant. Of course, this approach is closely related to York's[4] conformal method for treating the initial value problem.

Choose therefore $\mathscr{v}$ and the tensor density of weight $\tfrac{2}{3}$,

$$\bar\gamma^{ab} := \mathscr{v}^{2/3}\gamma^{ab}, \tag{2.48}$$

as gravitational configuration space variables, and as momenta the scalar

$$\bar\pi := \tfrac{2}{3}\mathscr{v}^{-1}\pi^r_r \tag{2.49}$$

and the tensor density of weight $\tfrac{1}{3}$,

$$\bar\pi_{ab} := \mathscr{v}^{-2/3}(\pi_{ab} - \tfrac{1}{3}\pi^r_r\gamma_{ab}). \tag{2.50}$$

The action then becomes, instead of (2.25),

$$k^{-1}c^2\int dt\int d^3x(\bar\pi_{ab}\dot{\bar\gamma}^{ab} + \bar\pi\mathscr{v} + kc^{-2}P_A\dot b^A - c^2N\mathscr{H} - X^a\mathscr{J}_a) \tag{2.51}$$

and leads to the evolution equations

$$\partial_t\bar\gamma^{ab} - \pounds_X\bar\gamma^{ab} = -2\mathscr{v}^{1/3}N\bar\pi^{ab}, \tag{2.52}$$

$$\partial_t\mathscr{v} - \pounds_X\mathscr{v} = -\tfrac{3}{4}\mathscr{v}N\bar\pi, \tag{2.53}$$

$$c^{-2}\mathscr{v}^{-1/3}(\partial_t\bar\pi_{ab} - \pounds_X\bar\pi_{ab} - 2\mathscr{v}^{-1/3}N\bar\pi^r_a\bar\pi_{rb}) = -N(R_{ab} - \tfrac{1}{3}R\gamma_{ab}) + c^{-2}(\nabla_a\nabla_b\Phi - \tfrac{1}{3}\nabla^2\Phi\gamma_{ab}) + \tfrac{1}{2}kc^{-4}Nf^{-1}V^{-1}\rho(v_av_b - \tfrac{1}{3}|v|^2\gamma_{ab}), \tag{2.54}$$

$$\partial_t\bar\pi - \pounds_X\bar\pi = -\tfrac{4}{3}\nabla^2\Phi + N[\tfrac{1}{3}c^2R + \mathscr{v}^{-2/3}\bar\pi_{rs}\bar\pi^{rs} + \tfrac{3}{8}\bar\pi^2 + kc^{-2}(p + \tfrac{1}{3}\rho f^{-1}V^{-1}|v|^2)], \tag{2.55'}$$

together with (2.32) and (2.34), and to the constraints

$$\mathscr{H} = -\mathscr{v}R + c^{-2}\mathscr{v}(\mathscr{v}^{-2/3}\bar\pi^{rs}\bar\pi_{rs} - \tfrac{3}{8}\bar\pi^2) + kc^{-4}\mathscr{v}(\mu + \rho V^{-1}f^{-1}|v|^2) = 0, \tag{2.56}$$

$$\mathscr{J}_a = -2\mathscr{v}^{-2/3}\nabla^r\bar\pi_{ra} - \mathscr{v}\partial_a\bar\pi - kc^{-2}f^{-1}V^{-1}v_i = 0. \tag{2.57}$$

(Indices are still raised with γ^{ab}.)

This is the new canonical form of the system of equations. If we add $\tfrac{1}{3}c^2\mathscr{v}^{-1}N\mathscr{H}$ to the right-hand side of Eq. (2.55') we obtain an evolution equation

$$\partial_t\bar\pi - \pounds_X\bar\pi = -\tfrac{4}{3}(\nabla^2\Phi - 4\pi G\rho N) + \tfrac{4}{3}N\mathscr{v}^{-2/3}\bar\pi^{rs}\bar\pi_{rs} + \tfrac{1}{4}N\bar\pi^2 + \tfrac{1}{3}kc^{-2}N(\rho\epsilon + 3p + 2\rho V^{-1}f^{-1}|v|^2) \tag{2.55}$$

which is equivalent to (2.55') *on the constraint set* and has a regular limit for $c \to \infty$.

3. THE GEOMETRIC NEWTONIAN THEORY

Classical hydrodynamics

In its classical form a gravitating barotropic perfect fluid is described by the system of equations

$$\partial_t\rho + \partial_i(\rho v^i) = 0, \tag{3.1}$$

$$\rho(\partial_t v_i + v^r\nabla_r v_i) + \partial_i p + \rho\partial_i\Phi = 0, \tag{3.2}$$

$$\nabla^r\nabla_r\Phi = 4\pi G\rho, \tag{3.3}$$

where ∇ is with respect to a flat 3-metric γ_{ij} which may be written in arbitrary coordinates but so that $\partial_t\gamma_{ij} = 0$. If the equation of state $p = p(\rho)$ is fixed and ρ and v are given on an initial time slice and the boundary conditions are such that the Laplace operator has an inverse, then the time evolution of the system is determined.

These equations can be derived from a (degenerate) Lagrangian density

$$\mathscr{L} = \rho[\tfrac{1}{2}|v|^2 - \epsilon(\rho) - \Phi\,] - (8\pi G)^{-1}\partial^r\Phi\partial_r\Phi \tag{3.4}$$

if one takes $b^A(t, x)$ and Φ as dynamical variables and defines ρ and v, similarly as in Sec. 2, by $\rho = m\det(b^A_i)$ and $\partial_t b^A + b^A_r v^r = 0$.

Since the metric and the coordinate system are fixed, however, one cannot compare this theory with the relativistic version without introducing there somewhat arbitrary coordinate (gauge) conditions. Our aim is to see explicitly how the general canonical or symplectic structure of the relativistic theory degenerates in the Newtonian limit and therefore we must *a priori* admit the metric as a dynamical variable in the Newtonian theory. In fact, the system (3.1)–(3.3) is not even invariant under time-dependent Euclidean transformations since the potential Φ does not transform as a scalar in that case. It is only correct in an "Aristotelian" space-time.[40]

Newtonian manifolds

We recall the main definitions of the geometric Newtonian gravitation theory as formulated in Refs. 14 and 15 where also further references to its long history can be found.[41]

A *Galilei manifold* is a space-time M with an equivalence class of frames singled out at each point that are related by the (homogeneous) Galilei group rather than the Lorentz group as for a Lorentz manifold. This class of frames $\{e_\alpha\}$ is determined by a contravariant symmetric positive semidefinite tensor field γ of rank 3 (the *space metric*) and a rank 1 tensor field $\mathbf{g}$ (the *time metric*) such that

$$\gamma^{ab}\mathbf{g}_{\beta\gamma} = 0 \tag{3.5}$$

and

$$\gamma = \delta^{\mathbf{ab}} e_{\mathbf{a}} \otimes e_{\mathbf{b}} \quad \text{and} \quad \mathbf{g} = \theta^{\mathbf{0}} \otimes \theta^{\mathbf{0}}. \tag{3.6}$$

We assume that the space-time M is time and space orientable. Then $\theta^{\mathbf{0}}$ is a Galilei invariant which we also call ψ. Also a volume element and thus the tensors $\epsilon^{\alpha\beta\gamma\delta}$ and $\epsilon_{\alpha\beta\gamma\delta}$ are still well defined. A vector $\mathbf{X}$ is called spacelike if $\psi(\mathbf{X}) = 0$, unit timelike if $\psi(\mathbf{X}) = 1$.

A Galilei manifold admits a compatible symmetric linear connection ∇ (i.e., $\nabla\gamma = 0$ and $\nabla\psi = 0$) if ψ is closed. Then, locally, $\psi = \mathbf{dt}$ and there exists a canonical foliation into spacelike hypersurfaces. Such a torsion-free Galilei connection is not fully determined by γ and ψ but is given by

$$\Gamma^\alpha_{\beta\gamma} = \overset{u}{\Gamma}{}^\alpha_{\beta\gamma} + \psi_{(\beta}\,\mathbf{F}_{\gamma)\lambda}\,\gamma^{\alpha\lambda} \tag{3.7}$$

in terms of a unit timelike vector field $\mathbf{u}$ and a 2-form $\mathbf{F}$ where $\overset{u}{\nabla}$ is the unique symmetric Galilei connection such that $\mathbf{u}$ is geodesic and irrotational, i.e., such that

$$\mathbf{u}^\lambda \overset{u}{\nabla}_\lambda \mathbf{u}^\alpha = 0 \quad \text{and} \quad \gamma^{\lambda[\alpha}\overset{u}{\nabla}_\lambda \mathbf{u}^{\beta]} = 0. \tag{3.8}$$

The curvature tensor of any symmetric Galilei connection satisfies

$$\gamma^{\lambda(\alpha}\mathbf{R}^{\beta)}{}_{\lambda\mu\nu} = 0 \quad \text{and} \quad \psi_\lambda\,\mathbf{R}^\lambda{}_{\alpha\beta\gamma} = 0. \tag{3.9}$$

If, moreover,

$$\mathbf{R}^\alpha{}_{\beta\lambda\delta}\gamma^{\gamma\lambda} = \mathbf{R}^\gamma{}_{\delta\lambda\beta}\,\gamma^{\alpha\lambda}, \tag{3.10}$$

the connection is called *Newtonian*. The condition (3.10) is equivalent to $\mathbf{dF} = 0$ in (3.7), and it is not hard to show[7] that if a family of symmetric Lorentz connections has a Galilean limit then that limit is Newtonian. We will only consider Newtonian structures from now on. They can be described by (γ,ψ,Γ) together with the constraints (3.10) and

$$\nabla\gamma = 0, \quad \nabla\psi = 0, \tag{3.11}$$

or locally, via (3.7) by $(\gamma,\psi,\mathbf{u},\mathbf{A})$ where $\mathbf{F} = \mathbf{dA}$, or even, when convenient, by $(\gamma,\psi,\mathbf{u},\mathbf{A},\Gamma)$ subject to[15]

$$\mathbf{F}_{\alpha\beta} + 2\overset{u}{\gamma}_{\lambda[\alpha}\nabla_{\beta]}\,\mathbf{u}^\lambda = 0, \tag{3.12}$$

where $\overset{u}{\gamma}$ is defined by

$$\overset{u}{\gamma}_{\alpha\lambda}\,\mathbf{u}^\lambda = 0 \quad \text{and} \quad \overset{u}{\gamma}_{\alpha\lambda}\gamma^{\lambda\beta} = \delta^\beta_\alpha - \psi_\alpha\mathbf{u}^\beta. \tag{3.13}$$

Clearly, these latter two descriptions of the Newtonian space-time contain more information than needed. Both $\mathbf{u}$ and $\mathbf{A}$ are some sort of gauge fields; either of the two could be fixed, though not in a canonical way. On the other hand, it seems quite hopeless to achieve a Lagrangian or Hamiltonian formalism in terms of γ, ψ, and Γ alone. When we describe the geometry by $(\gamma,\psi,\mathbf{u},\mathbf{A})$, however, we have—in addition to the invariance of the formalism under space-time diffeomorphisms—the gauge transformations[18]

$$\begin{aligned}\mathbf{u}^\alpha &\mapsto \mathbf{u}^\alpha + \gamma^{\alpha\lambda}\mathbf{w}_\lambda,\\ \mathbf{A}_\alpha &\mapsto \mathbf{A}_\alpha + \partial_\alpha\,\chi + \mathbf{w}_\alpha - \tfrac{1}{2}\,\psi_\alpha(\gamma^{\lambda\mu}\mathbf{w}_\lambda\mathbf{w}_\mu + 2\mathbf{u}^\lambda\mathbf{w}_\lambda)\end{aligned} \tag{3.14}$$

which are parametrized by a scalar χ and a 1-form $\mathbf{w} = \mathbf{w}_\lambda\,dx^\lambda$. (There is a connection between this gauge group and the extended Galilei (or Bargmann) bundle.[42,18])

In Ref. 14 it was shown that the field equations

$$\mathbf{R}_{\alpha\beta} = 4\pi G\rho\mathbf{g}_{\alpha\beta} \tag{3.15}$$

imposed on a Newtonian manifold lead to a theory that is only slightly more general than classical Newtonian gravitation and agrees exactly with the latter if suitable global conditions are added.

In contradistinction to Einstein's theory the field equations (3.15) do not imply the conservation equations for matter. On the other hand, it was shown in Ref. 15 that the conservation equations of matter can be obtained essentially by Noether's theorem if the matter Lagrangian $\mathscr{L}_m$ depends on the gravitational field $(\gamma,\psi,\mathbf{u},\mathbf{A})$ and is invariant (possibly up to a divergence) under the gauge transformations (3.14). As in particle mechanics the mass density ρ, or here better the matter flow vector $\mathbf{j}$, is independent of the stress energy tensor $\mathbf{T}^\alpha{}_\beta$ so that the conservation equations are

$$\nabla_\lambda\,\mathbf{j}^\lambda = 0 \tag{3.16}$$

and

$$\nabla_\lambda\mathbf{T}^\lambda{}_\alpha = \rho\overset{v}{\gamma}_{\alpha\lambda}\mathbf{v}^\mu\nabla_\mu\mathbf{v}^\lambda, \tag{3.17}$$

where $\mathbf{j} = \rho\mathbf{v}$, $\psi(\mathbf{v}) = 1$, and $\mathbf{T}^{[\alpha}{}_\lambda\gamma^{\beta]\lambda} = 0$. Equation (3.16) is the matter conservation (continuity) equation, (3.17) comprises energy and momentum conservation.[7]

For a perfect fluid

$$\mathbf{T}^\alpha{}_\beta = \epsilon\mathbf{j}^\alpha\psi_\beta - p(\delta^\alpha_\beta - \mathbf{v}^\alpha\psi_\beta), \tag{3.18}$$

where we consider the internal energy per unit mass ϵ and the pressure p as functions of ρ, related by $p = \rho^2 d\epsilon/d\rho$.

As mentioned already we can consider Eqs. (3.17) for the perfect fluid also as the field equations for the variables $b^A(x)$ if $\mathbf{j}$ is of the form (2.6).[43]

In fact, the matter part of the Lagrangian (3.4) can be generalized to

$$\mathscr{L}_m = (\tfrac{1}{2}\rho \overset{u}{\boldsymbol{\gamma}}_{\alpha\beta}\mathbf{v}^\alpha\mathbf{v}^\beta + \mathbf{A}_\alpha \mathbf{j}^\alpha - \rho\epsilon)\omega, \tag{3.19}$$

where $\mathbf{j}, \rho$, and $\mathbf{v}$ are now defined by (2.6) and (2.7) in terms of the fields $b^A(x)$ and $\omega\, \mathbf{d}^4\mathbf{x}$ is the Galilean space-time volume element. The Lagrangian $\mathscr{L}_m$ is invariant under the gauge transformations (3.14) up to a divergence and it can be shown that the Euler–Lagrange equations (with respect to the variables b^A) are equivalent to (3.17) with $\mathbf{T}^\alpha{}_\beta$ defined by (3.18). No field Lagrangian is known, however, that also yields Eqs. (3.15).

(3 + 1)-dimensional formalism

Our next step is to cast this geometrical theory again into a (3 + 1)-dimensional form in a way as similar as possible to the relativistic ADM formalism. However, since a Newtonian manifold has a canonical foliation into spacelike hypersurfaces $\iota_t : \Sigma \to M$ defined by $\iota_t{}^*\psi = 0$ we have hardly any choice but to adopt this time slicing. In fact, if there is to be a field of Galilei frames $\{\mathbf{e}_\alpha\}$ adapted to this foliation as the field of Lorentz frames in Sec. 2, then this foliation must be the canonical one.

Using the natural adapted local coordinates (t, x^a) such that $\boldsymbol{\theta}^0 = dt$ we have fixed the lapse function N to be the constant 1 while $\gamma = \delta^{ab} e_a \otimes e_b$ restricts to the inverse metric γ^{-1} of the slices. But we can still introduce an arbitrary (kinematical) shift vector Y by choosing, like in (2.4),

$$\partial_t = e_0 + Y = e_0 + Y^i\partial_i. \tag{3.20}$$

Without loss of generality we can now identify e_0 with the vector $\mathbf{u}$ and calculate the general Newtonian connection coefficients in this coordinate system according to (3.7) where we let $\mathbf{F} = \mathbf{dA}$. This gives[14]

$$\boldsymbol{\Gamma}^0_{\alpha\beta} = 0, \quad \boldsymbol{\Gamma}^a_{00} = -\nabla^a(Y^2) + \gamma^{ab}[\partial_t(Y_b + A_b) - \partial_b \mathbf{A}_0], \tag{3.21}$$

$$\boldsymbol{\Gamma}^a_{0b} = \gamma^{ar}[\partial_{[b}(A_{r]} + Y_{r]}) + \tfrac{1}{2}\dot{\gamma}_{br}], \quad \boldsymbol{\Gamma}^a_{bc} = \Gamma^a_{bc}.$$

We observe that it will be convenient to introduce a *dynamical shift vector*

$$X := (Y^a + \gamma^{ar}A_r)\partial_a. \tag{3.22}$$

If we also introduce, in analogy with (2.13),

$$K_{ab} := -\tfrac{1}{2}(\dot{\gamma}_{ab} - \pounds_X \gamma_{ab}), \tag{3.23}$$

then the gravitational field equations (3.15) become

$$R_{ab}(\gamma) = 0, \tag{3.24}$$

$$\nabla_r K^r_a - \nabla_a K^r_r = 0, \tag{3.25}$$

$$\nabla^2\Phi - 4\pi G\rho + \partial_t K^r_r - \pounds_X K^r_r - K^{rs}K_{rs} = 0, \tag{3.26}$$

where

$$\Phi := -A_0 + X^r A_r - \tfrac{1}{2}A^r A_r. \tag{3.27}$$

If we split the 4-velocity as in (2.14), but using the dynamical shift vector

$$\mathbf{v} = \partial_t + (v^a - X^a)\partial_a, \tag{3.28}$$

then it is found by a straightforward calculation that (3.16) and (3.17) become

$$\dot{\rho} - \pounds_X \rho + \nabla_r(\rho v^r) - K^r_r \rho = 0 \tag{3.29}$$

and

$$\rho(\dot{v}_a - \pounds_X v_a + v^r\nabla_r v_a) + \partial_a p + \rho\partial_a\Phi = 0, \tag{3.30}$$

respectively.

It is possible to replace (3.29) and (3.30) by Hamiltonian equations for $\dot{b}^A$ and $\dot{P}_A$ similar to (2.32) and (2.34) which can be derived from the matter Lagrangian (3.19).

It is already clear that the tensor K_{ab} takes the role of a second fundamental form of the hypersurfaces (although the standard definition of the latter is not available in the Newtonian case since no surface normal is defined). While the lapse function of the relativistic theory is trivial here its dynamical role is taken over by the quantity Φ which will be seen to coincide in the limit $c \to \infty$ with the relativistic Φ defined in (2.39). Finally it is clear that the relativistic shift vector must correspond to the dynamical shift vector of the Newtonian theory.

New variables and the relativistic-Newtonian correspondence

In order to see precisely how the Newtonian equations are obtained from the relativistic ones in the limit $c \to \infty$ we now simply have to rewrite Eqs. (3.24)–(3.26) and (3.29), (3.30) in terms of variables corresponding to the relativistic tensor densities. Define ω and $\bar{\gamma}^{ab}$ as in Sec. 2 and let

$$\bar{\pi} := \tfrac{4}{3}K^r_r \tag{3.31}$$

and

$$\bar{\pi}_{ab} := \omega^{1/3}(\tfrac{1}{3}K^r_r\gamma_{ab} - K_{ab}). \tag{3.32}$$

We now write the gravitational equations (3.32)–(3.26) in the form

$$\partial_t \bar{\gamma}^{ab} - \pounds_X \bar{\gamma}^{ab} = -2\omega^{1/3}\bar{\pi}^{ab}, \tag{3.33}$$

$$\partial_t \omega - \pounds_X \omega = -\tfrac{3}{4}\omega\bar{\pi}, \tag{3.34}$$

$$\mathscr{R}_{ab} := -R_{ab} + \tfrac{1}{3}R\gamma_{ab} = 0, \tag{3.35}$$

$$\mathscr{H} := -\omega R = 0, \tag{3.36}$$

$$\mathscr{J}_a := -2\omega^{2/3}\nabla^r\bar{\pi}_{ra} - \omega\partial_a\bar{\pi} = 0, \tag{3.37}$$

$$\begin{aligned}\partial_t \bar{\pi} - \pounds_X \bar{\pi} = &-\tfrac{4}{3}(\nabla^2\Phi - 4\pi G\rho)\\ &+ \tfrac{4}{3}\omega^{-2/3}\bar{\pi}^{rs}\bar{\pi}_{rs} + \tfrac{1}{4}\bar{\pi}^2.\end{aligned} \tag{3.38}$$

It is now obvious that the system of equations (3.33)–(3.38) is the limit for $c^{-2} \to 0$ (and $N \to 1$) of the corresponding relativistic system (2.52)–(2.57). Also the hydrodynamical equations (2.43) and (2.44) go over into (3.29) and (3.30) if we observe that $V = (1 - |v|^2c^{-2})^{1/2}$ and also f tend to 1. (We have assumed that all physical and geometrical quantities introduced remain regular and that the equation of state is $\mu = \rho c^2 + \epsilon(\rho)$ with the function $\epsilon(\rho)$ either independent of c or having a regular limit for $c^{-2} \to 0$.)

4. GEOMETRICAL INTERPRETATION

Fischer and Marsden have interpreted the canonical formalism for general relativity geometrically in terms of symplectic structures on Banach manifolds. (See, for example, the review paper.[3]) We show here how in that framework the limit to the Newtonian theory can be understood. This formalism is necessarily global, the Banach manifolds not being defined without precise asymptotic conditions (except for compact Σ_t where the Newtonian limit will not normally exist). We will not discuss the technical analytical aspects here, but simply outline the geometrical structure.

Limit of the Hamiltonian structure?

First we review the relativistic case closely following Ref. 3. Let the configuration manifold Q be the set of pairs (γ, b) where γ is a Riemannian metric on Σ[44] and b a map from Σ into B. We still write Q if we parametrize this set instead by $q = (\bar{\gamma}, \omega, b)$. Let TQ be the L_2-cotangent bundle of Q, i.e., the set $(q,\omega) = (\bar{\gamma}^{ij}, \omega, b^A, \bar{\pi}_{ij}, \bar{\pi}, P_A)$ with the pairing between $T_q Q = \{(k^{ij}, l, B^A)\}$ and $T_q{}^*Q$ being given by

$$\langle \omega, K \rangle = \int_\Sigma (\bar{\pi}_{ij}\, k^{ij} + \bar{\pi} l + P_A\, B^A) d^3x. \tag{4.1}$$

It follows that Hamilton's equations of motion deriving from the action principle (2.25) are equivalent to saying that the time flow generator $\mathscr{X}$ on T^*Q is given by

$$\mathscr{X} \lrcorner\, \Omega_c = \Psi^*(Z), \tag{4.2}$$

where Ω_c is the standard symplectic form on the cotangent bundle, the map

$$\Psi = (c^2\mathscr{H}, \mathscr{J}): T^*Q \to C_d^\infty(\Sigma) \times \Lambda_d^1(\Sigma) = : \mathscr{N}, \tag{4.3}$$

and Z an arbitrary cross section of $T^*\mathscr{N}$. Here we follow the notation of Ref. 3 where C_d^∞ denotes the set of scalar densities on Σ and Λ_d^1 the set of 1-form densities so that $Z = (N, X)$ is a pair consisting of a scalar function and a vector field on Σ.

The constraint equations mean that the motion is confined to the submanifold[46] $\mathscr{P}_1 : = \Psi^{-1}(0)$ so that we must also have

$$\Psi_*(\mathscr{X}) = 0, \tag{4.4}$$

which can, in principle, impose additional restrictions on $\mathscr{X}$. But it is well known that in view of (2.37), Eq. (4.4) is automatically satisfied in this case.

On the other hand, the time flow $\mathscr{X}$ on $\mathscr{P}_1$ is still determined only up to arbitrary Z which must be fixed or "factored out" before one obtains the true "space of motions" or "space of solutions" or "space of all degrees of freedom." In principle, this is done by a generalization of the symplectic reduction method of Marsden and Weinstein[47] described in Ref. 48. In practice, one chooses gauge conditions. The result is a symplectic manifold of the true gravitational degrees of freedom $\times$ the manifold of maps $\Sigma \to T^*B$ which must be further reduced by factoring out the diffeomorphisms on B.

What happens when we approach the limit $c^{-2} \to 0$? We have described our phase space Q in terms of quantities that all have regular limits and the system of evolution and constraint Eqs. (2.32), (2.34), (2.52)–(2.57) transforms correctly into the corresponding Newtonian system for $c \to \infty$. But the gravitational Newtonian system (3.33)–(3.38) is no longer of the form (4.2) for a standard symplectic form on a cotangent bundle. Also the map Ψ used in (4.2) has no regular limit.

Constraint algorithm for the Newtonian equations

It appears that even in terms of these carefully chosen, and in the limit well-behaved, variables the Hamiltonian structure (Ω_c, Ψ) on T^*Q has no Newtonian limit. Nevertheless the system (3.33)–(3.38) of evolution and constraint equations of the Newtonian theory can be studied just like a degenerate Hamiltonian system. (See, for example, Gotay and Nester.[49]) But this is somewhat cumbersome to carry out, especially if no gauge conditions are imposed. We will simply show how it means that degrees of freedom connected with the conformal 3-metric $\bar{\gamma}^{ij}$ are frozen out.

The constraints $\mathscr{R}_{ij} = 0$ and $\mathscr{H} = 0$ must hold for all t so that also their time derivatives must vanish. This leads to the *secondary constraints*

$$\Pi_{ij} : = \omega^{-1/3}\nabla^r\nabla_r\bar{\pi}_{ij} - \tfrac{3}{4}(\nabla_i\nabla_j\bar{\pi} - \tfrac{1}{3}\nabla^r\nabla_r\bar{\pi}\gamma_{ij}) = 0 \tag{4.5}$$

(where the primary constraints have been used). This however, represents an elliptic system for $\bar{\pi}_{ij}$ when $\bar{\pi}$ is known which will, whenever the asymptotic conditions are such that the Laplacian has an inverse, uniquely determine $\bar{\pi}_{ij}$ in terms of $\bar{\pi}$. No further condition on $\bar{\pi}$ arises since the divergence of Π_{ij} and the Laplacian of $\mathscr{J}_a$ lead to the same equation.

Gauge conditions

Of course, all this is unnecessary since one can use nonrotating Cartesian coordinates and put $\gamma_{ab} = \delta_{ab}$ throughout. Let us see, however, how these coordinate conditions follow naturally from the maximal slicing and minimal distortion condition discussed in Sec. 2. Again we assume appropriate asymptotic conditions to have an invertible Laplacian.

The *maximal slicing condition* has the limit

$$\mathscr{T} : = \bar{\pi} = 0, \tag{4.6}$$

so that (4.5) now implies

$$\bar{\pi}_{ij} = 0 \tag{4.7}$$

and (3.38) now becomes the *Poisson equation*

$$\nabla^2\Phi = 4\pi G\rho. \tag{4.8}$$

The shift vector field X is still arbitrary, however, and determines the time derivative of the components γ_{ab} of the (flat) metric of the time slices. The *minimal distortion condition*, finally, leads to [see (2.47)] the elliptic equation

$$\nabla^r\nabla_r X^i + \tfrac{1}{3}\nabla^i\nabla_r X^r = 0 \tag{4.9}$$

on Euclidean 3-space which implies $X^i = 0$ if one requires that X^i fall off sufficiently fast at infinity.[50]

At this stage we have recovered the standard classical description of a gravitating perfect fluid. In this form the theory admits again a symplectic description, even a (degenerate) Lagrangian one as was mentioned at the beginning of Sec. 3.

5. THE LOWEST-ORDER POST-NEWTONIAN APPROXIMATION

Let $\lambda \mapsto (\mathbf{g}_{\alpha\beta}(\lambda), \mathbf{T}_{\alpha\beta}(\lambda))$ or $\lambda \mapsto (\bar{\gamma}^{ij}, \mu, \bar{\pi}, \rho, v_i, \Phi, X^i)$ be a one-parameter family of solutions of the relativistic field equations to a given equation of state $\epsilon = \epsilon(\lambda, \rho)$ on a space-time diffeomorphic to $\mathbb{R}^4$ satisfying the appropriate asymptotic falloff conditions. We assume that for $\lambda \to 0$ the solution approaches a given Newtonian solution in the standard gauge, i.e., such that for $\lambda = 0$,

$$\gamma_{ii} \text{ flat}, \quad \partial_t \gamma_{ij} = 0, \quad \bar{\pi}_{ij} = 0, \quad \bar{\pi} = 0, \quad X^i = 0, \tag{5.1}$$

and ρ, v_i, and Φ satisfy (3.1)–(3.3).

To linearize the equations on the Newtonian background we write

$$\delta F := \left. \frac{\partial F}{\partial \lambda} \right|_{\lambda = 0}, \tag{5.2}$$

where F is any of the above tensor densities describing the field. For small positive values, at least, λ must be a continuous monotonic function of c^{-1}, so that without loss of generality we can assume that $\lambda = c^{-\alpha}$ $(\alpha > 0)$. With this identification we will now differentiate Eqs. (2.52)–(2.57) with respect to λ, put λ equal to 0 and use (5.1) and (5.2). It turns out that for $\alpha > 2$ one would get conditions on the Newtonian background field rather than on the variations, which is not what is wanted. For $0 < \lambda < 2$ we get the same system as below for $\lambda = 2$ but without the inhomogeneous terms. We will therefore state the linearized equations only for $\lambda = 2$ and, moreover, in tensorial form in terms of the variables

$$\bar{c}^{ij} := \mu^{-2/3} \delta \bar{\gamma}^{ij}, \quad \nu := \mu^{-1} \delta \mu, \quad \bar{q}_{ij} := \mu^{-1/3} \delta \bar{\pi}_{ij}. \tag{5.3}$$

Then they are

$$\partial_t \bar{c}^{ij} + 2\bar{q}^{ij} + 2(\nabla^{(i} \delta X^{j)} - \tfrac{1}{3} \nabla_r \delta X^r \gamma^{ij}) = 0, \tag{5.4}$$

$$\partial_t \nu - \nabla_r \delta X^r + \tfrac{3}{4} \delta \bar{\pi} = 0, \tag{5.5}$$

$$\begin{aligned} &\tfrac{1}{2} \nabla^r \nabla_r \bar{c}_{ij} - \nabla_{(i} \nabla^r c_{j)r} + \tfrac{1}{3} \nabla_r \nabla_s \bar{c}^{rs} \gamma_{ij} \\ &\quad - \tfrac{1}{3}(\nabla_i \nabla_j \nu - \tfrac{1}{3} \nabla^r \nabla_r \nu \gamma_{ij}) = -\nabla_i \nabla_j \Phi + \tfrac{1}{3} \nabla^2 \Phi \gamma_{ij}, \end{aligned} \tag{5.6}$$

$$\begin{aligned} &\partial_t \delta \bar{\pi} + \tfrac{4}{3}(\partial_r \Phi \nabla_s \bar{c}^{sr} + \nabla_r \nabla_s \Phi \bar{c}^{rs} + \tfrac{1}{3} \partial^r \Phi \partial_r \nu - \tfrac{2}{3} \nabla^2 \Phi \nu) \\ &\quad + \tfrac{4}{3}(\nabla^2 \delta \Phi - 4\pi G \delta \rho) = \tfrac{1}{3} k \rho (\Phi + \epsilon + 3 p \rho^{-1} + 2|v|^2), \end{aligned} \tag{5.7}$$

$$\nabla_r \nabla_s \bar{c}^{rs} + \tfrac{4}{3} \nabla^r \nabla_r \nu = -k\rho, \tag{5.8}$$

$$-2\nabla^r \bar{q}_{ri} - \partial_i \delta \bar{\pi} = k \rho v_i. \tag{5.9}$$

One should add the variations of the matter and momentum balance equations. But it turns out that they simply amount to evolution equations for $\delta\rho$ and δv_i. Since it is rather arbitrary, how the matter variables and the equations of state should vary with c, no insight is gained from looking at these equations in such a general framework. We will simply assume that $\delta\rho$ and δv_i are arbitrarily prescribed on some initial time slice and that they evolve according to these linearized matter equations.

Some of the above variations may not be truly relativistic terms. Variations of the form $\delta F = \pounds_\xi F$ where ξ is an arbitrary asymptotically vanishing vector field on each time slice are simply due to infinitesimal coordinate changes in the Newtonian background. We will factor them out as follows.

Let, for any dynamical variable F,

$$\delta F = \hat{\delta F} + \pounds_\xi F, \tag{5.10}$$

where the vector field ξ is chosen such that

$$\bar{c}^{ij} = \hat{c}^{ij} - 2\nabla^{(i} \xi^{j)} + \tfrac{2}{3} \nabla_r \xi^r \gamma^{ij} \tag{5.11}$$

and

$$\nabla_r \hat{c}^{ri} = 0. \tag{5.12}$$

Under the asymptotic conditions we assume there is a unique such vector field for any $\bar{c}^{ij}$, namely the solution of the elliptic equation

$$\nabla^r \nabla_r \xi^i + \tfrac{1}{3} \nabla^i (\nabla_r \xi^r) = -\nabla_r \bar{c}^{ri}. \tag{5.13}$$

(For the precise statement and the proof see Christodoulou and O'Murchadha.[45]) Explicitly (5.10) means that we put

$$\begin{aligned} &\nu = \hat{\nu} + \nabla_r \xi^r, \quad \delta\Phi = \hat{\delta\Phi} + \pounds_\xi \Phi, \quad \delta\rho = \hat{\delta\rho} + \pounds_\xi \rho, \\ &\delta v_i = \hat{\delta v_i} + \pounds_\xi v_i. \end{aligned} \tag{5.14}$$

For the variations of $\bar{\pi}_{ij}$ and $\bar{\pi}$ which correspond to the time derivatives of the metric we choose, however,

$$\bar{q}_{ij} = \hat{q}_{ij} - \nabla_{(i} W_{j)} + \tfrac{1}{3} \nabla^r W_r \gamma_{ij}, \tag{5.15}$$

and

$$\delta \bar{\pi} = \hat{q} - \tfrac{4}{3} \nabla^r W_r, \tag{5.16}$$

where

$$W_i := \delta X_i - \partial_t \xi_i. \tag{5.17}$$

(Recall that we work on the flat background metric and that $\partial_t \gamma_{ij} = 0$.)

Equations (5.4)–(5.9) now take the form

$$\partial_t \hat{c}^{ij} + 2\hat{q}^{ij} = 0, \tag{5.18}$$

$$\partial_t \hat{\nu} + \tfrac{3}{4} \hat{q} = 0, \tag{5.19}$$

$$\tfrac{1}{2} \nabla^r \nabla_r \hat{c}_{ij} - \tfrac{1}{3} \nabla_i \nabla_j \hat{\nu} = \nabla_i \nabla_j \Phi, \tag{5.20}$$

$$\begin{aligned} &\partial_t (\hat{q} + \nabla^r W_r) + \tfrac{4}{3}(\nabla^3 \delta\Phi - 4\pi G \delta\rho \\ &\qquad + \nabla_r \nabla_s \Phi \hat{c}^{rs} - \partial^r \Phi \partial_r \hat{\nu} - \tfrac{1}{6} k \rho \hat{\nu}) \\ &\quad = \tfrac{1}{4} k \rho (\Phi + \epsilon + 3 p \rho^{-1} + 2|v|^2), \end{aligned} \tag{5.21}$$

$$\nabla^r \nabla_r \hat{\nu} = -\tfrac{3}{4} k \rho, \tag{5.22}$$

$$\nabla_i \hat{q} - 2\nabla^r \nabla_{[r} W_{i]} = -k \rho v_i. \tag{5.23}$$

This system turns out to be consistent and can largely be solved explicitly. Differentiating (5.20) with respect to t and using (5.18) gives

$$\nabla^r \nabla_r \hat{q}_{ij} - \tfrac{1}{4} \nabla_i \nabla_j \hat{q} = -\nabla_i \nabla_j \partial_t \Phi, \tag{5.24}$$

and from the traces of this we conclude that

$$\hat{q} = 4 \partial_t \Phi. \tag{5.25}$$

Equation (5.22) gives $\nabla^2 \hat{\nu} = -12 G \pi \rho = -3 \nabla^2 \Phi$, whence

$$\hat{\nu} = -3\Phi. \tag{5.26}$$

Using (5.25) and (5.26) in (5.24) and (5.20), respectively, we see that

$$\hat{q}_{ij} = 0 \quad \text{and} \quad \hat{c}^{ij} = 0. \tag{5.27}$$

We decompose now W_i into a gradient and a transverse part,[45]

$$W_i = \partial_i W + \widehat{W}_i, \quad \nabla^r \widehat{W}_r = 0. \tag{5.28}$$

Then (5.23) gives

$$\nabla^r \nabla_r \widehat{W}_i = 4\partial_i \partial_t \Phi + k\rho v_i \tag{5.29}$$

which determines the transverse part of W_i. Finally Eq. (5.21) now takes the form

$$\begin{aligned}\nabla^2 \widehat{\delta\Phi} - 4\pi G \widehat{\delta\rho} + \partial_t \nabla^2 W \\ = 4\pi G\rho(\epsilon - \Phi + 3p\rho^{-1} + 2|v|^2) \\ - 3(\partial_t \partial_t \Phi + \gamma^{rs}\partial_r \Phi \partial_s \Phi).\end{aligned} \tag{5.30}$$

We see that *the vector field ξ^i and the scalar W can be arbitrarily chosen on space-time. Then if $\delta\rho$ and δv_i are given on an initial time slice all variations are uniquely determined.* In other words this means that the first-order shift vector δX^i and a scalar function (instead of the lapse) can be chosen freely as we expect since we have not yet imposed any gauge conditions.

The *maximal slicing condition* requires that $\delta\bar{\pi} = 0$, whence, by (5.16) and (5.25),

$$\nabla^2 W = -3\partial_t \Phi \tag{5.31}$$

which determines W and therefore W_i uniquely. If also the *minimal distortion condition* is imposed we have $\nabla^r \nabla_r \partial_t \xi^i + \frac{1}{3}\nabla^i \partial_r \xi^r = 0$ and thus $\partial_t \xi^i = 0$. In this case the first-order post-Newtonian approximation is completely determined by the Newtonian background field except for an arbitrary *time-independent* infinitesimal coordinate transformation on the flat background.

If we had chosen $\lambda = c^{-1}$ all equations for the variations would have been homogeneous and it is easily seen that then no nontrivial first-order variations exist.

We have thus rederived the usual results for the first post-Newtonian approximation in a systematic manner and in a very flexible form. Instead of imposing the maximal slicing and minimal distortion conditions it would still be possible to use (four-dimensional) *harmonic coordinates*. But this condition looks much less attractive in our geometrical framework although it is compatible with maximal slicing on the Newtonian background (only). It also implies that Cartesian coordinates must be used in the Newtonian background provided the usual asymptotic conditions are imposed.

ACKNOWLEDGMENTS

We thank J. Ehlers and M. Walker for discussions and for helpful comments on the manuscript. The first author (H.P.K.) also acknowledges hospitality and support of the Max-Planck-Institut für Physik and Astrophysik.

[1] R. Arnowitt, S. Deser, and C. W. Misner, *Gravitation*, edited by L. Witten (Wiley, New York, 1962), pp. 227–265.

[2] J. Isenberg and J. Nester, *General Relativity and Gravitation*, Vol. 1, edited by A. Held (Plenum, New York, 1980), pp. 23–97.

[3] A. E. Fischer and J. E. Marsden, *General Relativity*, edited by S. W. Hawkins and W. Israel (Cambridge U. P., New York, 1979), pp. 138–211.

[4] J. W. York, Jr., *Sources of Gravitational Radiation*, edited by L. L. Smarr (Cambridge U. P., New York, 1979), pp. 83–126.

[5] K. Friedrichs, Math. Annalen **98**, 566 (1927).

[6] G. Daucourt, Acta Phys. Polon. **65**, 637–646 (1964).

[7] H. P. Künzle, Gen. Rel. Grav. **7**, 445–457 (1976).

[8] J. Ehlers, *Grundlagenprobleme der modernen Physik*, edited by J. Nitsch *et al.* (Bibiliographisches Institut, Mannheim, 1981), pp. 65–84.

[9] S. Chandrasekhar, Astrophys. J. **142**, 1488–1521 (1965).

[10] S. Chandrasekhar and F. P. Esposito, Astrophys. J. **160**, 153–179 (1970).

[11] J. L. Anderson and T. C. DeCanio, Gen. Rel. Grav. **6**, 197–238 (1975).

[12] G. D. Kerlick, Gen. Rel. Grav. **12**, 467–482 (1980).

[13] T. Futamase and B. F. Schutz, Phys. Rev. D (to appear).

[14] H. P. Künzle, Ann. Inst. Henri Poincaré **27**, 337–362 (1972).

[15] C. Duval and H. P. Künzle, Rep. Math. Phys. **13**, 351–368 (1978).

[16] M. LeBellac and J. M. Levy-Leblond, Nuovo Cimento B **14**, 217–233 (1973).

[17] K. Kuchar, Phys. Rev. D **22**, 1285–1299 (1980).

[18] C. Duval and H. P. Künzle Gen. Rel. Grav. (to appear).

[19] A. H. Taub, Proc. Sympos. Appl. Math. **1**, 148–157 (1949).

[20] A. H. Taub, Commun. Math. Phys. **15**, 235–254 (1969).

[21] B. F. Schutz, Jr., Phys. Rev. D **2**, 2762–2773 (1970).

[22] B. F. Schutz, Jr., Phys. Rev. D **4**, 3559–3566 (1971).

[23] J. R. Ray, J. Math. Phys. **13**, 1451–1453 (1972) and Nuovo Cimento B **71**, 19–26 (1982).

[24] V. Moncrief, Phys. Rev. D **16**, 1702–1705 (1977).

[25] J. Demaret and V. Moncrief, Phys. Rev. D **21**, 2785–2793 (1980).

[26] R. L. Seliger and G. B. Whitham, Proc. R. Soc. London A **305**, 1–25 (1968).

[27] D. Lynden-Bell and J. Katz, Proc. R. Soc. London A **378**, 179–205 (1981).

[28] J. Katz and D. Lynden-Bell, Proc. R. Soc. London A **381**, 263–274 (1982).

[29] S. D. Mobbs, Proc. R. Soc. London A **381**, 457–468 (1982).

[30] J.-M. Souriau, *Géometrie et Relativité* (Hermann, Paris, 1964).

[31] J. Kijowski and W. M. Tulczyjew, *A Symplectic Framework for Field Theories* (Springer, Berlin, 1979).

[32] W. Noll, *The Foundations of Mechanics and Thermodynamics* (Springer, Heidelberg, 1974).

[33] J. Kijowski and W. M. Tulczyjew, "Relativistic hydrodynamics of isentropic flows," preprint, Calgary, 1981.

[34] M. Cantor, Bull. Amer. Math. Soc. (New Ser.) **5**, 235–262 (1981).

[35] A. Weinstein, "The local structure of Poisson manifolds," preprint, Berkeley, 1982.

[36] Bold indices are frame labels, Greek indices run from 0 to 3, and Latin indices from 1 to 3. Upper case Latin indices refer to the material space. Bold symbols, other than indices, refer to four-dimensional objects.

[37] We follow Ehlers[8] in giving every tensor a unique physical dimension including the frame field vectors, namely $[\theta^0] = T$, $[\theta^a] = L$, $[e_0] = T^{-1}$, $[e_a] = L^{-1}$ so that *frame* components of tensors may have different physical dimensions. All coordinates for space-time and the material space are dimensionless.

[38] We define K_{ab} this way so that it has a Newtonian limit.

[39] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

[40] A. Trautman, Rep. Math. Phys. **1**, 29–62 (1970).

[41] For a more recent account see Ref. 8.

[42] C. Duval and H. P. Künzle, C. R. Acad. Sci. Paris A **285**, 813–816 (1977).

[43] The matter flow vector **j** plays a more fundamental role in the Newtonian than in the relativistic theory. Its existence can be related to the additional gauge group independent of the matter field theory. In relativistic hydrodynamics it is specifically related to the fluid theory.

[44] Satisfying asymptotic conditions like its components in an asymptotical Cartesian chart differing from the Euclidean values by functions in a suitable weighted Sobolev space (Ref. 45).

[45] D. Christodoulou and N. O'Murchadha, Commun. Math. Phys. **80**, 271–300 (1981).

[46] Fischer and Marsden prove, in the case of pure gravitation, that $\mathscr{P}_1$ is a Banach submanifold of T^*Q.

[47] J. E. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121–130 (1974).

[48] J. E. Marsden, *Lectures on Geometric Methods in Mathematical Physics* (SIAM, Philadelphia, 1981).

[49] M. J. Gotay and J. M. Nester, Ann. Inst. Henri Poincaré **30**, 129–142 (1979) and **32**, 1–13 (1980).

[50] Recall from (3.22) that X incorporates "effects" due to rotating coordinate frames. They can only be eliminated by asymptotic conditions.

Separation of variables and symmetry operators for the neutrino and Dirac equations in the space-times admitting a two-parameter abelian orthogonally transitive isometry group and a pair of shearfree geodesic null congruences[a)]

N. Kamran and R. G. McLenaghan
Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(Received 28 June 1983; accepted for publication 7 October 1983)

We show that there exist a coordinate system and null tetrad for the space-times admitting a two-parameter abelian orthogonally transitive isometry group and a pair of shearfree geodesic null congruences in which the neutrino equation is solvable by separation of variables if and only if the Weyl tensor is Petrov type D. The massive Dirac equation is separable if in addition the conformal factor satisfies a certain functional equation. As a corollary, we deduce that the neutrino equation is separable in a canonical system of coordinates and tetrad for the solution of Einstein's type D vacuum or electrovac field equations with cosmological constant admitting a nonsingular aligned Maxwell field and that the Dirac equation is separable only in the subclass of Carter's $[\widetilde{A}]$ solutions and the Debever–McLenaghan null orbit solution A_0. We also compute the symmetry operators which arise from the above separability properties.

PACS numbers: 04.20.Jb, 03.65.Ge, 02.20.Km

1. INTRODUCTION

Central to the analytic solution of the relativistic hydrogen atom problem[1,2] stands the fact that in Minkowski space-time, the Dirac equation for a central potential is solvable by separation of variables in spherical coordinates, the symmetry operator[3] which underlies this separability property being the total angular momentum operator well known to Dirac himself.[4]

In curved space-time, although the separability properties of scalar equations such as the Hamilton-Jacobi and Klein–Gordon equations have been studied in detail by Carter[5] in the physically important situation of a charged rotating black hole described by the Kerr–Newman solution, it is only in the last decade that significant progress has been achieved in understanding the separability of higher spin wave equations. In particular, the separability of the Weyl neutrino equation (spin one-half), Maxwell's equations (spin one), and the perturbed Einstein gravitational field equations (spin two) has been established in the Kerr background by Teukolsky[6] and, using an analogous method, in the seven-parameter Plebański–Demiański background by Dudley and Finley.[7] All attempts to extend Teukolsky's separation method, where previously decoupled single-component equations are separated, to the Dirac equation for a massive charged spin one-half particle were unsuccessful until Chandrasekhar[8] ingeniously performed separation prior to decoupling, establishing the separability of the Dirac equation in the Kerr background. Chandrasekhar's result, which was immediately extended to the Kerr–Newman background by Page[9] and Toop[10] and to the Kinnersley case II vacuum solutions by Güven,[11] was subsequently analyzed by Carter and McLenaghan[12,13] who constructed an operator commuting with the Dirac operator in the Kerr–Newman background, admitting the separated solutions as eigenfunctions with the separation constant as eigenvalue, which is the symmetry operator associated to this separability, generalizing the total angular momentum operator existing in flat space-time. The symmetry operators for the Dirac operator on an aribtrary curved background were given a tensorial interpretation by Carter and McLenaghan[14] and a tensorial characterization by McLenaghan and Spindel.[15] The symmetry operators for the neutrino operator have recently been characterized tensorially by Kamran and McLenaghan.[16]

The purpose of this paper is to perform a systematic study of the separability properties of the neutrino and massive charged Dirac equations and an explicit computation of the associated symmetry operators in the class of Lorentzian spaces characterized by the existence of a two-dimensional abelian group of local isometries acting orthogonally transitively and the existence of two geodesic and shearfree real null congruences. The motivation behind this choice of background lies in a theorem of Debever, Kamran, and McLenaghan[17,18] which gives a canonical form for the metric and Maxwell field for the class $\mathfrak{D}$ of solutions of Einstein's vacuum and electrovac field equations with cosmological constant, for Petrov type D, with a nonsingular aligned Maxwell field. This canonical form, in which the Einstein–Maxwell equations have been integrated and a single expression for the general solution has been given,[19] is the starting point of our study, our aim being to impose only those conditions that are required for separability rather than work in the explicit form of the integrated solutions. Section 2 is devoted to the explicit statement of our hypotheses and results. Sections 3, 4, 5, and 6 contain the proofs of our theorems.

2. HYPOTHESIS AND STATEMENT OF RESULTS

We are investigating the separability properties of the neutrino equation

[a)] This work was supported in part by a grant from the National Sciences and Engineering Research Council of Canada.

 0022-2488/84/041019-09$02.50

$$\mathbf{H}_N\psi := i\gamma^k\nabla_k\psi = 0, \tag{2.1a}$$

and the Dirac equation

$$\mathbf{H}_D\psi := [i\gamma^k(\nabla_k - ieA_k) - \sqrt{2}\mu_e I]\psi = 0, \tag{2.1b}$$

where $\{\gamma_k\}$ is a set of Dirac matrices associated to a Lorentzian metric g_{ij}, ∇_k denotes the covariant differentiation operator[20] on four-spinors corresponding to the choice of γ^k's and to the Levi-Cività connection of g_{ij}, and where $A = A_k\, dx^k$ is a 1-form field, in the class of Lorentzian spaces admitting a metric of the form

$$ds^2 = 2(\theta^1\theta^2 - \theta^3\theta^4), \tag{2.2a}$$

where

$$\theta^1 = (\sqrt{2}T(w,x))^{-1}|Z(w,x)|^{1/2}[fW(w)Z(w,x)^{-1} \times(\epsilon_1 du + m(x)dv) + g^{-2}W(w)^{-1}dw], \tag{2.2b}$$

$$\theta^2 = (\sqrt{2}T(w,x))^{-1}|Z(w,x)|^{1/2}[W(w)Z(w,x)^{-1} \times(\epsilon_1 du + m(x)dv) - fg^{-2}W(w)^{-1}dw], \tag{2.2c}$$

$$\theta^3 = (\sqrt{2}T(w,x))^{-1}|Z(w,x)|^{1/2}[X(x)Z(w,x)^{-1} \times(\epsilon_2 du + p(w)dv) + iX(x)^{-1}dx], \tag{2.2d}$$

$$\theta^4 = \bar\theta^3, \tag{2.2e}$$

with

$$A_k dx^k = T(w,x)(2Z(w,x))^{-1/2}[g^{-2}H(w)W(w)^{-1} \times(f\theta^1 + \theta^2) + G(x)X(x)^{-1}(\theta^3 + \theta^4)], \tag{2.2f}$$

where

$$Z(w,x) = \epsilon_1 p(w) - \epsilon_2 m(x), \quad g = ((1+f^2)/2)^{1/2}, \tag{2.3}$$

and where ϵ_1, ϵ_2, and f are real constants satisfying $\epsilon_1^2 + \epsilon_2^2 \neq 0$ and all functions are real valued.

The Lorentzian spaces whose metric is given by (2.2) and (2.3) are characterized[21] by the following properties:

(i) There exists a two-dimensional abelian group of local isometries acting orthogonally transitively.

(ii) There exist two geodesic and shearfree real null congruences.

The motivation for the study of the separability properties of Eqs. (2.1) in the Lorentzian spaces whose metric is given by (2.2) and (2.3) with the 1-form field $A = A_k\, dx^k$ by (2.3) lies in the following result: Let $\mathfrak{D}$ denote the class of solutions of Einstein's vacuum and electrovac field equations with cosmological constant, which may be written as

$$R_{ij} - \tfrac{1}{2}Rg_{ij} + \lambda g_{ij} = F_{ik}F_j{}^k - \tfrac{1}{4}g_{ij}F_{rs}F^{rs}, \tag{2.4a}$$

$$F_{ik;}{}^k = 0, \quad F_{[ij;k]} = 0; \quad F_{ij} = 2A_{[j;i]}, \tag{2.4b}$$

where we permit the cosmological constant λ and the electromagnetic field tensor F_{ij} to vanish,that satisfy the following conditions:

H1. The Weyl tensor is everywhere of Petrov type D.
H2. If the Maxwell field tensor F_{ij} is nonzero, it is nonsingular with its principal null directions aligned with the repeated principal null directions of the Weyl tensor.
H3. The hypothesis of the generalized Goldberg–Sachs theorem is satisfied, insuring that the null congruences associated to the principal directions of the Weyl tensor are geodesic and shearfree.

Then, we have the following result, proved in Debever, Kamran, and McLenaghan (DKM).[22]

Theorem 1: For every solution in $\mathfrak{D}$, there exists a system of local coordinates (u,v,w,x) and a null basis of 1-forms $\{\theta^a\}$ such that the metric and self-dual Maxwell field take the form

$$ds^2 = 2(\theta^1\theta^2 - \theta^3\theta^4), \tag{2.5a}$$

$$\overset{+}{F} = B(w,x)(\theta^1\wedge\theta^2 - \theta^3\wedge\theta^4), \tag{2.5b}$$

where B is a complex valued function, with the 1-forms θ^a and a real vector potential A given by (2.2) and (2.3).

The field equations (2.4) have been integrated for the class $\mathfrak{D}$ and a single expression has been given in DKM for the general solution in the tetrad and coordinates of Theorem 1.

We shall work in the general context of the metric and 1-form (2.2) rather than in the explicit form of the integrated $\mathfrak{D}$ solutions, imposing those conditions that are necessary and sufficient for separability and subsequently determining those subclasses of $\mathfrak{D}$ satisfying the separability requirements.

The main results of this paper are given in the following theorems:

Theorem 2: In the Lorentzian space of Theorem 1, the neutrino equation (2.1a) admits, in the Weyl representation, an R- separable[23] solution of the form

$$\psi = \begin{pmatrix} P_0 \\ P_1 \\ \overline{Q}^{0'} \\ \overline{Q}^{1'} \end{pmatrix} = e^{i(\alpha u + \beta v)}T^{3/2}Z^{-1/4}\begin{pmatrix} e^{i\mathscr{B}}R_1(x)S_2(w) \\ e^{i\mathscr{B}}R_2(x)S_1(w) \\ e^{-i\mathscr{B}}R_1(x)S_1(w) \\ e^{-i\mathscr{B}}R_2(x)S_2(w) \end{pmatrix}, \tag{2.6a}$$

where α and β are arbitrary real constants and where

$$d\mathscr{B} = (4Z)^{-1}(\epsilon_1 m'(x)dw + \epsilon_2 p'(w)dx), \tag{2.6b}$$

if and only if the Petrov type D condition H_1, which insures the existence of $\mathscr{B}$[24] as given in (2.6b), is satisfied.

Corollary: In all space-times in the class $\mathfrak{D}$, the neutrino equation (2.1a) admits, in the Weyl representation, an R-separable solution of the form (2.6a).

Special cases of our corollary have been obtained using a different approach, where starting from the Weyl neutrino equation in terms of two-spinors, previously decoupled second-order equations are separated, by Teukolsky[25] for the Kerr solution and Dudley and Finley[26] for the seven-parameter Plebański–Demiański solutions. It should be noted that this decoupling prior to separation procedure seems to require either the vacuum $\mathfrak{D}$ equations or all but two of the electrovac $\mathfrak{D}$ equations[27] rather than the weaker conformally invariant Petrov type D equation appearing in Theorem 2. Although Theorem 2 and its corollary deal with the separability of the neutrino equation in $\mathfrak{D}$ and hence could be entirely formulated in terms of a single two-spinor, we employ a four-spinor formalism to be able to compare the above results with the following separability theorems for the massive charged Dirac equation, which may be formulated either in terms of four-spinors or a pair of two-spinors.

Theorem 3: In the Lorentzian space of Theorem 1, the Dirac equation (2.1b) admits, in the Weyl representation, an R-separable solution of the form given in (2.6a) and (2.6b) if and only if:

(i) The Petrov type D condition (H_1), which insures the existence of $\mathscr{B}$ as given in (2.6b), is satisfied.

(ii) There exist real valued functions $h(w)$ and $g(x)$ such that

$$Z^{1/2}T^{-1}e^{2i\mathscr{B}} = h(w) + ig(x). \tag{2.7}$$

Theorem 3 enables us to characterize those space-times in $\mathfrak{D}$ admitting a separable Dirac equation through the following corollary.

Corollary: Within the class $\mathfrak{D}$ of space-times, the Dirac equation (2.1b) admits, in the Weyl representation, an R-separable solution of the form given in Eqs. (2.6a) and (2.6b) only in Carter's $[\widetilde{A}]$ family of solutions and in the A_0 null orbit solution of Debever and McLenaghan.[28]

Our corollary becomes an immediate consequence of Theorem 3 once the values of Z, T, and $\mathscr{B}$ as given in DKM, which contains the result of the integration of the field equation (2.2) in the canonical (u,v,w,x) class of coordinates of Theorem 1, are substituted in Eq. (2.7). Special cases of our corollary have been obtained for the Kerr solution by Chandrasekhar,[29] whose approach was subsequently applied by Page[30] and Toop[31] to show separability of the Dirac equation in the Kerr–Newman solution and Güven[32] in the Kinnersley case II vacuum solutions.

There is a well-known connection between separation of variables and constants of the motion. In the case of the Hamilton–Jacobi equation for the massive charged particle orbits, it has been established in DKM that Carter's $[\widetilde{A}]$ family of solutions and the A_0 null orbit solution are those space-times in $\mathfrak{D}$ where one has separation. This separation of variables gives rise to a quadratic first integral of the equations of motion which implies the existence of a (0,2)-Killing tensor and whose Poisson bracket with the Hamiltonian vanishes. In the case of the Hamilton–Jacobi equation for the zero rest mass particles (null geodesics) it has been shown by Debever and McLenaghan[33] (in the electrovac case) and Czapor and McLenaghan[34] (in the vacuum case) that a separable coordinate system exists for every solution in $\mathfrak{D}$. This separation of variables gives in turn rise to a function on the cotangent bundle which is also quadratic in the momenta, whose existence yields a (0,2)-conformal Killing tensor, and whose Poisson bracket with the Hamiltonian will be *proportional* to this Hamiltonian.

When studying the motion of test particles in the first-quantized limit rather than in the classical description provided by the Hamilton–Jacobi equation, the connection between separation of variables and constants of motion is reflected in *commutation* relations rather than Poisson bracket relations (for a general discussion of this point, see Carter[35]). In the case of the Klein–Gordon equation for charged (massive or massless) spin zero particles, the separability properties in the class $\mathfrak{D}$ have been studied in detail in DKM. The constants of the motion appear there as symmetry operators admitting the separated solutions as eigenfunctions with the separation constant as eigenvalue. Explicitly, denoting by $\mathbf{H}_K$ the massless *conformally invariant* Klein–Gordon operator, for which R-separability was established in DKM for the entire $\mathfrak{D}$ class, we have a symmetry operator $\mathbf{K}_k$ defined by

$$\mathbf{K}_K = TU_K^{-1}(U_{Kx}\mathbf{W}_{Kw} - U_{Kw}\mathbf{W}_{Kx})T^{-1}, \tag{2.8a}$$

where

$$U_{Kx} = -\epsilon_2 mT^{-2},\ U_{Kw} = \epsilon_1 pT^{-2},\ U_K = U_{Kw} + U_{Kx}, \tag{2.8b}$$

$$\begin{aligned}\mathbf{W}_{Kx} =& X^2\partial_x^2 + (X^2)'\partial_x + X^{-2}[(-m\partial_u + \epsilon_1\partial_v)^2 \\ &+ ieG(-m\partial_u + \epsilon_1\partial_v)] \\ &+ GX^{-2}(ie(-m\partial_u + \epsilon_1\partial_v) - e^2G) + \tfrac{1}{6}(X^2)'',\end{aligned} \tag{2.8c}$$

$$\begin{aligned}\mathbf{W}_{Kw} =& fW^2\partial_w^2 + (f(W^2)' - g^{-2}(1-f^2)(p\partial_u - \epsilon_2\partial_v))\partial_w \\ &+ g^{-4}W^{-2}[-f(p\partial_u - \epsilon_2\partial_v)^2 \\ &- g^2ieH(p\partial_u - \epsilon_2\partial_v)] \\ &- 2^{-1}g^{-2}(1-f^2)p'\partial_u \\ &+ HW^{-2}(-ieg^{-2}(p\partial_u - e_2\partial_v) \\ &+ e^2g^{-4}fH) + \tfrac{1}{6}f(W^2)''.\end{aligned} \tag{2.8d}$$

If the wave function takes the R-separable form

$$\psi(u,v,w,x) = e^{i(\alpha u + \beta v)}\psi_1(w)\psi_2(x)T(w,x), \tag{2.8e}$$

the symmetry operator $\mathbf{K}_k$ satisfies the eigenvalue equation

$$\mathbf{K}_K\psi = \lambda_K\psi, \tag{2.8f}$$

where λ_K denotes the separation constant arising from the separation of variables, and the commutation relation

$$[\mathbf{K}_K, \mathbf{H}_K] = r_K\mathbf{H}_K, \tag{2.8g}$$

where

$$r_K = TU_K^{-1}\{[U_{Kx}, \mathbf{W}_{Kw}] - [U_{Kw}, \mathbf{W}_{Kw}]\}T^{-1}, \tag{2.8h}$$

and where use has been made in the expressions of those $\mathfrak{D}$ field equations required for separability. It can be checked that in the massive case, one obtains in the subclass of $\mathfrak{D}$ where one has separability, that is the Carter $[\widetilde{A}]$ and the A_0 null orbit solutions, a symmetry operator which reduces to a commuting operator and still satisfies the eigenvalue equation (2.8f).

Central to the Poisson bracket discussion of the constants of the motion for the charged particle orbits in $\mathfrak{D}$ and the commutator discussion of the constants of the motion for the first-quantized spin zero test particles in $\mathfrak{D}$, as they arise from the separability of the Hamilton–Jacobi and Klein–Gordon equations, stands a lemma, first stated by Carter[36] in the Poisson bracket case and subsequently employed by Carter and McLenaghan[37] in the commutator case, that will also be applied to the derivation of the symmetry operators associated to the separability of the neutrino and Dirac equations in $\mathfrak{D}$.

Lemma: Let M and N be matrix differential operators and P, P_1, P_2 be nonsingular matrices not involving any differentiation operators. Then

$$
\begin{aligned}
[P^{-1}(M+N),&P^{-1}(P_1N-P_2M)]\\
&=P^{-1}[M,P^{-1}(P_1+P_2)]N\\
&\quad-P^{-1}[N,P^{-1}(P_1+P_2)]M\\
&\quad+P^{-2}(P_1+P_2)[M,N]+P^{-1}\{[N,P_1]-[M,P_2]\\
&\quad-NP^{-1}[P_1,P]-MP^{-1}[P_2,P]\}P^{-1}(M+N).
\end{aligned}
\tag{2.9}
$$

The commutation relation (2.8g) can be deduced from this lemma by letting

$$
P=U_k,\quad P_1=U_{Kx},\quad P_2=U_{Kw},\quad M=\mathbf{W}_{Kx},\quad N=\mathbf{W}_{Kw}
\tag{2.10}
$$

in Eq. (2.9).

The separability Theorems 2 and 3 and their corollaries give rise to *matrix* symmetry operators for the neutrino and Dirac equations, which we will compute using the above Lemma. The expressions of these symmetry operators are given in the following theorems.

Theorem 4: In the single expression given in DKM for the class $\mathfrak{D}$ solutions, in the tetrad and coordinates of Theorem 1, let $\mathbf{K}_N$ be the operator defined by

$$
\mathbf{K}_N=S_NU_N^{-1}(U_{Nx}\mathbf{W}_{Nw}-U_{Nw}\mathbf{W}_{Nx})S_N^{-1},
\tag{2.11a}
$$

where

$$
\begin{aligned}
S_N&\\
=T^{3/2}\,&\mathrm{diag}\{[b^2(cw\cos\gamma+k)-i(cx\sin\gamma+l)]^{-1/2},\\
&[b^2(cw\cos\gamma+k)-i(cx\sin\gamma+l)]^{-1/2},\\
&[b^2(cw\cos\gamma+k)+i(cx\sin\gamma+l)]^{-1/2},\\
&[b^2(cw\cos\gamma+k)+i(cx\sin\gamma+l)]^{-1/2}\},
\end{aligned}
\tag{2.11b}
$$

$$
U_{Nx}=T^{-1}i(cx\sin\gamma+l)\mathrm{diag}(1,-1,1,-1),
\tag{2.11c}
$$

$$
U_{Nw}=T^{-1}b^2(cw\cos\gamma+k)\mathrm{diag}(1,-1,-1,1),
\tag{2.11d}
$$

$$
U_N=U_{Nx}+U_{Nw},
\tag{2.11e}
$$

$$
\mathbf{W}_{Nx}=\begin{pmatrix}0&0&0&L_x^-\\0&0&-L_x^+&0\\0&L_x^-&0&0\\-L_x^+&0&0&0\end{pmatrix},
$$

$$
\mathbf{W}_{Nw}=\begin{pmatrix}0&0&L_w^+&0\\0&0&0&-L_w^-\\-L_w^-&0&0&0\\0&L_w^+&0&0\end{pmatrix},
\tag{2.11f}
$$

$$
L_x^+=2^{-1/2}(iX\partial_x+X^{-1}(\epsilon_1\partial_v-m\partial_u)+i2^{-1}X'),
\tag{2.11g}
$$

$$
L_x^-=2^{-1/2}(-iX\partial_x+X^{-1}(\epsilon_1\partial_v-m\partial_u)-i2^{-1}X'),
\tag{2.11h}
$$

$$
L_w^+=2^{-1/2}(W\partial_w+fg^{-2}W^{-1}(p\partial_u-\epsilon_2\partial_v)+2^{-1}W'),
\tag{2.11i}
$$

$$
L_w^-=2^{-1/2}(-fW\partial_w+g^{-2}W^{-1}(p\partial_u-\epsilon_2\partial_v)-2^{-1}fW').
\tag{2.11j}
$$

The operator $\mathbf{K}_N$ has the following properties:

(i) The operator $\mathbf{K}_N$ is a symmetry operator for the neutrino operator $\mathbf{H}_N$, that is

$$
[\mathbf{K}_N,\mathbf{H}_N]=r_N\mathbf{H}_N,
\tag{2.12a}
$$

where

$$
r_N=S_NU_N^{-1}\{[U_{Nx},\mathbf{W}_{Nw}]-[U_{Nw},\mathbf{W}_{Nx}]\}S_N^{-1}.
\tag{2.12b}
$$

(ii) The R-separable solution (2.6) is an eigenspinor of $\mathbf{K}_N$ with the separation constant λ_N arising from the separation of variables as eigenvalue:

$$
\mathbf{K}_N\psi=\lambda_N\psi.
\tag{2.13}
$$

Now, for the Dirac equation, we have:

Theorem 5: In the expression given for the Carter $[\tilde{A}]$ solutions and the A_0 null orbit solution in DKM, which uses the tetrad and coordinates of Theorem 1, let $\mathbf{K}_D$ be the operator defined by

$$
\mathbf{K}_D=S_DU_D^{-1}(U_{Dx}\mathbf{W}_{Dw}-U_{Dw}\mathbf{W}_{Dx})S_D^{-1},
\tag{2.14a}
$$

where

$$
\begin{aligned}
&S_D=T^{-3/2}S_N,\quad U_{Dx}=TU_{Nx},\\
&U_{Dw}=TU_{Nw},\quad U_D=U_{Dx}+U_{Dw},
\end{aligned}
\tag{2.14b}
$$

$$
\mathbf{W}_{Dx}=\begin{pmatrix}\mu_e(cx\sin\gamma+l)&0&0&D_x^-\\0&-\mu_e(cx\sin\gamma+l)&-D_x^+&0\\0&D_x^-&\mu_e(cx\sin\gamma+l)&0\\-D_x^+&0&0&-\mu_e(cx\sin\gamma+l)\end{pmatrix},
$$

$$
\mathbf{W}_{Dx}=\begin{pmatrix}-i\mu_eb^2(cw\cos\gamma+k)&0&D_w^+&0\\0&i\mu_eb^2(cw\cos\gamma+k)&0&-D_w^-\\-D_w^-&0&i\mu_eb^2(cw\cos\gamma+k)&0\\0&D_w^+&0&-i\mu_eb^2(cw\cos\gamma+k)\end{pmatrix},
\tag{2.14c}
$$

$$
\begin{aligned}
D_x^+&=L_x^+-i2^{-1/2}eX^{-1}G,\\
D_x^-&=L_x^--i2^{-1/2}eX^{-1}G,
\end{aligned}
\tag{2.14d}
$$

$$
\begin{aligned}
D_w^+&=L_w^+-i2^{-1/2}fg^{-2}eW^{-1}H,\\
D_w^-&=L_w^--i2^{-1/2}g^{-2}eW^{-1}H.
\end{aligned}
\tag{2.14e}
$$

The operator $\mathbf{K}_D$ has the following properties:

(i) The operator $\mathbf{K}_D$ commutes with the Dirac operator $\mathbf{H}_D$:

$$
[\mathbf{K}_D,\mathbf{H}_D]=0.
\tag{2.15}
$$

(ii) The R-separable solution (2.6) is an eigenspinor of $\mathbf{K}_D$ with the separation constant λ_D arising from the separation of variables as eigenvalue:

$$\mathbf{K}_D \psi = \lambda_D \psi. \tag{2.16}$$

Operators commuting with the Dirac operator have previously been considered, the first example of which is given in flat space-time by the total angular momentum operator $\overline{J}$ [38] which is notably associated to the separability in spherical coordinates of the Dirac equation for a central potential and whose eigenvalues yield quantum numbers. First-order operators which commute with the Dirac operator on a curved background were obtained in the Kerr–Newman solution by Carter and McLenaghan,[39] using the above Lemma in their analysis of Chandrasekhar's separation of variables procedure. A tensorial characterization of the operators commuting with the Dirac operator on an arbitrary curved background was given by Carter and McLenaghan[40] and McLenaghan and Spindel.[41] The symmetry operators for the neutrino operator on an arbitrary curved space-time have been characterized tensorially by Kamran and McLenaghan.[42]

3. PROOF OF THEOREM 2

The essence of our proof lies in the analysis performed by Carter and McLenaghan[43] of Chandrasekhar's separation of variables procedure for the Dirac equation in the Kerr space-time.

In the Newman–Penrose spin coefficient formalism, the neutrino equation (2.1a) reads,[44] using the Weyl representation for ψ,

$$\begin{pmatrix} 0 & 0 & D+\bar{\epsilon}-\bar{\rho} & \delta+\bar{\pi}-\bar{\alpha} \\ 0 & 0 & \bar{\delta}+\bar{\beta}-\bar{\tau} & \Delta+\bar{\mu}-\bar{\gamma} \\ \Delta+\mu-\gamma & -(\delta+\beta-\tau) & 0 & 0 \\ -(\bar{\delta}+\pi-\alpha) & D+\epsilon-\rho & 0 & 0 \end{pmatrix} \begin{pmatrix} P_0 \\ P_1 \\ \bar{Q}^{0'} \\ \bar{Q}^{1'} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.1a}$$

In the tetrad and coordinates of Theorem 1, we have, letting

$$\theta^1 = n_i\, dx^i, \quad \theta^2 = l_i\, dx^i, \quad \theta^3 = -\bar{m}_i\, dx^i, \quad \theta^4 = -m_i\, dx^i, \tag{3.1b}$$

the following:

$$\begin{aligned} D+\epsilon-\rho &= T2^{-1/2}Z^{-1/2}[W\partial_w + fg^{-2}W^{-1}(p\partial_u - \epsilon_2\partial_v) \\ &\quad - \tfrac{1}{4}WZ^{-1}\epsilon_1(-p'+im') + \tfrac{1}{2}W' - \tfrac{3}{2}T_{,w}T^{-1}W], \end{aligned} \tag{3.2a}$$

$$\begin{aligned} \Delta-\gamma+\mu &= T2^{-1/2}Z^{-1/2}[-fW\partial_w + g^{-2}W^{-1}(p\partial_u - \epsilon_2\partial_v) \\ &\quad + \tfrac{1}{4}fWZ^{-1}\epsilon_1(-p'+im') - \tfrac{1}{2}fW' + \tfrac{3}{2}fT_{,w}T^{-1}W], \end{aligned} \tag{3.2b}$$

$$\delta+\beta-\tau = T2^{-1/2}Z^{-1/2}[-iX\partial_x + X^{-1}(\epsilon_1\partial_v - m\partial_u) - \tfrac{1}{4}XZ^{-1}\epsilon_2(p'-im') - \tfrac{1}{2}iX' + \tfrac{3}{2}iT_{,x}T^{-1}X], \tag{3.2c}$$

$$\bar{\delta}-\alpha+\pi = T2^{-1/2}Z^{-1/2}[iX\partial_x + X^{-1}(\epsilon_1\partial_v - m\partial_u) + \tfrac{1}{4}XZ^{-1}\epsilon_2(p'-im') + \tfrac{1}{2}iX' - \tfrac{3}{2}iT_{,x}T^{-1}X]. \tag{3.2d}$$

We first perform the four-spinor transformation defined by

$$\psi = S\psi', \tag{3.3a}$$

with

$$S = T^{3/2}Z^{-1/4}\,\mathrm{diag}(e^{i\mathscr{B}}, e^{i\mathscr{B}}, e^{-i\mathscr{B}}, e^{-i\mathscr{B}}), \tag{3.3b}$$

where

$$d\mathscr{B} = (4Z)^{-1}(\epsilon_1 m'\, dw + \epsilon_2 p'\, dx), \tag{3.3c}$$

the integrability condition of which is satisfied thanks to the Petrov type D condition (H1).[45]

The resulting operator acting on the transformed spinor ψ' is

$$S^{-1}\mathbf{H}_N S = TZ^{-1/2}\begin{pmatrix} 0 & 0 & e^{-2i\mathscr{B}}L_w^+ & e^{-2i\mathscr{B}}L_x^- \\ 0 & 0 & e^{-2i\mathscr{B}}L_x^+ & e^{-2i\mathscr{B}}L_w^- \\ e^{2i\mathscr{B}}L_w^- & -e^{2i\mathscr{B}}L_x^- & 0 & 0 \\ -e^{2i\mathscr{B}}L_x^+ & e^{2i\mathscr{B}}L_w^+ & 0 & 0 \end{pmatrix}, \tag{3.4}$$

where the operators L_x^+, L_x^-, L_w^+, and L_w^- are defined by Eqs. (2.11). We next multiply the operator which acts on the transformed spinor ψ' by the nonsingular separating matrix U defined by

$$U = Z^{1/2}T^{-1}\,\mathrm{diag}(e^{2i\mathscr{B}}, -e^{2i\mathscr{B}}, -e^{-2i\mathscr{B}}, e^{-2i\mathscr{B}}), \tag{3.5}$$

to obtain an operator $\mathbf{W}_N$ given by

$$\mathbf{W}_N = US^{-1}\mathbf{H}_N S = \begin{pmatrix} 0 & 0 & L_w^+ & L_x^- \\ 0 & 0 & -L_x^+ & -L_w^- \\ -L_w^- & L_x^- & 0 & 0 \\ -L_x^+ & L_w^+ & 0 & 0 \end{pmatrix}. \tag{3.6}$$

Now, taking into account the existence for all solutions in $\mathfrak{D}$ of a two-parameter abelian group of local isometries,[46] which manifests itself in the form given in Theorem 1 for the metrices of the $\mathfrak{D}$ class by u and v being ignorable coordinates, we may write

$$\psi'(u,v,w,x) = e^{i(\alpha u + \beta v)}\begin{pmatrix} F_1(w,x) \\ F_2(w,x) \\ G_1(w,x) \\ G_2(w,x)\end{pmatrix}, \tag{3.7}$$

where α and β are arbitrary constants.

When the form (3.7) for ψ' is substituted into the transformed equation $\mathbf{W}_N\psi' = 0$, we obtain the following system of first-order partial differential equations:

$$\mathring{L}_w^+ G_1 + \mathring{L}_x^- G_2 = 0, \quad -\mathring{L}_w^- F_1 + \mathring{L}_x^- F_2 = 0, \tag{3.8a}$$

$$\mathring{L}_x^+ G_1 + \mathring{L}_w^- G_2 = 0, \quad -\mathring{L}_x^+ F_1 + \mathring{L}_w^+ F_2 = 0, \tag{3.8b}$$

where $\mathring{L}_x^+$, $\mathring{L}_x^-$, $\mathring{L}_w^+$, and $\mathring{L}_w^-$ denote the operators given in Eqs. (2.11) where the derivatives with respect to ignorable coordinates $\partial/\partial u$ and $\partial/\partial v$ have been, respectively, replaced by $i\alpha$ and $i\beta$. Following Chandrasekhar,[47] we seek separable solutions of the form

$$F_1(w,x) = R_1(x)S_2(w), \quad G_1(w,x) = R_1(x)S_1(w), \tag{3.9a}$$

$$F_2(w,x) = R_2(x)S_1(w), \quad G_2(w,x) = R_2(x)S_2(w), \tag{3.9b}$$

which, when substituted into the system (3.8), yield

$$R_1(x)\mathring{L}_w^+ S_1(w) + S_2(w)\mathring{L}_x^- R_2(x) = 0, \tag{3.10a}$$

$$S_1(w)\mathring{L}_x^+ R_1(x) + R_2(x)\mathring{L}_w^- S_2(w) = 0, \tag{3.10b}$$

$$-R_1(x)\mathring{L}_w^- S_2(w) + S_1(w)\mathring{L}_x^- R_2(x) = 0, \tag{3.10c}$$

$$-S_2(w)\mathring{L}_x^+ R_1(x) + R_2(x)\mathring{L}_w^+ S_1(w) = 0. \tag{3.10d}$$

These relations give then immediately

$$\mathring{L}_w^+ S_1(w) = \lambda_1 S_2(w), \quad \mathring{L}_w^- S_2(w) = \lambda_3 S_1(w), \tag{3.11a}$$

$$\mathring{L}_x^- R_2(x) = -\lambda_1 R_1(x), \quad \mathring{L}_x^- R_2(x) = \lambda_3 R_1(x), \tag{3.11b}$$

$$\mathring{L}_w^- S_2(w) = \lambda_2 S_1(w), \quad \mathring{L}_w^+ S_1(w) = \lambda_4 S_2(w), \tag{3.11c}$$

$$\mathring{L}_x^+ R_1(x) = -\lambda_2 R_2(x), \quad \mathring{L}_x^+ R_1(x) = \lambda_4 R_2(x), \tag{3.11d}$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are the separation constants, which for consistency must satisfy

$$\lambda_1 = \lambda_4 = -\lambda_2 = -\lambda_3 (= \lambda_N). \tag{3.12}$$

Equations (3.11) are thus equivalent to the system of coupled first-order *ordinary differential equations*

$$\mathring{L}_w^+ S_1(w) = \lambda_N S_2(w), \quad \mathring{L}_w^- S_2(w) = -\lambda_N S_1(w), \tag{3.13a}$$

$$\mathring{L}_x^- R_2(x) = -\lambda_N R_1(x), \quad \mathring{L}_x^+ R_1(x) = \lambda_N R_2(x), \tag{3.13b}$$

and the proof of Theorem 2 is complete. Note that the first-order system (3.13) can be written as a system of *decoupled* second-order ordinary differential equations which reads as follows:

$$\mathring{L}_w^+ \mathring{L}_w^- S_2(w) = -\lambda_N^2 S_2(w), \quad \mathring{L}_x^- \mathring{L}_x^+ R_1(x) = -\lambda_N^2 R_1(x), \tag{3.14a}$$

$$\mathring{L}_w^- \mathring{L}_w^+ S_1(w) = -\lambda_N^2 S_1(w), \quad \mathring{L}_x^+ \mathring{L}_x^- R_2(x) = -\lambda_N^2 R_2(x). \tag{3.14b}$$

It may be noted that one recovers as special cases of Eqs. (3.14) Teukolsky's[48] decoupled equations in the Kerr background and Dudley and Finley's[49] decoupled equations in the Plebański–Demiański background when the metric functions appearing in Theorem 1 are appropriately specialized.

4. PROOF OF THEOREM 3

The first steps of our proof are similar to those of Theorem 2. In the Newman–Penrose formalism, the Dirac equation (2.1b) reads,[50] using the Weyl representation for ψ,

$$\begin{pmatrix} -i\mu_e & 0 & D+\bar{\epsilon}-\bar{\rho}-eA_1 & \delta+\bar{\pi}-\bar{\alpha}-ieA_3 \\ 0 & -i\mu_e & \bar{\delta}+\bar{\beta}-\bar{\tau}-ieA_4 & \Delta+\bar{\mu}-\bar{\gamma}-ieA_2 \\ \Delta+\mu-\gamma-ieA_2 & -(\delta+\beta-\tau)+ieA_3 & -i\mu_e & 0 \\ -(\bar{\delta}+\pi-\alpha)+ieA_4 & D+\epsilon-\rho-ieA_1 & 0 & -i\mu_e \end{pmatrix}\begin{pmatrix} P_0 \\ P_1 \\ \bar{Q}^{0'} \\ \bar{Q}^{1'}\end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0\end{pmatrix}, \tag{4.1}$$

where A_a are the components of the vector potential 1-form in the basis $\{\theta^a\}$. Having written Eq. (4.1) with the tetrad, coordinates, and vector potential given in Eqs. (2.3), (2.4), (2.5), and (3.1b) we perform, as we did for the neutrino equation, the four-spinor transformation defined by Eq. (3.3) and we multiply on the left the Dirac operator $S^{-1}\mathbf{H}_D S$ acting on the transformed spinor ψ' by the separating matrix U given in (3.5) to obtain an operator $\mathbf{W}_D$ given by

$$\mathbf{W}_D = US^{-1}\mathbf{H}_D S = \begin{pmatrix} -i\mu_e Z^{1/2}T^{-1}e^{2i\mathscr{B}} & 0 & D_w^+ & D_x^- \\ 0 & i\mu_e Z^{1/2}T^{-1}e^{2i\mathscr{B}} & -D_x^- & -D_w^- \\ -D_w^- & D_x^- & i\mu_e Z^{1/2}T^{-1}e^{-2i\mathscr{B}} & 0 \\ -D_x^+ & D_w^+ & 0 & -i\mu_e Z^{1/2}T^{-1}e^{-2i\mathscr{B}} \end{pmatrix}, \tag{4.2}$$

where the operators D_x^+, D_x^-, D_w^+, and D_w^- are those given in Eqs. (2.14).

A necessary condition for the separability of the transformed equation $\mathbf{W}_D\psi' = 0$ is that each diagonal entry in the expression given in Eq. (4.2) for the operator $\mathbf{W}_D$ split into the sum of a function of x and a function of w, in other words that there exist real valued functions $f(x)$, $g(x)$, $h(w)$, and $k(w)$ such that

$$Z^{1/2}T^{-1}e^{2i\mathscr{B}} = [f(x) + ig(x)] + [h(w) + ik(w)]. \tag{4.3}$$

However, this condition is not sufficient for separability with ψ' given (following Chandrasekhar) by

$$\psi(u,v,w,x) = e^{i(\alpha u + \beta v)} \begin{pmatrix} H_1(x)K_2(w) \\ H_2(x)K_1(w) \\ H_1(x)K_1(w) \\ H_2(x)K_2(w) \end{pmatrix}. \tag{4.4}$$

Indeed, substituting Eqs. (4.3) and (4.4) into $\mathbf{W}_D \psi' = 0$ with $\mathbf{W}_D$ as given by (4.2), we obtain, in the same way in which we obtained Eqs. (3.11), the following equations:

$$\mathring{D}_w^+ K_1(w) - i\mu_e(h(w) + ik(w))K_2(w) = \mu_1 K_2(w), \tag{4.5a}$$

$$\mathring{D}_x^- H_2(x) - i\mu_e(f(x) + ig(x))H_1(x) = -\mu_1 H_1(x), \tag{4.5b}$$

$$\mathring{D}_w^- K_2(w) - i\mu_e(h(w) + ik(w))K_1(w) = \mu_2 K_1(w), \tag{4.5c}$$

$$\mathring{D}_x^+ H_1(x) - i\mu_e(f(x) + ig(x))H_2(x) = -\mu_2 H_2(x), \tag{4.5d}$$

$$\mathring{D}_w^- K_2(w) - i\mu_e(h(w) - ik(w))K_1(w) = \mu_3 K_1(w), \tag{4.5e}$$

$$-\mathring{D}_x^- H_2(x) - i\mu_e(f(x) - ig(x))H_1(x) = -\mu_3 H_1(x), \tag{4.5f}$$

$$\mathring{D}_w^+ K_1(w) - i\mu_e(h(w) - ik(w))K_2(w) = \mu_4 K_2(w), \tag{4.5g}$$

$$-\mathring{D}_x^+ H_1(x) - i\mu_e(f(x) - ig(x))H_2(x) = -\mu_4 H_2(x), \tag{4.5h}$$

where $\mathring{D}_x^+$, $\mathring{D}_x^-$, $\mathring{D}_w^+$, and $\mathring{D}_w^-$ denote the operators given in Eqs. (2.14) where the derivatives with respect to the ignorable coordinates $\partial/\partial u$ and $\partial/\partial v$ have been, respectively, replaced by $i\alpha$ and $i\beta$.

Subtracting Eq. (4.5g) from Eq. (4.5a), we obtain

$$k(w) \equiv c_1, \tag{4.6a}$$

where c_1 is a real constant and adding Eq. (4.5h) to Eq. (4.5d) we obtain

$$f(x) \equiv c_2, \tag{4.6b}$$

where c_2 is a real constant. Defining now new functions $\tilde{g}(x)$ and $\tilde{h}(w)$ by

$$\tilde{g}(x) = g(x) + c_1, \quad \tilde{h}(w) = h(w) + c_2, \tag{4.7}$$

the condition (4.3) reduces, dropping tildes, to the following:

$$Z^{1/2}T^{-1}e^{2i\mathscr{B}} = h(w) + ig(x), \tag{4.8}$$

which along with the Petrov type D condition H1 constitutes a necessary and sufficient condition for the existence of a separable solution of the form given in Eqs. (2.6); indeed, substituting Eq. (4.8) into Eq. (4.2) yields, by a consistency argument similar to that used for the neutrino equation, the following system of coupled first-order *ordinary differential equations*:

$$\begin{aligned} &\mathring{D}_w^+ K_1(w) - i\mu_e h(w)K_2(w) = \lambda_D K_2(w), \\ &\mathring{D}_w^- K_2(w) - i\mu_e h(w)K_1(w) = -\lambda_D K_1(w), \end{aligned} \tag{4.9a}$$

$$\begin{aligned} &\mathring{D}_x^- H_2(x) + \mu_e g(x)H_1(x) = -\lambda_D H_1(x), \\ &\mathring{D}_x^+ H_1(x) + \mu_e g(x)H_2(x) = \lambda_D H_2(x). \end{aligned} \tag{4.9b}$$

This completes the proof of Theorem 3.

It should be noted that one recovers as special cases of Eqs. (4.9) the separated equations of Chandrasekhar[51] in the Kerr solution, of Carter and McLenaghan[52] in the Kerr–Newman solution and of Güven[53] in the Kinnersley case II vacuum solutions. One also notes that the first-order system (4.9) can be written as a system of *decoupled* second-order ordinary differential equations which reads as follows:

$$\mathring{D}_w^- \mathring{D}_w^+ K_1(w) - \frac{i\mu_e(\mathring{D}_w^- h(w))}{\lambda_D + i\mu_e h(w)} \mathring{D}_w^+ K_1(w) + (\lambda_D^2 + \mu_e^2 h(w)^2)K_1(w) = 0, \tag{4.10a}$$

$$\mathring{D}_w^+ \mathring{D}_w^- K_2(w) - \frac{i\mu_e(\mathring{D}_w^+ h(w))}{-\lambda_D + i\mu_e h(w)} \mathring{D}_w^- K_2(w) + (\lambda_D^2 + \mu_e^2 h(w)^2)K_2(w) = 0, \tag{4.10b}$$

$$\mathring{D}_x^- \mathring{D}_x^+ H_1(x) - \frac{\mu_e(\mathring{D}_x^- g(x))}{-\lambda_D + \mu_e g(x)} \mathring{D}_x^+ H_1(x) + (\lambda_D^2 - \mu_e^2 g(x)^2)H_1(x) = 0, \tag{4.10c}$$

$$\mathring{D}_x^+ \mathring{D}_x^- H_2(x) - \frac{\mu_e(\mathring{D}_x^+ g(x))}{\lambda_D + \mu_e g(x)} \mathring{D}_x^- H_2(x) + (\lambda_D^2 - \mu_e^2 g(x)^2)H_2(x) = 0. \tag{4.10d}$$

We again remark that one recovers as special cases of Eqs. (4.10) the decoupled second-order equations of Chandrasekhar[54] in the Kerr solution, of Page[55] and Carter and McLenaghan[56] in the Kerr–Newman solution, and of Güven[57] in the Kinnersley case II vacuum solutions.

5. PROOF OF THEOREM 4

The main tool in our proof is the identity (2.9). If we go back to Eq. (3.6), we see that we have a splitting

$$\mathbf{W}_N = \mathbf{W}_{Nw} + \mathbf{W}_{Nx}, \tag{5.1}$$

where $\mathbf{W}_{Nw}$ and $\mathbf{W}_{Nx}$ are given in Eqs. (2.11). It should be noted that except for derivatives with respect to the ignorable coordinates $\partial/\partial u$ and $\partial/\partial v$, $\mathbf{W}_{Nw}$ depends only on w and $\mathbf{W}_{Nx}$ depends only on x, moreover it may be readily checked that

$$[\mathbf{W}_{Nw}, \mathbf{W}_{Nx}] = 0. \tag{5.2}$$

Now, if we refer ourselves to the single expression presented in Ref. 18 for the general solution of the class $\mathfrak{D}$ field equations, we have as solution to Eqs. (2.6c)

$$\begin{aligned}\mathscr{B} = (i/4)[&\ln(b^2(cw\cos\gamma + k) - i(cx\sin\gamma + l)) \\ &- \ln(b^2(cw\cos\gamma + k) + i(cx\sin\gamma + l))],\end{aligned} \tag{5.3}$$

which implies that the matrix U_N, obtained upon substitution of $\mathscr{B}$'s expression (4.3) into the definition (3.5) of the matrix U, splits as follows

$$U_N = U_{Nw} + U_{Nx}, \tag{5.4a}$$

with U_{Nw} and U_{Nx} given by Eqs. (2.11).

Also, one checks immediately that

$$[U_{Nw}, U_{Nx}] = 0. \tag{5.4b}$$

Thus, if we set in the identity (2.9)

$$P = U_N, \quad P_1 = U_{Nx}, \quad P_2 = U_{Nw}, \quad M = W_{Nx}, \quad N_{Nw}, \tag{5.5}$$

we obtain part (i) of Theorem 4, thanks to Eqs. (5.2) and (5.4). To prove part (ii) of Theorem 4, we proceed as follows. Equations (3.13) are equivalent, using Eqs. (3.9), to the following relations:

$$\mathring{L}_w^+ F_2(w,x) = \lambda_N G_2(w,x), \quad \mathring{L}_x^+ F_1(w,x) = \lambda_N G_2(w,x), \tag{5.6a}$$

$$\mathring{L}_w^+ G_1(w,x) = \lambda_N F_1(w,x), \quad \mathring{L}_x^+ G_1(w,x) = \lambda_N F_2(w,x), \tag{5.6b}$$

$$\mathring{L}_x^- G_2(w,x) = -\lambda_N F_1(w,x), \quad \mathring{L}_w^- G_2(w,x) = -\lambda_N F_2(w,x), \tag{5.6c}$$

$$\mathring{L}_x^- F_2(w,x) = -\lambda_N G_1(w,x), \quad \mathring{L}_w^- F_1(w,x) = -\lambda_N G_1(w,x). \tag{5.6d}$$

Now, on account of Eqs. (5.1b) and (3.7), it is easily seen that Eqs. (5.6) are equivalent to

$$\mathbf{W}_{Nw}\psi' = \lambda_N\psi', \quad \mathbf{W}_{Nx}\psi' = -\lambda_N\psi'. \tag{5.7}$$

Equations (5.7) enable us to show as follows that ψ is an eigenspinor of $\mathbf{K}_N$ with the separation constant λ_N as eigenvalue. Let us first note that S_N as defined in (2.11b) is obtained by replacing $\mathscr{B}$ with its expression (5.3) for the class $\mathfrak{D}$ solutions in the definition (3.3b) of S. We have then from (2.11a), (3.3a), (5.4a), and (5.7):

$$\begin{aligned}\mathbf{K}_N\psi &= SU_N^{-1}(U_{Nx}\mathbf{W}_{Nw}\psi' - U_{Nw}\mathbf{W}_{Nx}\psi') \\ &= \lambda_N SU_N^{-1}(U_{Nx} + U_{Nw})\psi' = \lambda_N\psi.\end{aligned} \tag{5.8}$$

6. PROOF OF THEOREM 5

If we consider the separability condition (4.8) and we refer ourselves to the single expression presented in DKM for the general solution of the class $\mathfrak{D}$ field equations, we see that in the tetrad and class of coordinates of Theorem 1, we have separability for those solutions with

$$T(w,x) = 1, \tag{6.1}$$

in the above-mentioned single expression, that is, the Carter $[\widetilde{A}]$ solutions and the A_0 null orbit solution listed in the Corollary to Theorem 3.

Using now Eq. (6.1) along with the expression (5.3) for $\mathscr{B}$ and the fact[58] that

$$Z(w,x) = b^4(cw\cos\gamma + k)^2 + (cx\sin\gamma + l)^2, \tag{6.2}$$

in the above-mentioned single expression, we see from (4.2) that there is a splitting

$$\mathbf{W}_D = \mathbf{W}_{Dw} + \mathbf{W}_{Dx}, \tag{6.3}$$

where $\mathbf{W}_{Dw}$ and $\mathbf{W}_{Dx}$ are as given in Eqs. (2.14).

It should be noted that in analogy with the neutrino case, $\mathbf{W}_{Dw}$ and $\mathbf{W}_{Dx}$ depend, except for derivatives with respect to the ignorable coordinates $\partial/\partial u$ and $\partial/\partial v$, only on w and x, respectively. We also have

$$[\mathbf{W}_{Dw}, \mathbf{W}_{Dx}] = 0. \tag{6.4}$$

Now, the matrix U_D, obtained upon substitution of $\mathscr{B}$'s expression (5.3) and the separability condition (6.1) into the definition (3.5) of the matrix U, splits as follows:

$$U_D = U_{Dw} + U_{Dx}, \tag{6.5}$$

with U_{Dw} and U_{Dx} given by Eqs. (2.14). It may then be verified by direct calculation that

$$[U_{Dw}, U_{Dx}] = 0, \quad [U_{Dw}, \mathbf{W}_{Dx}] = 0, \quad [U_{Dx}, \mathbf{W}_{Dw}] = 0. \tag{6.6}$$

It is then straightforward to obtain part (i) of Theorem 5 by applying the Identity (2.9) with

$$\begin{aligned}&P = U_D, \quad P_1 = U_{Dx}, \quad P_2 = U_{Dw}, \\ &M = \mathbf{W}_{Dx}, \quad N = \mathbf{W}_{Dw},\end{aligned} \tag{6.7}$$

and using Eqs. (6.4), (6.5), and (6.6).

To prove part (ii) of Theorem 5, we note first that Eqs. (6.2), (6.1), (5.3), and (4.8) yield

$$h(w) = b^2(cw\cos\gamma + k), \quad g(x) = cx\sin\gamma + l, \tag{6.8}$$

so that Eqs. (4.9) imply, using Eqs. (3.7), (4.4), and (6.8) that

$$\mathring{D}_w^+ F_2(w,x) - i\mu_e b^2(cw\cos\gamma + k)G_2(w,x) = \lambda_D G_2(w,x), \tag{6.9a}$$

$$\mathring{D}_x^+ F_1(w,x) + \mu_e(cx\sin\gamma + l)G_2(w,x) = \lambda_D G_2(w,x), \tag{6.9b}$$

$$\mathring{D}_w^+ G_1(w,x) - i\mu_e b^2(cw\cos\gamma + k)F_1(w,x) = \lambda_D F_1(w,x), \tag{6.9c}$$

$$\mathring{D}_x^+ G_1(w,x) + \mu_e(cx\sin\gamma + l)F_2(w,x) = \lambda_D F_2(w,x), \tag{6.9d}$$

$$\mathring{D}_x^- G_2(w,x) + \mu_e(cx\sin\gamma + l)F_1(w,x) = -\lambda_D F_1(w,x), \tag{6.9e}$$

$$\mathring{D}_w^- G_2(w,x) - i\mu_e b^2(cw\cos\gamma + k)F_2(w,x) = -\lambda_D F_2(w,x), \tag{6.9f}$$

$$\mathring{D}_x^- F_2(w,x) + \mu_e(cx\sin\gamma + l)G_1(w,x) = -\lambda_D G_1(w,x), \tag{6.9g}$$

$$\mathring{D}_w^- F_1(w,x) - i\mu_e b^2(cw\cos\gamma + k)G_1(w,x) = -\lambda_D G_1(w,x). \tag{6.9h}$$

Now, Eqs. (6.9), using the expressions (2.14c) for $\mathbf{W}_{Dw}$ and $\mathbf{W}_{Dx}$ and Eq. (3.7), are equivalent to

$$\mathbf{W}_{Dw}\psi' = \lambda_D\psi', \quad \mathbf{W}_{Dx}\psi' = -\lambda_D\psi'. \tag{6.10}$$

We then have from Eqs. (2.14a), (3.3a), (6.5), and (6.10)

$$\begin{aligned}\mathbf{K}_D\psi &= SU_D^{-1}(U_{Dx}\mathbf{W}_{Dw}\psi' - U_{Dw}\mathbf{W}_{Dx}\psi') \\ &= \lambda_D SU_D^{-1}(U_{Dx} + U_{Dw})\psi' = \lambda_D\psi.\end{aligned} \tag{6.11}$$

Let us finally note that the expression of the operators $\mathbf{W}_{Dw}$ and W_{Dx} verifying the property (6.10) was computed in the Kerr–Newman background by Carter and McLenaghan[59] in their analysis of Chandrasekhar's separation of variables procedure.

[1]P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford. U. P., London, 1958), 4th ed., p. 269.

[2]L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1955), 2nd ed., p. 485.

[3]W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, MA, 1977), p. 2.

[4]Reference 1, p. 267.

[5]B. Carter, Phys. Rev. **174**, 1559 (1968); Commun. Math. Phys. **10**, 280 (1968).

[6]S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).

[7]A. L. Dudley and J. D. Finley, III, J. Math. Phys. **20**, 311 (1979).

[8]S. Chandrasekhar, Proc. R. Soc. London Ser. A **349**, 571 (1976).

[9]D. N. Page, Phys. Rev. D **14**, 1509 (1976).

[10]N. Toop, preprint, D. A. M. T. P., Cambridge (1976).

[11]R. Güven, Proc. R. Soc. London Ser. A **356**, 465 (1977).

[12]B. Carter and R. G.McLenaghan, Phys. Rev. D **19**, 1093 (1979).

[13]B. Carter and R. G.McLenaghan, in *Proceedings of the Second Marcel Grossmann Meeting on Recent Developments of General Relativity Trieste, 1979*), edited by Ruffinni (North-Holland, Amsterdam, 1982).

[14]Reference 12.

[15]R. G. McLenaghan and Ph. Spindel, Phys. Rev. D **20**, 409 (1979); Bull. Soc. Math. Belg. **XXXI**, 65 (1979).

[16]N. Kamran and R. G. McLenaghan, Lett. Math. Phys. **7**, 381 (1983).

[17]R. Debever, N. Kamran, and R. G. McLenaghan, J. Math. Phys. (in press).

[18]The results of Ref. 17 were announced by the same authors in Phys. Lett. **93A**, 399 (1983) and in Bull. Cl. Sci. Acad. Roy. Belg. **LXVIII**, 592 (1982).

[19]Reference 17.

[20]A. Lichnerowicz, in *Relativity, Groups and Topology*, edited by B. S.

DeWitt and C. DeWitt (Gordon and Breach, New York, 1963), p. 823.
[21]R. Debever, Bull. Cl. Sci. Acad. Roy. Belg. **LV**, 8 (1969).
[22]Reference 17.
[23]P. Moon and D. E. Spencer, *Field Theory Handbook* (Springer-Verlag, New York, 1971), 2nd ed., p. 96.
[24]Reference 17, Eq. (4.2).
[25]Reference 6.
[26]Reference 7.
[27]J. M. Stewart and M. Walker, Proc. R. Soc. London Ser. A **341**, 49 (1974).
[28]R. Debever and R. G. McLenaghan, J. Math. Phys. **22**, 1711 (1981).
[29]Reference 8.
[30]Reference 9.
[31]Reference 10.
[32]Reference 11.
[33]Reference 28.
[34]S. R. Czapor and R. G.McLenaghan, J. Math. Phys. **23**, 2159 (1982).
[35]B. Carter, Phys. Rev. D **16**, 3395 (1977).
[36]B. Carter, in *Black Holes*, edited by B. DeWitt and C. DeWitt (Gordon and Breach, New York, 1973), p. 118.
[37]Reference 12.
[38]Reference 4.
[39]References 12 and 13.
[40]Reference 12.
[41]Reference 15.
[42]Reference 16.
[43]Reference 12.
[44]Reference 8, Eqs. (9) to (13).
[45]Reference 17, Eqs. (4.1) and (4.2).
[46]This was demonstrated in the electrovac case in Ref. 28 and in the vacuum case in Ref. 34.
[47]Reference 8, Eqs. (30).
[48]Reference 6, Eqs. (4.9) and (4.10) with $|s| = \frac{1}{2}$.
[49]Reference 7, Eqs. (5.18a) and (5.18b) with $|s| = \frac{1}{2}$.
[50]Reference 9, Eqs. (3)–(6).
[51]Reference 8, Eqs. (40) and (41).
[52]Reference 13, Eqs. (4.9) and (4.10).
[53]Reference 11, Eqs. (47)–(50).
[54]Reference 8, Eqs. (44) and (45).
[55]Reference 9, Eqs. (19) and (20).
[56]Reference 13, Eqs. (4.13) and (4.14).
[57]Reference 11, Eqs. (54) and (55).
[58]Reference 17, Eqs. (2.6a)–(2.6c).
[59]Reference 13, Eq. (4.5).

Time-asymmetric initial data for the N black hole problem in general relativity

Anil D. Kulkarni
Center for Relativity, The University of Texas, Austin, Texas 78712

(Received 18 May 1983; accepted for publication 17 June 1983)

The time-asymmetric initial-value problem for N-body systems consisting of N black holes is described. The spacelike hypersurface for setting up the initial data is chosen to be the Einstein–Rosen manifold with N bridges. An explicit procedure in the form of an infinite series is given to construct solutions of the momentum constraints on the Einstein–Rosen manifold, starting from any solution of the momentum constraints in three-dimensional Euclidean space. A sufficient condition for the absolute convergence of a wide range of these solutions is derived. In particular, the solution representing N black holes with arbitrary momenta and spins is presented. The formulation of the Hamiltonian constraint as a boundary-value problem is discussed briefly.

PACS numbers: 04.20.Jb, 02.90. + p, 97.60.Lf

I. INTRODUCTION

The N-body problem has been of interest since the beginning of general relativity. However, because of its complexity, the study of this problem has been mainly confined to situations where perturbation theory or other approximation techniques can be used; i.e., where the speeds of the interacting objects are small or gravity is weak. But to understand fully the implications of general relativity, the strong-field, high-speed situations do need to be analyzed. An impetus to these studies also comes from relativistic celestial processes. In the analysis of these situations, the full nonlinear equations of general relativity must be solved.

Recent developments[1,2] show that the solutions can be obtained by the Cauchy formulation of these problems, in which the first step is to choose a spacelike hypersurface and to set up initial data on it satisfying the constraint equations of general relativity.[3] The initial data consist of the 3-metric and the extrinsic curvature of the initial hypersurface, and the energy and momentum densities on it. From this data set, the data on a successive hypersurface is obtained by integrating the evolution equations numerically. By repeating this procedure, the geometry of spacetime is constructed.

In this paper, I describe the initial-value problem for N-body systems consisting of N black holes. The corresponding problem for noncollapsed objects will, in general, depend on their stress-energy tensor. Instead of choosing a particular stress-energy tensor, for this work I bypass this issue by considering black holes (which also simplifies the problem). For this system, the initial hypersurface is chosen to be the singularity-free and matter-free model of particles proposed by Einstein and Rosen,[4] i.e., the Einstein–Rosen manifold M_N (Fig. 1). It consists of two asymptotically flat universes (sheets) connected by N bridges (throats). The symmetry between the two sheets of this manifold makes the numerical integration of the evolution equations particularly convenient, since it suffices to evolve the data on just one sheet; the evolution on the other sheet follows by symmetry.

This model was given further attention by Misner[5] and Lindquist.[6] Misner[5] solved the initial-value problem on M_N for N particles momentarily at rest, and Lindquist[6] extended the solution to the case of N charged particles. Their data were originally interpreted as geometric models of particles, but later as black holes. All these data were obtained by taking the extrinsic curvature of M_N to be zero (time symmetric data), which necessarily yielded particles momentarily at rest. Such a data set for two identical black holes was evolved numerically by Eppley and Smarr,[7] who retained the full nonlinear nature of the head-on collision of two black holes.

Problems such as the scattering of two spinning black holes or the dynamics of a binary system of black holes orbiting each other are more interesting. But they would require initial data describing holes with arbitrary spins and momenta. That is, the extrinsic curvature of M_N can no longer be taken to be zero. For the case of a single moving and spinning black hole, assuming the Einstein–Rosen manifold M_1 to be conformally flat, Bowen and York[8] have explicitly solved the momentum constraints and have formulated the Hamiltonian constraint as a well-posed boundary-value problem. These results are extended to N black holes elsewhere.[9] The solutions (K_{ij}) of the momentum constraints for the problem of N black holes have several interesting mathematical features, and these features, especially the existence and convergence properties of the infinite sum for K_{ij}, are the main topics of this paper. In addition to the purely mathematical and aesthetic reasons, the study of convergence is

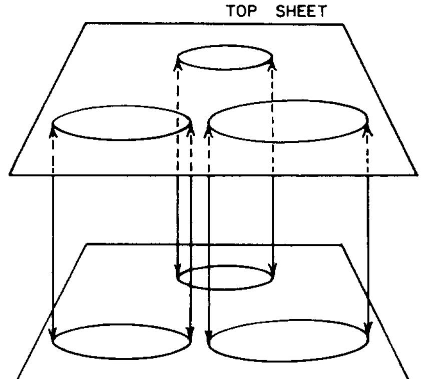


FIG. 1. The Einstein–Rosen manifold M_N with three bridges ($N = 3$) and with one dimension suppressed. Topologically, it consists of two copies of three-dimensional Euclidean spaces (top and bottom sheets) with the interiors of N nonintersecting spheres removed and their boundaries identified (as indicated by the arrows). These boundaries constitute the bridges connecting the two sheets (universes).

 0022-2488/84/041028-07$02.50

also motivated by the fact the data will be used in numerical work. For such work, the series will be truncated, and one would like to have an estimate of the accuracy of such a truncation.

The organization of this paper is as follows: First I define the Einstein–Rosen manifold M_N as a point set and describe its differentiable structure in terms of coordinate patches. Then I address the question of constructing initial data on M_N. The data must preserve the symmetry between the two sheets of this manifold. The resulting symmetry conditions on the data are derived assuming M_N to be conformally flat. Starting from any solution of the momentum constraints in the Euclidean space E^3, I give a systematic procedure, in the form of an infintie series, to construct solutions of the momentum constraints on M_N satisfying the symmetry conditions. I find a sufficient condition (which is not very restrictive) for the absolute convergence of a wide range of these solutions and examine its implications for various values of N. The problem of representing N black holes with arbitrary spins and momenta is addressed, and finally, the remaining part of the initial-value problem, the Hamiltonian constraint, is discussed briefly.

II. FORMULATION OF THE PROBLEM

The initial hypersurface M_N and its differentiable structure (which turns out to be analytic, C^ω) must be described in a computationally convenient form. Let $p = (p^1, p^2, p^3)$ represent points in a three-dimensional Euclidean space E^3. Draw N noninteresting spheres in E^3 of radii a_α and centers at c_α $(\alpha = 1, ..., N)$. Then, as point sets, the top sheet Y and the bottom sheet Z of M_N can be defined as follows (they are topologically the same; Fig. 2):

$$Y = Z = \{p\in E^3: |p - c_\alpha| > a_\alpha, \quad \alpha = 1,...,N\}.$$

Let $\bar{Y} = Y\cup B$, where $B = \cup_{\alpha=1}^N B_\alpha$, and $B_\alpha = \{p\in E^3: |p - c_\alpha| = a_\alpha\}$. Define $\bar{Z}$ similarly. Then $M_N = \bar{Y}\cup\bar{Z}$ with boundaries of $\bar{Y}$ and $\bar{Z}$ identified. The B_α's represent the bridges between the two sheets.

Having described M_N as a point set, I describe its topology in terms of the basic neighborhoods of its points. Since Y is open in E^3, every point $y\in Y$ has neighborhoods lying in Y; these define its neighborhoods in M_N. The same applies to every $z\in Z$. The basic neighborhoods of points in B are defined as follows.

Let $p\in B$ and let $B'(p,r)\subset E^3$ be the open ball with center p and radius $r > 0$. Then $B\cap B'(p,r)$ is the boundary of $N_y(p,r) = \bar{Y}\cap B'(p,r)$ as well as that of $N_z(p,r) = \bar{Z}\cap B'(p,r)$ (Fig. 2). Let $N(p,r) = N_y(p,r)\cup N_z(p,r)$ with boundaries identified. Then $\{N(p,r): r > 0\}$ constitute a collection of basic neighborhoods of p. This structure makes M_N a topological manifold.

M_N is covered with coordinate patches by defining C^ω-related coordinate maps ψ_α, $\alpha = 1, ..., N$. A particular map ψ_α will cover the two sheets (Y and Z) continuously through the αth bridge, and its range will be a subset of E^3 (whose points will be labeled as x). Let

$$X = \{X\in E^3: |x - c_\alpha| > a_\alpha, \quad \alpha = 1,...,N\}$$

and

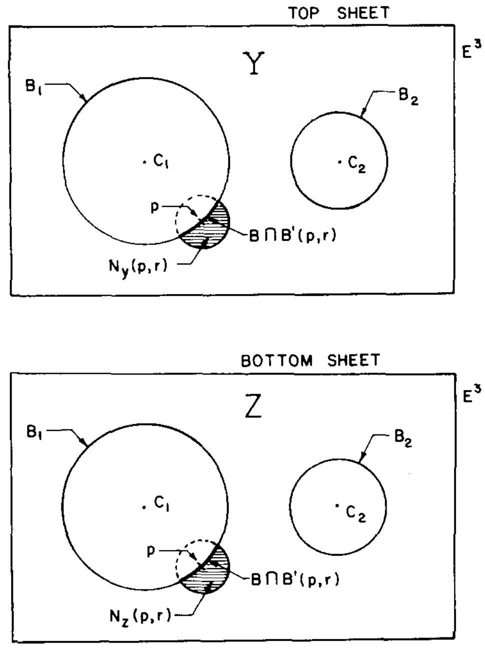


FIG. 2. A two-dimensional analog of the point sets Y and Z (with $N = 2$). The shaded regions $N_y(p,r)$ and $N_z(p,r)$ include $B\cap B'(p,r)$, their boundary.

$$S_\alpha = \{x\in E^3: |x - c_\alpha| = a_\alpha\}, \quad \alpha = 1,...,N.$$

Also, in this space, define the inversion maps $J_\alpha: E^3 - \{c_\alpha\} \to E^3 - \{c_\alpha\}$ by

$$J_\alpha(x) \equiv J_\alpha x = (a_\alpha^2/r_\alpha)n_\alpha + c_\alpha, \tag{1}$$

where

$$r_\alpha = |x - c_\alpha| \quad \text{and} \quad n_\alpha = (x - c_\alpha)/r_\alpha. \tag{2}$$

These maps are one-one and onto, and they map the regions $r_\alpha > a_\alpha$ and $0 < r_\alpha < a_\alpha$ onto each other; hence the name inversion map. Also, $J_\alpha^2 =$ identity. Thus $J_\alpha^{-1} = J_\alpha$. Because of these properties, X, S_α, and $J_\alpha[X] \equiv I_\alpha$ are disjoint sets (Fig. 3). So ψ_α's can be defined in such a way that $\psi_\alpha[Y] = X$, $\psi_\alpha[B_\alpha] = S_\alpha$, and ψ_α, and $\psi_\alpha[Z] = I_\alpha$; to be precise,

$$\psi_\alpha(p) = \begin{cases} (p^1, p^2, p^3)\in (X\cup S_\alpha) & \text{if } p\in (Y\cup B_\alpha), \\ J_\alpha(p^1, p^2, p^3)\in I_\alpha & \text{if } p\in Z. \end{cases}$$

These maps are homeomorphisms and are analytically related to each other. Thus M_N becomes an analytic manifold.

The initial-value problem on M_N is now equivalent to setting up data in the regions $(X\cup S_\alpha\cup I_\alpha)$, representing the coordinate patches $Y\cup B_\alpha\cup Z$, $\alpha = 1, ..., N$. Those parts of M_N on which the patches overlap will receive more than one data set. Hence, for consistency, all such overlapping data (that are geometric objects) must be related to each other by the pull-back map. To formulate these conditions precisely, first note that the initial data may be set up on the single set $X' \equiv \cup_{\alpha=1}^N (X\cup S_\alpha\cup I_\alpha)$. Then the bottom sheet Z receives N sets of data from I_α's, $\alpha = 1, ..., N$, whereas all other parts of M_N receive precisely one. Since I_α is $\{J_\alpha x: x\in X\}$, for every $x\in X$, the points $J_\alpha x$ $(\alpha = 1, ..., N)$ assign data to the same point in Z. Hence the consistency requirements on the data must be, for every $x\in X$ and $\alpha, \beta = 1, ..., N$,

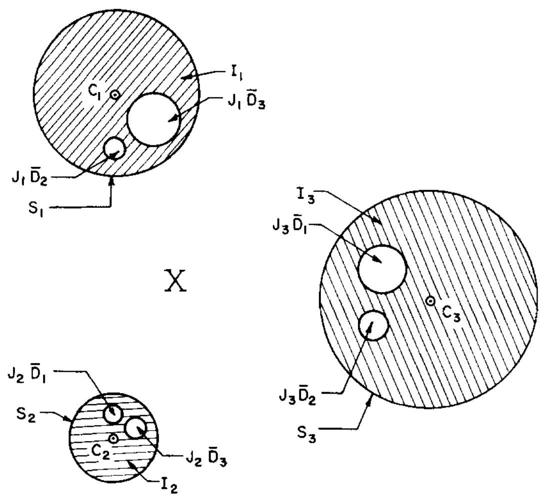


FIG. 3. A two-dimensional illustration of the range of the coordinate maps $\psi_\alpha, \alpha = 1, 2, 3$. In this case, three ψ_α's cover M_3. X is the region outside all spheres, $I_\alpha \equiv J_\alpha[X]$ is the shaded region excluding c_α, and X' is the union of X, the spheres and the shaded regions. $\bar{D}_\alpha$ [Def. (18)] is S_α together with its interior. Thus, for example, $J_2\bar{D}_1$ is the unshaded circle together with its boundary, as shown in the figure.

$$\begin{aligned}(\text{data at } (J_\beta x)) &= (\psi_\alpha \circ \psi_\beta^{-1})^*(\text{data at } (J_\alpha x)) \\ &= (J_\alpha \circ J_\beta)^*(\text{dat at } (J_\alpha x)).\end{aligned} \tag{3}$$

Here $(\psi_\alpha \circ \psi_\beta^{-1})^*$ is the pull-back associated with the map $(\psi_\alpha \circ \psi_\beta^{-1})$: $X \cup I_\beta \rightarrow X \cup I_\alpha$, and $J_\alpha \circ J_\beta$ is the restriction of $\psi_\alpha \circ \psi_\beta^{-1}$ onto I_β.

Now recall that the symmetry between the two sheets of the Einstein–Rosen manifold makes it convenient to evolve on the computer. For this reason, as well as from the aesthetic point of view, the data must also preserve this symmetry. That is, the data at $y \in Y$ and $z \in Z$ with $(y^1, y^2, y^3) = (z^1, z^2, z^3)$ must be the same or at least closely related (to be made precise). [I distinguish the 3-tuple (z^1, z^2, z^3) associated with z as an element of $Z \subset E^3$ from its coordinates $\psi_\alpha(z) \in I_\alpha$. I make a similar distinction for y.] To obtain related data at y and z, the data at $x \in X$ and $(J_\alpha x) \in I_\alpha$ must also be related, where $x = \psi_\alpha(y)$; [then $J_\alpha x = \psi_\alpha(z)$, since $\psi_\alpha(z) = J_\alpha(z^1, z^2, z^3) = J_\alpha(y^1, y^2, y^3) = J_\alpha(x^1, x^2, x^3) = J_\alpha x$]. Since J_α maps X onto I_α, it is natural to compare the data at x and $J_\alpha x$ using the pull-back map J_α^*. Hence the required symmetry conditions are, for every $x \in X$ and $\alpha = 1, ..., N$.

$$(\text{data at } x \in X) = \pm J_\alpha^*(\text{data at } J_\alpha x)). \tag{4}$$

The minus sign can occur since it gives physically meaningful results for some data functions (which may be seen by looking at the electric field[6] and the extrinsic curvature[10,11] in this language). Also note that if a geometric object at x is proportional to the object at $J_\alpha x$, then the constant of proportionality must be ± 1 (since $J_\alpha^2 =$ identity).

The symmetry conditions (4) incorporate the consistency requirements (3): Since $J_\beta^{-1} = J_\beta$, (4) is equivalent to

$$(\text{data at } (J_\beta x)) = \pm J_\beta^*(\text{data at } x \in X).$$

Therefore it follows that

$$\begin{aligned}(\text{data at } (J_\beta x)) &= \pm J_\beta^*(\text{data at } x \in X) \\ &= \pm J_\beta^*[\pm J_\alpha^*(\text{data at } (J_\alpha x))] \quad \text{from (4)} \\ &= (J_\alpha \circ J_\beta)^*(\text{data at } (J_\alpha x)).\end{aligned}$$

Thus the initial-value problem is reduced to obtaining the data on X' subject to the symmetry conditions (4).

The data must also satisfy the vacuum constraint equations of general relativity ($c = G = 1$)

$$\bar{R} + (\bar{g}_{pq}\bar{K}^{pq})^2 - \bar{K}_{ij}\bar{K}^{ij} = 0,$$
$$\bar{\nabla}_j[\bar{K}^{ij} - \bar{g}^{ij}(\bar{g}_{pq}\bar{K}^{pq})] = 0,$$

where $\bar{g}_{ij}$ and $\bar{K}^{ij}$ are the 3-metric and the extrinsic curvature of M_N, $\bar{R}$ its scalar curvature, and $\bar{\nabla}_j$ the covariant derivative associated with $\bar{g}_{ij}$. These equations can be simplified using the conformal technique, as, for example, in Ref. 12: Take tr $\bar{K} \equiv \bar{g}_{pq}\bar{K}^{pq} \equiv 0$ (maximal initial slice), and make the conformal transformations $\bar{g}_{ij} = \varphi^4 g_{ij}$ and $\bar{K}^{ij} = \varphi^{-10}K^{ij}$ ($\bar{K}_{ij} = \varphi^{-2}K_{ij}$). Then the constraint equations become an uncoupled system

$$\nabla_j K^{ij} = 0, \tag{5}$$
$$8(\Delta\varphi) = -\varphi^{-7}K_{ij}K^{ij}, \tag{6}$$

where ∇_j and Δ are convariant derivative and the Laplacian associated with g_{ij}.

For simplicity, I shall assume M_N to be conformally flat, i.e., $g_{ij} = f_{ij}, f =$ flat 3-metric. (Unfortunately, the techniques described here do not readily adapt to the most general 3-metric.) Then the initial data reduce to φ and K_{ij}. Since the symmetry conditions (4) are applicable only to the physical data ($\bar{g}_{ij}$ and $\bar{K}_{ij}$), the corresponding conditions on φ and K_{ij} must be derived from the relations $\bar{g}_{ij} = \varphi^4 f_{ij}$ and $\bar{K}_{ij} = \varphi^{-2}K_{ij}$.[8,10] For $\bar{g}_{ij}$, the conditions are

$$\begin{aligned}\bar{g}_{ij}(x) &= \pm J_\alpha^*[\bar{g}_{pq}(J_\alpha x)] \\ &= \pm\left[\frac{\partial}{\partial x^i}(J_\alpha x)^p\right]\left[\frac{\partial}{\partial x^j}(J_\alpha x)^q\right]\bar{g}_{pq}(J_\alpha x).\end{aligned} \tag{7}$$

From Eqs. (1) and (2) in Cartesian coordinates, it is seen that

$$\begin{aligned}(\partial/\partial x^i)(J_\alpha x)^p &= (a_\alpha/r_\alpha)^2(\delta_{ip} - 2n_\alpha^i n_\alpha^p) \\ &\equiv (a_\alpha/r_\alpha)^2(R_x^\alpha)_{ip}.\end{aligned} \tag{8}$$

Note that the matrices R_x^α are symmetric and orthogonal. The subscript x in R_x^α specifies that R^α are to be evaluated at the point x. Then substitution of $\bar{g}_{ij} = \varphi^4\delta_{ij}$ in (7) gives

$$\begin{aligned}\delta_{ij}\varphi^4(x) &= \pm(a_\alpha/r_\alpha)^4(R_x^\alpha)_{ip}(R_x^\alpha)_{jq}\delta_{pq}\varphi^4(J_\alpha x) \\ &= \pm\delta_{ij}(a_\alpha/r_\alpha)^4\varphi^4(J_\alpha x).\end{aligned}$$

It is now clear that the plus sign must be chosen in (7), for otherwise φ would vanish. Then the symmetry conditions on φ read: for every $x \in X$ and $\alpha = 1, ..., N$,

$$\varphi(x) = \pm(a_\alpha/r_\alpha)\varphi(J_\alpha x). \tag{9}$$

Here too, the minus sign is ruled out since it leads to $\varphi \equiv 0$ on the bridges S_α (the $r_\alpha = a_\alpha$ surfaces).

Next I consider $\bar{K}_{ij} = \varphi^{-2}K_{ij}$. Following the same steps as above, and using Eq. (9) with a plus sign gives

$$K_{ij}(x) = \pm(a_\alpha/r_\alpha)^6(R_x^\alpha)_{ip}(R_x^\alpha)_{jq}K_{pq}(J_\alpha x).$$

These conditions can be expressed concisely in matrix notation as

$$K(x) = \pm(a_\alpha/r_\alpha)^6 R_x^\alpha K(J_\alpha x) R_x^\alpha \tag{10}$$

for every $x \in X$ and $\alpha = 1, ..., N$.

Thus the initial-value problem on M_N consists of finding φ and K_{ij} on X' satisfying the constraint equations (5) and (6) and symmetry conditions (9) and (10).

III. EXISTENCE OF SOLUTIONS OF THE MOMENTUM CONSTRAINTS

In this section, I shall give a constructive method of finding solutions (K_{ij}) of the momentum constraints having the symmetry (10). Recall that K_{ij}'s are taken to be tracefree and the 3-metric is assumed to be conformally flat. Then, in Cartesian coordinates, the momentum constraints (5) take the form

$$K_{ij,j} = 0. \tag{11}$$

The matrices R^{α}_{x} ($\alpha = 1, ..., N$) are now used to define operators $\mathfrak{R}_\alpha$, in terms of which K_{ij}'s satisfying (11) and (10) will be constructed. The operators $\mathfrak{R}_\alpha$ are defined as follows: Let $M_{ij}(x)$ be a tensor field on E^3. Then

$$(\mathfrak{R}_\alpha M)(x) = \pm (a_\alpha/r_\alpha)^6 R^{\alpha}_{x} M(J_\alpha x) R^{\alpha}_{x}. \tag{12}$$

These operators have the following four properties.

(1) $\mathfrak{R}_\alpha \mathfrak{R}_\alpha = \mathfrak{I}$, the identity operator (no sum over α).

Proof:

$$\begin{aligned}[\mathfrak{R}_\alpha(\mathfrak{R}_\alpha M)](x) &= \pm (a_\alpha/r_\alpha)^6 R^{\alpha}_{x}[(\mathfrak{R}_\alpha M)(J_\alpha x)] R^{\alpha}_{x} \\ &= \pm (a_\alpha/r_\alpha)^6 R^{\alpha}_{x}[\pm (r_\alpha/a_\alpha)^6 \\ &\quad \times R^{\alpha}_{J_\alpha x} M(x) R^{\alpha}_{J_\alpha x}] R^{\alpha}_{x} \\ &= M(x),\end{aligned}$$

where the facts J^2_α = identity, $R^{\alpha}_{x} = R^{\alpha}_{J_\alpha x}$, and $(R^{\alpha}_{x})^2 = \mathrm{I}$ have been used.

(2) $(\mathfrak{R}_\alpha M)^t = \mathfrak{R}_\alpha(M^t)$, where M^t is the transpose of M.

Proof:

$$\begin{aligned}(\mathfrak{R}_\alpha M)^t(x) &= \pm (a_\alpha/r_\alpha)^6 [R^{\alpha}_{x} M(J_\alpha x) R^{\alpha}_{x}]^t \\ &= \pm (a_\alpha/r_\alpha)^6 R^{\alpha}_{x} M^t(J_\alpha x) R^{\alpha}_{x} \\ &\qquad\qquad (\text{since } (R^{\alpha}_{x})^t = R^{\alpha}_{x}) \\ &= [\mathfrak{R}_\alpha(M^t)](x).\end{aligned}$$

(3) $[\mathrm{tr}(\mathfrak{R}_\alpha M)](x) = \pm (a_\alpha/r_\alpha)^6 [(\mathrm{tr}\, M)(J_\alpha x)]$.

Proof:

$$\begin{aligned}[\mathrm{tr}(\mathfrak{R}_\alpha M)](x) &= \pm (a_\alpha/r_\alpha)^6 \mathrm{tr}[R^{\alpha}_{x} M(J_\alpha x) R^{\alpha}_{x}] \\ &= \pm (a_\alpha/r_\alpha)^6 \mathrm{tr}[M(J_\alpha x) R^{\alpha}_{x} R^{\alpha}_{x}] \\ &= \pm (a_\alpha/r_\alpha)^6 [(\mathrm{tr}\, M)(J_\alpha x)].\end{aligned}$$

(4) $$\begin{aligned}[(\mathfrak{R}_\alpha M)_{ij,j}](x) &= \pm (a_\alpha/r_\alpha)^8 (R^{\alpha}_{x})_{ip} [M_{pq,q}(J_\alpha x)] \\ &\quad \mp 2(a^6_\alpha/r^7_\alpha) n^i_\alpha [(\mathrm{tr}\, M)(J_\alpha x)].\end{aligned}$$

Proof:

$$\begin{aligned}&[(\mathfrak{R}_\alpha M)_{ij,j}](x) \\ &= \pm (\partial/\partial x^j)[(a_\alpha/r_\alpha)^6 (R^{\alpha}_{x})_{ip} M_{pq}(J_\alpha x)(R^{\alpha}_{x})_{qj}] \\ &= \pm M_{pq}(J_\alpha x)(\partial/\partial x^j)[(a_\alpha/r_\alpha)^6 (R^{\alpha}_{x})_{ip}(R^{\alpha}_{x})_{qj}] \\ &\quad \pm (a_\alpha/r_\alpha)^6 (R^{\alpha}_{x})_{ip}(R^{\alpha}_{x})_{qj} [(\partial/\partial x^j)(J_\alpha x)^s] M_{pq,s}(J_\alpha x) \\ &= \mp 2(a^6_\alpha/r^7_\alpha) n^i_\alpha [(\mathrm{tr}\, M)(J_\alpha x)] \\ &\quad \pm (a_\alpha/r_\alpha)^8 (R^{\alpha}_{x})_{ip} [M_{pq,q}(J_\alpha x)].\end{aligned}$$

From the last three properties it follows that, if $M_{ij}(x)$ is symmetric and tracefree, and if $M_{ij,j} = 0$, then $(\mathfrak{R}_\alpha M)$ is also symmetric and tracefree, and $(\mathfrak{R}_\alpha M)_{ij,j} = 0$. Thus if $M_{ij}(x)$ is symmetric and tracefree, and satisfies the momentum constraints (11), then for any combination of $\mathfrak{R}_\alpha$'s, say $\mathfrak{R}_\beta ... \mathfrak{R}_\lambda$, $(\mathfrak{R}_\beta ... \mathfrak{R}_\lambda M)$ is also symmetric and tracefree, and satisfied the momentum constraints.

A combination $\mathfrak{R}$ of $\mathfrak{R}_\alpha$'s can now be constructed such that $\mathfrak{R}M$ satisfies the symmetry conditions (10). First express these conditions as $K(x) = (\mathfrak{R}_\alpha K)(x)$ for all $\alpha = 1, ..., N$. Then $\mathfrak{R}M$ must satisfy $\mathfrak{R}M = \mathfrak{R}_\alpha(\mathfrak{R}M)$ for all α. So it suffices to find an $\mathfrak{R}$ such that $\mathfrak{R} = \mathfrak{R}_\alpha \mathfrak{R}$ for all α. From the fact $\mathfrak{R}_\alpha \mathfrak{R}_\alpha = \mathfrak{I}$, it is seen that the following definition of $\mathfrak{R}$ has that property:

$$\mathfrak{R} = \mathfrak{I} + \sum_{\{\alpha_i\}} \left(\prod_{i=1}^{m} \mathfrak{R}_{\alpha_i} \right), \tag{13}$$

where each index α_i can take any of the values $1, ..., N$, and the sum $\Sigma_{\{\alpha_i\}}$ extends over all finite sequences $\{\alpha_1, ..., \alpha_m\}$ of length $m = 1, 2, 3, ...$, subject to the restriction $\alpha_{i+i} \neq \alpha_i$.

To clarify the notation, I shall give explicit forms of $\mathfrak{R}$ in three special cases. First, if only one particle is present, i.e., for $N = 1$,

$$\mathfrak{R} = \mathfrak{I} + \mathfrak{R}_1.$$

For the two-particle ($N = 2$) case,

$$\begin{aligned}\mathfrak{R} = \mathfrak{I} &+ \mathfrak{R}_1 + \mathfrak{R}_1\mathfrak{R}_2 + \mathfrak{R}_1\mathfrak{R}_2\mathfrak{R}_1 + \cdots \\ &+ \mathfrak{R}_2 + \mathfrak{R}_2\mathfrak{R}_1 + \mathfrak{R}_2\mathfrak{R}_1\mathfrak{R}_2 + \cdots.\end{aligned}$$

For the three-particle ($N = 3$) case,

$$\begin{aligned}\mathfrak{R} = \mathfrak{I} &+ \mathfrak{R}_1 + \mathfrak{R}_1(\mathfrak{R}_2 + \mathfrak{R}_3) + \mathfrak{R}_1\mathfrak{R}_2(\mathfrak{R}_3 \\ &+ \mathfrak{R}_1) + \mathfrak{R}_1\mathfrak{R}_3(\mathfrak{R}_1 + \mathfrak{R}_2) + \cdots \\ &+ \mathfrak{R}_2 + \mathfrak{R}_2(\mathfrak{R}_3 + \mathfrak{R}_1) + \mathfrak{R}_2\mathfrak{R}_3(\mathfrak{R}_1 \\ &+ \mathfrak{R}_2) + \mathfrak{R}_2\mathfrak{R}_1(\mathfrak{R}_2 + \mathfrak{R}_3) + \cdots \\ &+ \mathfrak{R}_3 + \mathfrak{R}_3(\mathfrak{R}_1 + \mathfrak{R}_2) + \mathfrak{R}_3\mathfrak{R}_1(\mathfrak{R}_2 \\ &+ \mathfrak{R}_3) + \mathfrak{R}_3\mathfrak{R}_2(\mathfrak{R}_3 + \mathfrak{R}_1) + \cdots.\end{aligned}$$

Thus if M is a solution of the momentum constraints, so will $\mathfrak{R}M$ be. In addition, $\mathfrak{R}M$ will also satisfy the symmetry conditions (10). However, it must be shown that $\mathfrak{R}M$ is convergent on X'. To study this problem, I assume M to be a symmetric and tracefree solution of the momentum constraints, and to be of the form

$$M = \sum_{b=1}^{B} \sum_{\lambda=1}^{N} r_\lambda^{-p_b} M^{(b)}_\lambda, \tag{14}$$

where $p_b \geqslant 0$ are constants, and the matrices $M^{(b)}_\lambda$ are smooth functions in the region $E^3 - \cup^N_{\alpha=1}\{c_\alpha\}$ with the boundedness property

$$\sum_{i,j=1}^{3} [(M^{(b)}_\lambda(x))_{ij}]^2 \equiv \mathrm{tr}(M^{(b)}_\lambda)^2(x) < A^2 < \infty$$

for every $x \in E^3 - \cup^N_{\alpha=1}\{c_\alpha\}$, and for all λ's and b's.

I write the operator $\mathfrak{R}$ as a sum of factors of a given order:

$$\mathfrak{R} = \mathfrak{I} + \mathfrak{S}_1 + \mathfrak{S}_2 + \mathfrak{S}_3 + \cdots, \tag{15}$$

where $\mathfrak{S}_1 = \Sigma\mathfrak{R}_{\alpha_1}$, $\mathfrak{S}_2 = \Sigma'\mathfrak{R}_{\alpha_2}\mathfrak{R}_{\alpha_1}$, $\mathfrak{S}_3 = \Sigma'\mathfrak{R}_{\alpha_3}\mathfrak{R}_{\alpha_2}\mathfrak{R}_{\alpha_1}$, ..., and Σ' is the sum over all α_i's with the restriction $\alpha_{i+1} \neq \alpha_i$. To determine the domain of convergence of $\mathfrak{R}M$, I shall first examine each $\mathfrak{S}_m M$, ($m = 1, 2, 3, ...$), and then the sum $(\mathfrak{I} + \mathfrak{S}_1 + \mathfrak{S}_2 + \cdots)M$.

I start with $\mathfrak{S}_4$, which consists of $N(N-1)^3$ terms of the form $\mathfrak{R}_{\alpha_4}\mathfrak{R}_{\alpha_3}\mathfrak{R}_{\alpha_2}\mathfrak{R}_{\alpha_1}$ with $\alpha_4 \neq \alpha_3$, $\alpha_3 \neq \alpha_2$, and $\alpha_2 \neq \alpha_1$.

These are the simplest terms that illustrate the arguments to be used for subsequent higher-order terms. Now consider some particular term in $\mathfrak{S}_4$, say $\mathfrak{R}_\nu \mathfrak{R}_\mu \mathfrak{R}_\beta \mathfrak{R}_\alpha$, with $\nu \neq \mu$, $\mu \neq \beta$, $\beta \neq \alpha$; ν, μ, β, α fixed. From Definition (12) of $\mathfrak{R}_\alpha$, for any $x \in X'$,

$$\begin{aligned}[t][\mathfrak{R}_\nu \mathfrak{R}_\mu \mathfrak{R}_\beta \mathfrak{R}_\alpha M](x) &= \pm (a_\nu / r_\nu)^6 R^{\nu}_{x} [\mathfrak{R}_\mu \mathfrak{R}_\beta \mathfrak{R}_\alpha M](J_\nu x)(R^{\nu}_{x})' \\ &= (a_\nu/r_\nu)^6_x (a_\mu/r_\mu)^6_{J_\nu x} R^{\nu}_{x} R^{\mu}_{J_\nu x} \\ &\quad \times [\mathfrak{R}_\beta \mathfrak{R}_\alpha M](J_{\mu\nu} x)(R^{\nu}_{x} R^{\mu}_{J_\nu x})',\end{aligned}$$

where $J_{\mu\nu} x \equiv J_\mu (J_\nu x)$ and $R^{\mu}_{J_\nu x}$ means R^{μ} evaluated at $J_\nu x$. Similarly, $(a_\mu / r_\mu)^6_{J_\nu x} = a_\mu^6 / |J_\nu x - c_\mu|^6$. I continue this procedure and write

$$R \equiv R^{\nu}_{x} R^{\mu}_{J_\nu x} R^{\beta}_{J_{\mu\nu} x} R^{\alpha}_{J_{\beta\mu\nu} x}, \tag{16}$$

with the result

$$\begin{aligned}[t][\mathfrak{R}_\nu \mathfrak{R}_\mu \mathfrak{R}_\beta \mathfrak{R}_\alpha M](x) = \sum_{b=1}^{B} \sum_{\lambda=1}^{N} & (a_\nu/r_\nu)^6_x (a_\mu/r_\mu)^6_{J_\nu x} \\ & \times (a_\beta/r_\beta)^6_{J_{\mu\nu} x} (a_\alpha/r_\alpha)^6_{J_{\beta\mu\nu} x} \\ & \times (r_\lambda)^{-p_b}_{J_{\alpha\beta\mu\nu} x} R M^{(b)}_{\lambda} (J_{\alpha\beta\mu\nu} x) R'.\end{aligned} \tag{17}$$

First consider the coefficient

$$(a_\nu/r_\nu)^6_x (a_\mu/r_\mu)^6_{J_\nu x} (a_\beta/r_\beta)^6_{J_{\mu\nu} x} (a_\alpha/r_\alpha)^6_{J_{\beta\mu\nu} x}.$$

Let

$$D_\lambda = \{p \in E^3 : 0 < |p - c_\lambda| < a_\lambda\}, \quad \lambda = 1,\ldots,N. \tag{18}$$

Then the closure of D_λ is $\bar{D}_\lambda = D_\lambda \cup S_\lambda \cup \{c_\lambda\}$. Now fix some $x \in X'$. Either $|x - c_\nu| \geqslant a_\nu$ or $|x - c_\nu| < a_\nu$. If $|x - c_\nu| \geqslant a_\nu$, then $J_\nu x \in \bar{D}_\nu$. Hence $|J_\nu x - c_\mu| > a_\mu$ since $\nu \neq \mu$; i.e., $(a_\mu / r_\mu)_{J_\nu x} < 1$. Furthermore, since $\mu \neq \beta$ and $\beta \neq \alpha$, $J_{\mu\nu} x \in J_\mu \bar{D}_\nu \subset \bar{D}_\nu$, $J_{\beta\mu\nu} x \in J_\beta \bar{D}_\mu$, and $J_{\alpha\beta\mu\nu} x \in J_\alpha \bar{D}_\beta$ (Fig. 3). Similarly, if $|x - c_\nu| < a_\nu$, then $x \in I_\nu = J_\nu[X]$ since $x \in X'$. Hence $J_\nu x \in X$, and thus $|J_\nu x - c_\mu| > a_\mu$; i.e., $(a_\mu/r_\mu)_{J_\nu x} < 1$. Furthermore, $J_{\mu\nu} x \in I_\mu \subset \bar{D}_\mu$, $J_{\beta\mu\nu} x \in J_\beta \bar{D}_\mu$, and $J_{\alpha\beta\mu\nu} x \in J_\alpha \bar{D}_\beta$. Thus for any $x \in X'$, it follows that $(a_\mu/r_\mu)_{J_\nu x} < 1$, $J_{\mu\nu} x \in \bar{D}_\mu$, $J_{\beta\mu\nu} x \in J_\beta \bar{D}_\mu$, and $J_{\alpha\beta\mu\nu} x \in J_\alpha \bar{D}_\beta$.

Now observe that since the spheres S_λ ($\lambda = 1, \ldots, N$) are nonintersecting, for any $\sigma, \tau, \sigma \neq \tau$,

$$d_{\sigma\tau} \equiv |c_\sigma - c_\tau| - (a_\sigma + a_\tau) > 0. \tag{19}$$

($d_{\sigma\tau}$ is the smallest distance between the surfaces of the spheres S_σ and S_τ.) Then, for any $p \in \bar{D}_\mu$ and for $\beta \neq \mu$,

$$\frac{a_\beta}{|p - c_\beta|} \leqslant \frac{a_\beta}{|c_\mu - c_\beta| - a_\mu} = \frac{1}{1 + (d_{\mu\beta}/a_\beta)} \leqslant \epsilon < 1,$$

where

$$\epsilon = \max_{\substack{\sigma,\tau \\ \sigma \neq \tau}} [1 + (d_{\sigma\tau}/a_\tau)]^{-1} = \left[1 + \min_{\substack{\sigma,\tau \\ \sigma \neq \tau}} (d_{\sigma\tau}/a_\tau)\right]^{-1}. \tag{20}$$

Here $(a_\beta / r_\beta)_{J_{\mu\nu} x} = \epsilon < 1$, since $J_{\mu\nu} x \in \bar{D}_\mu$ and $\mu \neq \beta$. Similarly, $(a_\alpha/r_\alpha)_{J_{\beta\mu\nu} x} \leqslant \epsilon < 1$.

Since X' is an open set and $c_\lambda \notin X'$ ($\lambda = 1, \ldots, N$), there exists a constant $\rho > 0$ such that $|x - c_\lambda| > \rho$ for all λ's. (ρ will, of course, depend on x.) In view of these inequalities, let $a = \max_\lambda a_\lambda$, so that

$$(a_\nu/r_\nu)^6_x (a_\mu/r_\mu)^6_{J_\nu x} (a_\beta/r_\beta)^6_{J_{\mu\nu} x} (a_\alpha/r_\alpha)^6_{J_{\beta\mu\nu} x} \leqslant (a/\rho)^6 . 1 . (\epsilon^6)^2 < \infty. \tag{21}$$

To obtain a bound on the term $(r_\lambda)^{-p_b}_{J_{\alpha\beta\mu\nu} x}$, note that for any $p \in J_\sigma \bar{D}_\tau$, $\sigma \neq \tau$, $|p - c_\lambda| > 0$ for all λ (Fig. 3). In fact,

$$\epsilon_1 \equiv \min\{|p - c_\lambda| : p \in J_\sigma \bar{D}_\tau, \quad \sigma \neq \tau, \quad 1 \leqslant \sigma, \quad \tau,\lambda \leqslant N\} > 0.$$

Since $J_{\alpha\beta\mu\nu} x \in J_\alpha \bar{D}_\beta$ for every $x \in X'$,

$$(r_\lambda)^{-p_b}_{J_{\alpha\beta\mu\nu} x} \leqslant \epsilon_1^{-p_b} \leqslant \max_{b'} (\epsilon_1^{-p_{b'}}) \equiv \epsilon' < \infty. \tag{22}$$

A bound on the matrix part of (17) can be obtained as follows: Since $(R^{\alpha}_{p})^2 = I$ and $\mathrm{tr}(M^{(b)}_{\lambda})^2(p) < A^2 < \infty$ for every $p \in E^3 - \cup_{\alpha=1}^{N} \{c_\alpha\}$, it is seen that

$$\mathrm{tr}[R M^{(b)}_{\lambda} (J_{\alpha\beta\mu\nu} x) R']^2 = \mathrm{tr}(M^{(b)}_{\lambda})^2 (J_{\alpha\beta\mu\nu} x) < A^2.$$

Hence, for every $x \in X'$ and for all i, j's,

$$|[R M^{(b)}_{\lambda} (J_{\alpha\beta\mu\nu} x) R']_{ij}| < A < \infty. \tag{23}$$

The inequalities (21)–(23) are used in (17) to get

$$|[\mathfrak{R}_\nu \mathfrak{R}_\mu \mathfrak{R}_\beta \mathfrak{R}_\alpha M]_{ij}(x)| \leqslant BNA\epsilon'(a/\rho)^6 (\epsilon^6)^2 < \infty,$$

where BN is the number of terms in (14).

The extension of this result to the higher-order terms is immediate. For the general term in $\mathfrak{S}_m$, it is seen that

$$|[\mathfrak{R}_{\alpha_m} \cdots \mathfrak{R}_{\alpha_1} M]_{ij}(x)| \leqslant BNA\epsilon'(a/\rho)^6 (\epsilon^6)^{m-2} < \infty.$$

Thus every term in the series $\mathfrak{R}M$ is finite on X'. But $\mathfrak{S}_m = \Sigma' \mathfrak{R}_{\alpha_m} \cdots R_{\alpha_1}$ consists of $N(N-1)^{m-1}$ terms (since $\alpha_{i+1} \neq \alpha_i$). Hence

$$|[\mathfrak{S}_m M]_{ij}(x)| \leqslant ABN^2(N-1)\epsilon'(a/\rho)^6 [(N-1)\epsilon^6]^{m-2}.$$

From this inequality and the expression (15) for $\mathfrak{R}$, it follows that the condition $(N-1)\epsilon^6 < 1$ guarantees the absolute convergence of $\mathfrak{R}M$. Also, if the series $\mathfrak{R}M$ is truncated beyond the first $K+1$ terms, the remainder is seen to be smaller than

$$\begin{aligned}[t]\sum_{m=K+1}^{\infty} & ABN^2(N-1)\epsilon'(a/\rho)^6 [(N-1)\epsilon^6]^{m-2} \\ &= ABN^2(N-1)\epsilon'(a/\rho)^6 \frac{[(N-1)\epsilon^6]^{K-1}}{1-(N-1)\epsilon^6}.\end{aligned}$$

This bound is an estimate of truncation error and should be kept in mind when the series is used in numerical work.

In view of Definition (20) of ϵ, the convergence condition can be reformulated as

$$\min_{\substack{\sigma,\tau \\ \sigma \neq \tau}} (d_{\sigma\tau}/a_\tau) > (N-1)^{1/6} - 1, \tag{24}$$

where $d_{\sigma\tau}$ is the distance between the surfaces of the spheres S_σ and S_τ. Note that for $N = 2$, (24) is automatically satisfied (since all $d_{\sigma\tau} > 0$). Hence for the problem of two black holes, the series $\mathfrak{R}M$ is always absolutely convergent. For $N = 3$, $\min(d_{\sigma\tau}/a_\tau) > (N-1)^{1/6} - 1 = .1225$ guarantees the absolute convergence of $\mathfrak{R}M$. In particular, if $a_1 = a_2 = a_3$, then the condition reduces to $\min(d_{\sigma\tau}) > a_1/8.166$. For $N = 4$, 65, and 1000, the corresponding values of $(N-)^{1/6} - 1$ are 0.201, 1, and 2.162, respectively. The rate of growth of $(N-1)^{1/6} - 1$ is quite slow, and, thus, the convergence requirement (24) is not too restrictive.

IV. APPLICATIONS AND THE CONFORMAL FACTOR

So far I have not made any specific choice of $M^{(b)}_{\lambda}$'s and p_b's in M. To represent black holes with arbitrary spins and momenta, the York solution[12] (in E^3) of the momentum constraints may be used:

$$M=\sum_{k=2}^{3}\sum_{\alpha=1}^{N} r_\alpha^{-k} M^{(k)}_{\alpha}, \tag{25a}$$

where

$$M^{(2)}_{\alpha}=\tfrac{3}{2}[P_\alpha n^t_\alpha+n_\alpha P^t_\alpha+(n_\alpha n^t_\alpha-I)(P^t_\alpha n_\alpha)] \tag{25b}$$

and

$$M^{(3)}_{\alpha}=-3[n_\alpha(S_\alpha n_\alpha)^t+(S_\alpha n_\alpha)n^t_\alpha]. \tag{25c}$$

P_α and S_α are constant vectors and antisymmetric matrices, respectively, and $(S_\alpha n_\alpha)_i\equiv(S_\alpha)_{ij}n^j_\alpha$. This solution represents N particles at sites c_α ($\alpha=1,\ldots,N$) with linear and (spin) angular momenta P^i_α and $Y^i_\alpha=\frac{1}{2}\epsilon_{ijk}(S_\alpha)_{jk}$. Thus $S_\alpha n_\alpha=n_\alpha\times Y_\alpha$, the cross product of the vectors n_α and Y_α. Then $M^{(3)}_{\alpha}$ takes the form

$$M^{(3)}_{\alpha}=-3[n_\alpha(n_\alpha\times Y_\alpha)^t+(n_\alpha\times Y_\alpha)n^t_\alpha]. \tag{25d}$$

This choice of M does have the same form as (14), and

$$\mathrm{tr}(M^{(2)}_{\alpha})^2=(9/2)[|P_\alpha|^2+2(n^t_\alpha P_\alpha)^2]\leqslant(27/2)|P_\alpha|^2,$$

$$\mathrm{tr}(M^{(3)}_{\alpha})^2=18[|Y_\alpha|^2-(n^t_\alpha Y_\alpha)^2]\leqslant 18|Y_\alpha|^2.$$

Hence

$$\mathrm{tr}(M^{(k)}_{\alpha})^2\leqslant A^2\equiv\max_{\lambda}[(27/2)|P_\lambda|^2,\quad 18|Y_\lambda|^2]<\infty. \tag{26}$$

Then, provided condition (24) is satisfied, the conclusion is that $\mathfrak{R}M$ is a smooth solution of the momentum constraints for the problem of N black holes with arbitrary spins and momenta.

As a special case, let $N=1$ and $S_1=0$ in (25); that is, $M=(3/2r_1^2)[P_1n^t_1+n_1P^t_1+(n_1n^t_1-I)(P^t_1n_1)]$. For $N=1$, the operator $\mathfrak{R}$ is $\mathfrak{I}+\mathfrak{R}_1$. Thus

$$\mathfrak{R}M=(\mathfrak{I}+\mathfrak{R}_1)M=M\mp(3a_1^2/2r_1^4)\times[P_1n^t_1+n_1P^t_1+(I-5n_1n^t_1)(P^t_1n_1)].$$

This solution is the same as the one-body solution (9) of Ref. 8. If the minus sign is taken, it becomes the Bowen solution,[10,11] which was found from a vector potential as in Ref. 12. An explicit calculation of $K=\mathfrak{R}M$ for the problem of two black holes and a discussion of its physical properties will be presented elsewhere.

The conformal factor $\varphi: X'\to\mathbb{R}$ ($\mathbb{R}$ is the set of all real numbers) must satisfy the symmetry conditions (9) as well as the Hamiltonian constraint (6) with $K_{ij}=(\mathfrak{R}M)_{ij}$. Furthermore, φ should be smooth, positive and asymptotically equal to unity. For the case of a single black hole, Bowen and York[8] have formulated a boundary-value problem on $\bar{X}$ (the region $r_1\geqslant a_1$) which gives the same result and is convenient for obtaining numerical solutions. Its generalization to the problem of N black holes is immediate. But I will out line it here for the sake of completeness.

One starts from the Hamiltonian constraint (6) on X' and sets up a boundary-value problem on $\bar{X}$. Its solution is extended to yield φ on X' with the desired properties: Note that the spherical polar coordinates $(r_\alpha,\theta_\alpha,\phi_\alpha)$ with their origin at c_α can be used to express the symmetry condition (9) as

$$\varphi(r_\alpha,\theta_\alpha,\phi_\alpha)=(a_\alpha/r_\alpha)\varphi(a_\alpha^2/r_\alpha,\theta_\alpha,\phi_\alpha). \tag{27}$$

Differentiating this equation with respect to r_α and setting $r_\alpha=a_\alpha$ yields $[(\partial\varphi/\partial r_\alpha)+(2a_\alpha)^{-1}\varphi]_{r_\alpha=a_\alpha}=0$. These equations are used as the inner boundary conditions in the fomulation of the following boundary-value problem on $\bar{X}$:

$$\left.\begin{aligned}&8(\Delta f)=-K_{ij}K^{ij}f^{-7},\quad f>0,\\ &\lim_{r_\alpha\to\infty}f=1\\ &\text{and}\\ &[(\partial f/\partial r_\alpha)+(2a_\alpha)^{-1}f]_{r_\alpha=a_\alpha}=0\quad\text{for all }\alpha.\end{aligned}\right\} \tag{28}$$

Conditions which ensure the existence and uniqueness of solutions have been discussed by Bowen and York[8] and York[13]. The inner boundary conditions show that the bridges S_α are extremal area surfaces. Explicit calculations over a wide range of momenta and spins have shown that surfaces to be minimal for the one-body problem,[14] and it would be interesting to show that they are minimal in general. The apparent horizons[15] associated with the data will not, in general, coincide with the extremal surfaces S_α, but will lie outside them, each with an image on the other sheet. This topic is treated for the one-hole problem in Bowen and York.[8]

To extend the solution f of (28) to X', first define $g:\cup_{\alpha=1}^{N}J_\alpha[\bar{X}]\to\mathbb{R}$ by

$$g(J_\alpha x)=(r_\alpha/a_\alpha)f(x) \tag{29}$$

for every $x\in\bar{X}$ and $\alpha=1,\ldots,N$. Note that $g=f$ on $\cup_{\alpha=1}^{N}S_\alpha$. Also, g satisfies the Hamiltonian constraint on $\cup_{\alpha=1}^{N}J_\alpha[X]$: Choose some point $(J_\alpha x)\in\cup_{\lambda=1}^{N}J_\lambda[X]$. Then, Eq. (8) implies

$$(\partial/\partial x^j)[g(J_\alpha x)]=(a_\alpha/r_\alpha)^2(R^{\alpha}_{x})_{ji}[\partial g(\xi)/\partial\xi^i]_{\xi=J_\alpha x}.$$

Since $R^{\alpha}_{x}R^{\alpha}_{x}=I$, it follows that

$$[\partial g(\xi)/\partial\xi^i]_{J_\alpha x}=(r_\alpha/a_\alpha)^2(R^{\alpha}_{x})_{ij}(\partial/\partial x^j)[g(J_\alpha x)].$$

From this equation, one obtains

$$[\Delta g]_{J_\alpha x}=(r_\alpha/a_\alpha)^4\Delta[g(J_\alpha x)]-2(r_\alpha^3/a_\alpha^4)n^i_\alpha(\partial/\partial x^i)[g(J_\alpha x)].$$

Thus, from Definition (29) of g, one finds

$$[\Delta g]_{J_\alpha x}=(r_\alpha/a_\alpha)^5[\Delta f]_x.$$

Since f satisfies the Hamiltonian constraint, and since the K_{ij}'s have the symmetry property (10), it is seen that

$$\begin{aligned}8[\Delta g]_{J_\alpha x}&=8(r_\alpha/a_\alpha)^5[\Delta f]_x\\&=-(r_\alpha/a_\alpha)^5[K_{ij}K^{ij}](x)f^{-7}(x)\\&=-(r_\alpha/a_\alpha)^5(a_\alpha/r_\alpha)^{12}[K_{ij}K^{ij}](J_\alpha x)\\&\quad\times(a_\alpha/r_\alpha)^{-7}g^{-7}(J_\alpha x)\\&=-[K_{ij}K^{ij}g^{-7}]_{J_\alpha x}.\end{aligned}$$

Thus g satisfies the Hamiltonian constraint on $\cup_{\alpha=1}^{N}J_\alpha[X]$.

Hence define $\varphi: X'\to\mathbb{R}$ by

$$\varphi(x)=\begin{cases}g(x)&\text{if}\quad x\in\bar{X},\\ g(x)&\text{if}\quad x\in\cup_{\alpha=1}^{N}J_\alpha[X].\end{cases} \tag{30}$$

Then φ satisfies the Hamiltonian constraint on X' and also has the desired symmetry (9): for $x \in X$,

$$\varphi(x) = f(x) = (a_\alpha / r_\alpha) g(J_\alpha x) = (a_\alpha / r_\alpha)\varphi(J_\alpha x).$$

Thus the problem of obtaining the conformal factor φ on X' is equivalent to solving the boundary-value problem (28) and extending its solution to X' using (29) and (30). The advantages of this formulation are that the symmetry condition (27) is automatically satisfied and the boundary-value problem (28) is convenient for obtaining numerical solutions. Such solutions have been obtained in the case of one black hole by York and Piran,[14] who also, in this case, determine the location of apparent horizon and give estimates of the gravitational radiation that could be produced by the hole in its evolution.

From Definition (29) of g expressed in spherical polar coordinates $(r_\alpha, \theta_\alpha, \phi_\alpha)$, it follows that the inner boundary condition in (28) is equivalent to the requirement $\partial f / \partial r_\alpha = \partial g / \partial r_\alpha$ at $r_\alpha = a_\alpha$. Thus the functions f and g, as well as their first radial derivatives, match on the boundary $\cup_{\alpha=1}^{N} S_\alpha$.

V. CONCLUSIONS

I have examined the time-asymmetric initial-value problem for N black holes. The spacelike hypersurface for setting up the initial data is taken to be the Einstein–Rosen manifold with N bridges. Thus the black holes are modeled as bridges between two asymptotically flat universes (sheets). I assume the initial 3-geometry to be conformally flat. Starting from any solution of the momentum constraints in three-dimensional Euclidean space, I have given an explicit procedure to construct solutions of the momentum constraints that preserve the symmetry between the two sheets of the Einstein–Rosen manifold. These solutions are in the form of infinite series. I have obtained a sufficient condition (which is not very restrictive) for the absolute convergence of a wide range of these solutions. This analysis also gives an estimate of the error involved in truncating these series. The problem of specifically representing N black holes with arbitrary spins and momenta is discussed. Finally, for completeness, the extension to the case of N black holes, of the formulation of the Hamiltonian constraint as a boundary-value problem is presented.

The symmetry between the two sheets of the Einstein–Rosen manifold, apart from being of aesthetic value, is also of practical importance, for the data on one sheet can be deduced from that on the other. Thereby it suffices to evolve the data on just one sheet, the evolution of the other sheet being identical to the first. It makes the numerical integration of the evolution equations particularly convenient. Consequently, the initial data representing black holes with linear and angular momenta (spin) can be evolved on the computer to study, with a greater reliability, processes such as the scattering of black holes, the collapse of the orbit of two black holes going around each other, and so on. Since these processes are expected to be strong sources of gravitational waves,[1] their study is also important in the program of the detection of gravitation waves. Thus the formulation of the time-asymmetric data on the Einstein–Rosen manifold makes it possible to study a large class of astrophysically interesting phenomena.

ACKNOWLEDGMENTS

It is a pleasure to thank L. C. Shepley for many stimulating discussions in the course of this work and for critically reading the manuscript. I am also grateful to J. Bowen, M. Cantor, and especially J. W. York, Jr. for helpful remarks and suggestions.

[1]*Sources of gravitational radiation*, edited by L. Smarr (Cambridge U. P., Cambridge, England, 1979).
[2]D. Eardley and L. Smarr, "Time functions in numerical relativity: marginally bound dust collapse," Phys. Rev. D **19**, 2239–2259 (1979); L. Smarr and J. W. York, Jr., "Kinematical conditions in the construction of spacetime," Phys. Rev. D **17**, 2529–2551 (1978).
[3]A. Lichnerowicz, "L'integration des équations de la gravitation relativiste et le problème des n corps," J. Math. Pures Appl. **23**, 37–63 (1944).
[4]A. Einstein and N. Rosen, "The particle problem in the theory of general relativity," Phys. Rev. **48**, 73–77 (1935).
[5]C. W. Misner, "The method of images in geometrostatics," Ann. Phys. (N.Y.) **24**, 102–117 (1963).
[6]R. W. Lindquist, "Initial-value problem on Einstein–Rosen manifolds," J. Math. Phys. **4**, 938–950 (1963).
[7]L. Smarr, "Gauge conditions, radiation formulae and the two black hole collision," Chapter in Ref. 1, pp. 245–274.
[8]J. M. Bowen and J. W. York, Jr., "Time-asymmetric initial data for black holes and black-hole collisions," Phys. Rev. D **21**, 2047–2056 (1980).
[9]A. D. Kulkarni, L. C. Shepley, and J. W. York, Jr., "Initial data for N black holes" (to be published).
[10]J. M. Bowen, "Initial-value problems on non-Euclidean topologies," Ph.D. thesis, University of North Carolina, Chapel Hill (1979).
[11]J. M. Bowen, "General form for the longitudinal momentum of a spherically symmetric source," Gen. Rel. Grav. **11**, 227–231 (1979).
[12]J. W. York, Jr., "Kinematics and dynamics of general relativity, Chapter in Ref. 1, pp. 83–126.
[13]J. W. York, Jr., private communication (1983).
[14]J. W. York, Jr. and T. Piran, "The initial-value problem and beyond," in *Spacetime and geometry*, edited by R. Matzner and L. C. Shepley (University of Texas Press, Austin, 1982), pp. 147–176.
[15]S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time* (Cambridge U. P., Cambridge, England, 1973).

The multipole expansion of stationary Einstein–Maxwell fields[a)]

Walter Simon
Institut für theoretische Physik, Universität Wien, Boltzmanngasse 5, A-1090 Wien, Austria

(Received 15 March 1983; accepted for publication 22 April 1983)

The stationary Einstein–Maxwell equations are rewritten in a form which permits the introduction of (Geroch-) multipole moments for asymptotically flat solutions. Some known theorems on the moments in the stationary case are generalized to include Einstein–Maxwell fields.

PACS numbers: 04.20.Jb

I. INTRODUCTION

Geroch has given a definition of multipole moments for a static, asymptotically flat solution of Einstein's equations.[1] His work was generalized to the stationary case by Hansen[2] and to the electrostatic case by Hoenselaers.[3] The present paper provides a generalization to the stationary Einstein–Maxwell (SEM) case.

Geroch's results and conjectures on this subject have stimulated a lot of interesting work (see, e.g, Refs. 4 and 5) which so far culminated in the proof that a given space-time can be uniquely characterized by its moments.[6–8] There is another paper[9] which shows that "in principle," (i.e., modulo convergence) a space-time with a given set of moments can be found.

Most of the above-mentioned approaches start with different assumptions and use different mathematical techniques to establish their special result; a unified description is lacking.

Moreover, it is not yet seen whether any of these approaches can be wedded with the "Theory of the axially symmetric case" (see Ref. 10 for a review) or with nonstationary expansions near spatial infinity.[11,12]

On a stationary 4-manifold, the timelike Killing vector induces a preferred 3-manifold (the trajectories) of the Killing field), and potentials (the norm and the twist potential[13]). The introduction of multipole moments, as described in Ref. 1, requires the existence of a compactification of this 3-space, such that, after a suitable conformal transformation, the 3-metric and the potentials become analytic in suitable coordinates. For the static and stationary gravitational fields one already knows that a delicate choice of variables must be made to meet these requirements. If moreover, an electromagnetic field is present, one should, of course, look for some "natural" generalization of the stationary variables. What "natural" means in this case is the following.

The variables should be well behaved near spatial infinity for every asymptotically flat or NUT stationary Einstein–Maxwell metric (called AFSEM). It is well known that there are general methods for transforming such solutions into each other.[10] Accordingly, the potentials must be chosen such that the good behavior is preserved under these transformations. If there are potentials that transform linearly, they seem to be favorable candidates, since linear combinations of well-behaved functions are also well behaved.

Of particular interest in the present case are Kinnersley's SU(2,1) transformations,[14] since they preserve a certain 3-metric which can be used as "background geometry."

(The choice of the background is not unique. One might be led to use just this one because of the existence of such a group.) It is, however, not desirable to linearize the whole SU(2,1) since some of their transformations destroy asymptotic flatness. In this paper we arrive at a formulation of the SEM equations where a U(1,1) group acts linearly on the AFSEM solutions. There is not a unique choice of variables with this property. Indeed, since U(1,1) has a quadratic invariant, one can choose suitable functions of the linearizing variables and of this invariant to obtain a new set of linearizing variables. This observation can be used to reduce the algebraic complexity of the field equations. The variables which are used in this paper are new and seem to be simplest. They are members of a one-parameter family of potentials which comprises, as special cases, most of the existing multipole fields.

Having obtained an appropriate form of the field equations, one must choose one of the above-mentioned ways[4–9] to deal with the multipoles. The present work provides the generalization of Ref. 9, enjoying, of course, the particular strengths and weaknesses of this approach. We are, in essence, performing a $1/r$ expansion of the potentials in some suitable coordinates by counting powers of r and the parity of spherical harmonics. The procedure, which is outlined in Sec. III (see Ref. 9 for details), is rather easy to handle and works with a "weak" definition of asymptotic flatness in terms of fall-off conditions on the potentials and on the 3-metric. The drawback of the approach is that we obtain C^k-variables, for arbitrary k, near spatial infinity, and no proof of analyticity. This means that we do not have access to the "uniqueness proofs" as obtained by the more geometric method in Refs. 6–8. There should, however, be no fundamental difficulty in applying also these approaches to the SEM equations in the form of Sec. II.

II. THE FIELD EQUATIONS

Greek indices take the values 0, 1, 2, 3, Latin indices 1, 2, 3. We consider a manifold X, which is topologically $I \times N$ where I is the t axis and N is diffeomorphic to $\mathbf{R}^3$ minus a ball. On X we are given a stationary metric

[a)] Research supported by "Fonds zur Förderung der wissenschaftlichen Forschung in Oesterreich," Project No. 4069.

$$ds^2 = \lambda\,(dt + \sigma_i dx^i)^2 - \lambda^{-1}\gamma_{ij} dx^i dx^j \tag{2.1}$$

and a stationary electromagnetic field $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$, $A_\mu = (e, A_i)$ which satisfy the SEM equations (without sources). λ, σ_i, e, A_i, and γ_{ij} (the metric on N) are smooth functions of x^i on N. Tensor indices will be moved with γ_{ij} and its inverse (γ^{ij}). Indices on coordinates will be moved with $\delta_{ij} = \delta^{ij} = \mathrm{diag}\,(1, 1, 1)$ ($x^i = x_i$). The covariant derivative with respect to γ_{ij} will be denoted by D_i, the covariant Laplace operator by $\Delta(\gamma) = D_i D^i$. The Ricci tensor of γ_{ij} is defined by $D^i D_j \alpha_i - D_j D^i \alpha_i = R_{jk}(\gamma)\alpha^k$, where α_i is an arbitrary vector, and the permutation symbol by $\epsilon_{ijk} = \epsilon_{[ijk]} = 0, \pm |\det \gamma_{ij}|^{1/2}$.

The source-free Maxwell equation implies that the vector field

$$f_i = \lambda \epsilon_i{}^{jk}(D_j A_k + \sigma_j D_k e) \tag{2.2}$$

is curl-free, i.e.,

$$D_{[i} f_{j]} = 0. \tag{2.3}$$

It follows that there exists a scalar field f such that $f_j = D_j f$. In terms of the complex potential $\Psi = e + if$, the source-free Einstein equations imply that

$$\omega_i = -\lambda^2 \epsilon_i{}^{jk} D_j \sigma_k + i(\Psi^* D_i \Psi - \Psi D_i \Psi^*) \tag{2.4}$$

is also curl-free:

$$D_{[i}\omega_{j]} = 0 \tag{2.5}$$

and gives rise to the existence of another field ω ($\omega_i = D_i \omega$). (The detailed calculations leading to (2.2)–(2.5) can be found in Israel and Wilson[15]).

Next, we define the Ernst potential

$$\mathscr{E} = \lambda + i\omega - \Psi\Psi^* \tag{2.6}$$

and the fields v and w by the ansatz

$$\mathscr{E} = \frac{1-w}{1+w}, \quad \Psi = \frac{v}{1+w}. \tag{2.7}$$

These fields become identical with the potentials introduced by Kinnersley[14] if u is put equal to 1 there. Finally, we introduce a complex 2-space whose elements are denoted by Y^A, $A = 1, 2$ with a metric and an ϵ-tensor

$$\eta_{AB} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \epsilon_{AB} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.8}$$

Capital Latin indices are moved with η_{AB}.

The manifold N with metric γ_{ij}, together with the 2-vector $Y^A = [v,w]$ at every point of N, form the arena for the subsequent discussion. We arrive at the following form of the SEM equations:

$$\Delta(\gamma) Y^A = 2\Theta^{-1}(Y^*_C D_i Y^C) D^i Y^A, \tag{2.9}$$

$$R_{ij}(\gamma) = 2\Theta^{-2}(Z_{(i} Z^*_{j)} - D_{(i} Y^C D_{j)} Y^*_C), \tag{2.10}$$

where

$$\Theta = (1 + Y^*_A Y^A) \tag{2.11}$$

and

$$Z_i = \epsilon_{AB} Y^A D_i Y^B. \tag{2.12}$$

These equations can be derived from the Lagrangian

$$\mathscr{L} = |\det \gamma|^{1/2} \gamma^{ij} [R_{ij}(\gamma) - 2\Theta^{-2}(Z_{(i} Z^*_{j)} - D_{(i} Y^C D_{j)} Y^*_C)]. \tag{2.13}$$

The Lagrangian and the field equations are invariant under the replacement

$$Y^{A'} = L^A{}_B Y^B, \tag{2.14}$$

where L is a 2×2 Lorentz matrix with constant coefficients

$$L^+ \eta L = \eta, \quad D_i L \equiv 0. \tag{2.15}$$

This is true because the vector Z_i, whose norm enters the Lagrangian (2.13), transforms as $Z'_i = (\det L) Z_i$. Clearly, the invariance group is U(1,1). The field equations (2.9) and (2.10) are (modulo misprints in Ref. 14) identical with the Kinnersley equations if we set $u = 1$. However, this is an incidence; one must not, in general, impose constraints on the Kinnersley equations.

We write some known solutions in the present variables.

(1) The Kerr solution reads

$$Y^A = [0, m(\rho + ia\cos\vartheta)^{-1}], \tag{2.16}$$

$$ds^2 = \frac{\rho^2 - m^2 + a^2\cos^2\vartheta}{\rho^2 - m^2 + a^2} d\rho^2 + (\rho^2 - m^2 + a^2\cos^2\vartheta) d\vartheta^2 + (\rho^2 - m^2 + a^2)\sin^2\vartheta\, d\varphi^2, \tag{2.17}$$

where $(t, \rho + m, \vartheta, \varphi)$ are the Boyer–Lindquist coordinates.

(2) The generalized Papapetrou–Majumdar class[15] reads

$$Y^A = [w, w], \tag{2.18}$$

$$ds^2 = dr^2 + r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2), \tag{2.19}$$

where w is an arbitrary Newtonian potential ($\Delta w = 0$).

III. THE ASYMPTOTIC EXPANSION

In order to state our requirements and our results, we start with the following definitions. The latter two are based on the "Geroch compactification" of a static space-time.[1]

Definition 1 (*asymptotic flatness*): A solution of (2.9) and (2.10) is called AFSEM if there exists a coordinate system x^i on N such that

$$Y^A = 0^\infty(r^{-1}), \quad \gamma_{ij} = \delta_{ij} + 0^\infty(r^{-1}) \tag{3.1}$$

[a field $\varphi(x)$ is said to be $0^\infty(f(r))$ if it is C^∞ and if there is a C^∞ function $f(r)$ such that

$$|\varphi(x^j)| \leqslant |f(r)|, \quad |\partial_i \varphi(x^j)| \leqslant \left| \frac{\partial f(r)}{\partial r} \right|, \ldots, \tag{3.2}$$

where $r^2 := x^i x_i$.]

Definition 2 (*asymptotic solution*): C^∞-functions $Y^{A(k)}$ (complex) and $\gamma_{ij}{}^{(k)}$ (real) on N are called asymptotic solutions of the SEM equations if there is a coordinate system where they fall off like (3.1) and where they satisfy (2.9) and (2.10) modulo remainder terms of order $0^\infty(r^{-(k+3)})$, i.e.,

$$\Delta(\gamma) Y^A - 2\Theta^{-1}(Y^*_C D_i Y^C) D^i Y^A = 0^\infty(r^{-(k+3)}), \tag{3.3}$$

$$R_{ij}(\gamma) - 2\Theta^{-2}(Z_{(i} Z^*_{j)} - D_{(i} Y^C D_{j)} Y^*_C) = 0^\infty(r^{-(k+3)}). \tag{3.4}$$

Definition 3 (C^k-*compactification*): The manifold $\widetilde{N}$ is called C^k-compactification of N, if $\widetilde{N}$ consists of N plus a point Λ and if there is a C^∞-function Ω and a C^k-metric $\tilde{\gamma}_{ij}$ on $\widetilde{N}$ such that

$$\tilde{\gamma}_{ij} = \Omega^2 \gamma_{ij}, \tag{3.5}$$

$$\Omega \,|_\Lambda = \widetilde{D}_i \Omega \,|_\Lambda = (\widetilde{D}_i \widetilde{D}_j \Omega - 2\tilde{\gamma}_{ij})\,|_\Lambda = 0. \tag{3.6}$$

Definition 4 (*multipole moments*)[1]: Assume that N has a C^k-compactification, and that $\widetilde{Y}^A = \Omega^{-1/2} Y^A$ extend to C^{k-1}-functions on $\widetilde{N}$.

From $\widetilde{Y}^A$, sets $P^A{}_{a_1 \cdots a_m}$ $(m \leqslant k-1)$ of trace-free, symmetric tensor fields can be defined by ($\mathscr{T}$ denotes the trace-free part with respect to $\tilde{\gamma}_{ij}$)

$$P^A = \widetilde{Y}^A, \tag{3.7}$$

$$P^A{}_{a_1 \cdots a_{n+1}} = \mathscr{T}\big[\widetilde{D}_{a_1} P^A_{a_2 \cdots a_{n+1}} - \tfrac{1}{2} n(2n-1) \widetilde{R}_{a_1 a_2}(\tilde{\gamma}) P^A_{a_3 \cdots a_{n+1}}\big]. \tag{3.8}$$

Their values at Λ are called multipole moments:

$$M^A{}_{a_1 \cdots a_m} = P^A{}_{a_1 \cdots a_m}|_\Lambda. \tag{3.9}$$

Our main results read as follows:

Theorem 1: Let k be an arbitrary integer $(\geqslant 0)$. Every AFSEM solution has a C^k-compactification $\widetilde{N} = N \cup \Lambda$, which admits C^{k-1}-extensions of $\widetilde{Y}^A = \Omega^{-1/2} Y^A$ to Λ.

There exists a coordinate system x^i on $\widetilde{N}$ such that

(a) Λ is located at $\bar{x}^i = 0$, $\Omega = \bar{r}^2$;

(b) $\tilde{\gamma}_{ij}$ and $\widetilde{Y}^A$ have Taylor expansions up to order k and $k-1$, respectively, the remainders fall off like $0^\infty(r^{k+1})$, $0^\infty(r^k)$;

(c) The Taylor coefficients are uniquely determined by the multipole moments only.

Theorem 2: Let $k \geqslant 0$ be arbitrary.

To every given set $\mathscr{M}_k = \{M^A{}_{a_1 \cdots a_{l-1}}, 1 \leqslant l \leqslant k\}$ of symmetric, trace-free matrices there exist a manifold N ($\mathbb{R}$ minus ball) with C^∞ fields $\gamma_{ij}{}^{(k)}$, $Y^{A(k)}$ which are uniquely determined by the parameters $\mathscr{M}_k$ such that

(a) $\gamma_{ij}{}^{(k)}$, $Y^{A(k)}$ are asymptotic solutions of the AFSEM equations;

(b) There is a C^k-compactification of N, with C^{k-1} fields $\widetilde{Y}^A$, whose multipole moments are $\mathscr{M}_k$.

The proofs of these theorems are identical with those in the stationary case, which is described in detail in Ref. 9, so we give only an outline here.

Outline of proofs of the theorems

(1) Given some AFSEM solution, we proceed by establishing $1/r$ expansions of Y^A, γ_{ij} inductively. Equations (2.9) and (2.10) can be rewritten in the form $\Delta Y^A = \cdots$, $\Delta \gamma_{ij} = \cdots$, with flat-space Laplacians on the left-hand side. We insert the known parts of the expansions into the right-hand side and invert the Laplacians. This works if we adapt the coordinate system in the k th step such that

$$\Lambda_i = \partial^j(\gamma_{ij} - \tfrac{1}{2}\delta_{ij}\delta^{kl}\gamma_{kl}) = 0^\infty(r^{-(k+2)}). \tag{3.10}$$

It is shown in Ref. 9 that the required coordinate transformations exist, that they map a neighborhood of Λ into itself, that they do not destroy the forms of Y^A and γ_{ij} up to the known order k, and that the Laplace operators can be inverted. Since we are interested in the general solution, we must admit a contribution from the homogeneous solution in Y^A and γ_{ij} of order $r^{-(k+1)}$. In Y^A, these contributions involve the 2^k-moments $M^A{}_{a_1 \cdots a_k}$; in γ_{ij}, it can be shown that they are "pure gauge."

Having established the expansions, the C^k-compactification and the C^{k-1}-fields of Theorem 1 are obtained, as in flat space, by inversion on the unit circle ($\bar{x}^i = r^{-2} x^i$).

(2) If we are given the moments, we can construct fields $Y^{A(k)}$, $\gamma_{ij}{}^{(k)}$ by the prescription which was established in the proof of Theorem 1. They are asymptotic solutions of (2.9) and (2.10) if we can satisfy the gauge condition (3.10) simultaneously. This can be shown with the help of the Bianchi identities.

In the coordinate system (3.10), the asymptotic expansions of the fields in terms of the multipole moments (3.9) start as follows (we chose $k = 3$):

$$Y^A = \frac{M^A}{r} + \frac{M^A_i x^i}{r^3} + \frac{1}{2}\frac{M^A_{ij} x^i x^j}{r^5} + 0^\infty\left(\frac{1}{r^4}\right), \tag{3.11}$$

$$\begin{aligned}\gamma_{ij} = \delta_{ij} &+ \frac{M^*_A M^A}{r^4}(\delta_{ij} r^2 - x^i x^j) \\ &+ \frac{M_A M^{A*}_k + M^*_A M^A_k}{r^6}\,[r^2(\delta_{k(i} x_{j)} + \delta_{ij} x^k) \\ &- 2x^i x^j x^k] + 0^\infty\left(\frac{1}{r^4}\right).\end{aligned} \tag{3.12}$$

Y^A will pick up trace terms in the higher orders.

We remark that this method of obtaining a $1/r$ expansion works (at least) for the following class of field equations. Let γ_{ij} be a positive definite, real C^∞-metric in three dimensions and U^A some C^∞-fields (real or complex). Assume that they fall off as in (3.1) and that they satisfy equations of the type $\Delta(\gamma) U^A = F^A$, $R_{ij}(\gamma) = \tau_{ij}$. The right-hand sides of these equations should be any finite or convergent infinite sum, where each term is some product of the U^A, their first derivatives, the complex conjugates of these, and of the metric. For everyone of those terms, count the total number of fields and add the total number of derivatives which occur there. The numbers that are obtained in this way should be:

For every term in F^A: Odd and $\geqslant 5$.

For every term in τ_{ij}: Even and $\geqslant 4$.

Moreover, for Theorem 2 to be true, the Bianchi identities for γ^{ij} should not impose constraints on the U^A, i.e., $D^i \tau_{ij} = \frac{1}{2} D_i \tau$ should be satisfied *by virtue* of the Poisson equations for the U^A (this is automatically true if the field equations can be derived from a Lagrangian). All these properties can be easily verified for the SEM equations (2.9) and (2.10) (one has to expand the Θ^{-1} in powers of $Y^A Y^*_A$).

For any fields γ_{ij} and U^A which satisfy equations with these properties, a treatment along the lines of Ref. 9 is straightforward; Definitions 1–4 can be applied, and Theorems 1 and 2 hold.

We remark that $\widetilde{U}^A$ and $\tilde{\gamma}_{ij}$ are very likely not only C^{k-1} (C^k) but, in suitably chosen coordinates, analytic near Λ. However, in the present approach, a proof of this seems to require a rather unnatural procedure; a geometric way to establish analyticity as well as the identity of two solutions with the same moments would be, e.g., to proceed as in Refs. 6–8, where these results have been obtained for static and stationary solutions.

IV. THE MULTIPOLE MOMENTS

In this section we discuss briefly the Y^A-fields and their moments. The real (imaginary) parts of v are called electric (magnetic) potentials, the real (imaginary) parts of w are called mass (angular, dual mass) potentials. There are the corresponding four sets of multipole moments, defined by (3.9). For the monopoles and dipoles we employ a proper notation:

$$\widetilde{Y}^A|_\Lambda = [M + iS, Q + iB], \tag{4.1}$$

$$\widetilde{D}_i \widetilde{Y}^A|_\Lambda = [M_i + iS_i, Q_i + iB_i]. \tag{4.2}$$

Having established that $\widetilde{Y}^A$ are C^k near spatial infinity, the same is, of course, true for any analytic function of the $\widetilde{Y}^A$. In particular, the fields

$$\phi_M + i\phi_S = -w(1 + vv^* - ww^*)^{-a}, \tag{4.3}$$

$$\phi_E + i\phi_B = v(1 + vv^* - ww^*)^{-a} \tag{4.4}$$

are, for $a = \frac{1}{4}, \frac{1}{2}, 1$, generalizations of the potentials introduced by Geroch,[1] Hoenselaers,[3] and Hansen,[2] respectively.

From the symmetry of the field equations (2.9) and (2.10) it is clear that they respect the presence of B and S. However, if we pass over to the four-dimensional description on X in terms of $g_{\mu\nu}$ and $F_{\mu\nu}$, our topological assumptions on X are incompatible with the existence of these monopoles. This can be seen from (2.2) and (2.4) which, together with Stokes' law, imply

$$\oint \lambda^{-1} f_i dS^i = 0, \tag{4.5}$$

$$\oint [\lambda^{-2}\omega_i - i(\Psi^* D_i \Psi - \Psi D_i \Psi^*)] dS^i \tag{4.6}$$

over any closed surface. Inserting the expansions (3.11) and (3.12) and taking the limit $r \to \infty$ of the integrals, it follows that $B = S = 0$.

We will now analyze the U(1, 1)-invariance of the Eqs. (2.9) and (2.10) in terms of the moments. Since γ_{ij} is not transformed, we can choose some compactification $\widetilde{N}^k$, and study the action of a Lorentz matrix $L^A{}_B$ on $\widetilde{Y}^B$. The four parameters of U(1, 1) correspond, in general, to the four different kinds of fields. In terms of the multipole moments this means that we may change the kind of the moments (e.g., charge a vacuum solution) but, since L is constant, the moments of a certain order are shuffled among themselves only.

The action of a U(1, 1) transformation on the multipole moments can readily be understood in terms of invariants like $\widetilde{Y}^*_A \widetilde{Y}^A$, $(\widetilde{D}_i \widetilde{Y}^A) Y^*_A$, etc. and in terms of the vector

$$\widetilde{Z}_i = \epsilon_{AB} \widetilde{Y}^A \widetilde{D}_i \widetilde{Y}^B = \Omega^{-1} \epsilon_{AB} Y^A D_i Y^B = \Omega^{-1} Z_i \tag{4.7}$$

(see 2.12) which transforms as $Z'_i = (\det L) Z_i$.

Consider, as an example, the case where we start with an uncharged solution (i.e., $\tilde{v} = 0$) or with a nonmassive charge (i.e., $\tilde{w} = 0$). It follows that $\widetilde{Z}_i = 0$ in both cases. Now apply any U(1, 1) transformation. Clearly, we stay within the subspace characterized by $Z_i = 0$, so the new solution $\widetilde{Y}'_A$, has either again $\tilde{v}' = 0$, or $\tilde{w}' = 0$, or $D_i(\ln \tilde{v}') = D_i(\ln \tilde{w}')$. If we assume, in the latter case, that $M' \neq 0$ and $Q' \neq 0$, we may take the limit at Λ of this equation, and if, in addition, $S' = B' = 0$, it follows that $B'_i = M'^{-1} Q' S'_i$, i.e., the gyromagnetic ratio is equal to 2. This result was obtained, in a nongeometric language, by Reina and Treves.[16]

ACKNOWLEDGMENT

I am grateful to R. Beig and to C. Hoenselaers for discussions.

[1] R. Geroch, J. Math. Phys. **11**, 2580 (1970).
[2] R. O. Hansen, J. Math. Phys. **15**, 46 (1974).
[3] C. Hoenselaers, Progr. Theor. Phys. **55**, 406 (1976).
[4] B. Xanthopoulos, J. Phys. A **12**, 1025 (1979).
[5] P. Kundu, J. Math. Phys. **22**, 1236 (1981).
[6] R. Beig and W. Simon, Comm. Math. Phys. **78**, 75 (1980).
[7] R. Beig and W. Simon, Proc. R. Soc. London, Ser. A **376**, 333 (1981).
[8] P. Kundu, J. Math. Phys. **22**, 2006 (1981).
[9] W. Simon and R. Beig, J. Math. Phys. **24**, 1163 (1983).
[10] D. Kramer *et al.*, *Exact Solutions of Einstein's Field Equations* (Cambridge, U.P., New York, 1979), §30.
[11] K. S. Thorne, Rev. Mod. Phys. **52**, 299 (1980).
[12] R. Beig and B. Schmidt, Comm. Math. Phys. **87**, 65 (1982).
[13] R. Geroch, J. Math. Phys. **12**, 918 (1970).
[14] W. Kinnersley, J. Math. Phys. **14**, 651 (1973).
[15] W. Israel and G. A. Wilson, J. Math. Phys. **13**, 865 (1972).
[16] C. Reina and A. Treves, Phys. Rev. D **11**, 3031 (1975).

A complete multipole expansion for a test particle in geometric fields

M. Toller[a)] and R. Vaia[a)]
Department of Physics, University of Trento, I-38100 Trento, Italy

(Received 2 June 1983; accepted for publication 30 September 1983)

We study the motion of a test particle interacting with a class of external fields including torsion, curvature, electromagnetic and Yang–Mills fields, using the definition of total energy momentum, angular momentum, charge and isotopic spin introduced in a preceding paper. We define a set of reduced multipole moments and we show that they give a complete description of the density and the flow of the quantities mentioned above. We write explicitly the exact evolution equations in terms of the reduced multipole moments and we show that they are the only consequences of the balance equations.

PACS numbers: 04.20.Me

1. INTRODUCTION

The motion of an extended test particle has been studied in a preceding paper[1] taking into account its interaction with a large class of geometric fields, including gravitation with torsion and curvature, electromagnetic and Yang–Mills gauge fields. The approach of that work is inspired by the treatment given by Dixon[2,3] of an extended body in general relativity and is based on the balance equations for energy, momentum, relativistic angular momentum, electric charge and isotopic spin. These quantities are treated in a symmetric way and called the components of the n-momentum.

The components $p_\alpha(s)$ of the total n-momentum of the particle measured with respect to the reference frame s have been defined exactly in Ref. 1 in such a way that a linear combination of them is exactly conserved whenever the geometric fields have a one-parameter symmetry group. In the same paper one can also find the exact equations which describe the dependence of the quantities $p_\alpha(s)$ on the reference frame s. These equations, together with some subsidiary conditions which define the center of the particle and some assumptions about its internal structure, can be used to describe the motion of the particle. The pole-dipole approximation has been treated in detail.

In the present paper we complete that investigation by deriving explicitly the whole multipole expansion. It is known[2,3] that the balance equations imply an infinite set of relations between the multipole moments. We give all these relations explicitly and we use them to define a set of "reduced" multipole moments, which are independent, apart from the evolution equations which determine the derivatives of the quantities $p_\alpha(s)$. We show that under very general conditions the reduced multipoles give a complete description of the density and the flow of n-momentum and that the information contained in the balance equations has been exploited completely. A more detailed treatment can be found in Ref. 4.

Since we are interested in the dependence of p_α on the reference frame s, we work in the space $\mathscr{S}$ of all the reference frames. We include in the definition of reference frame the choice of the internal gauge at the origin; therefore the dimension of the space $\mathscr{S}$ is $n = k + 10$, where k is the dimension of the internal gauge group.

As it has been shown in Refs. 5 and 6, the gravitational and the internal gauge fields can be described geometrically by means of n vector fields A_α in the space $\mathscr{S}$. $A_0,\dots,A_3$ represent infinitesimal parallel displacements, $A_4,\dots,A_9$ represent infinitesimal Lorentz transformations, and the other fields represent infinitesimal internal gauge transformations. We also use the differential 1-forms ω^β defined by

$$i_\alpha \omega^\beta = \delta^\beta_\alpha, \tag{1.1}$$

where i_α is the internal product operator corresponding to the vector field A_α and the structure coefficients $F^\gamma_{\alpha\beta}(s)$ defined by

$$[A_\alpha, A_\beta] = -F^\gamma_{\alpha\beta} A_\gamma. \tag{1.2}$$

The normal coordinates $\xi^\alpha(s,s')$ of the point s with respect to the point s' are defined by[1]

$$s = \exp(\xi^\alpha A_\alpha)s', \tag{1.3}$$

where $t \to \exp(tB)$ is the one-parameter group of mappings generated by the vector field B. They satisfy the differential equation

$$\xi^\gamma(s,s') L_\gamma \xi^\alpha(s,s') = \xi^\alpha(s,s'). \tag{1.4}$$

We have indicated by L_α the derivative in the direction of A_α with respect to the argument s. We consider also the matrices $D_\alpha^{\ \beta}(s,s')$ introduced to Ref. 1 defined by

$$\xi^\gamma(s,s') L_\gamma D_\alpha^{\ \beta}(s,s') = \xi^\gamma(s,s') F^{\ \beta}_{\gamma\delta}(s) D_\alpha^{\ \delta}(s,s'), \tag{1.5}$$

$$D_\alpha^{\ \beta}(s',s') = \delta^\beta_\alpha. \tag{1.6}$$

Both these quantities are defined if s belongs to a suitable neighborhood of s'.

The density and the flow of the n-momentum of matter are described by n differential 3-forms τ^M_α which satisfy the balance equations

$$\eta_\alpha = d\tau^M_\alpha - F^\gamma_{\alpha\beta}\omega^\beta \wedge \tau^M_\gamma = 0. \tag{1.7}$$

The definition of the total n-momentum suggested in Ref. 1 is

$$p_\alpha(s') = \int_{\mathscr{R}(s')} D_\alpha^{\ \beta}(s,s')\tau^M_\beta, \tag{1.8}$$

[a)] Associato all'Istituto Nazionale di Fisica Nucleare.

where $\mathscr{R}(s')$ is a three-dimensional integration surface defined by

$$\xi^{\alpha}(s,s') = 0 \quad \text{for } \alpha \neq 1,2,3,$$
$$|\xi^{k}(s,s')| \leqslant l \quad \text{for } k = 1,2,3. \tag{1.9}$$

We assume that the parameter l can be chosen in such a way that $\mathscr{R}(s')$ crosses the region $\mathscr{U}$ where the differential forms τ_{α}^{M} are not vanishing. A more precise discussion of these conditions can be found in Ref. 1. Note that the values 1, 2, 3 of the indices, which correspond to space translations, play a special role in the definition (1.9). In the following, the indices i, j, k take only these values.

2. THE GENERALIZED MOMENTS

In order to describe with more detail the density of n-momentum, we generalize the definition (1.8) by introducing the functionals

$$p_{\alpha}[\varphi,s'] = \int_{\mathscr{R}(s')} D_{\alpha}^{\beta}(s,s')\,\varphi\tau_{\beta}^{M}, \tag{2.1}$$

where φ is a C^{∞} function of the variables $\xi^{k}(s,s')$ ($k = 1,2,3$). If the function φ is a monomial, we get the multipole moments. Since we want to also describe the flow of n-momentum, we introduce the functionals

$$j_{\theta\alpha}[\chi,s'] = \int_{\mathscr{R}(s')} D_{\alpha}^{\beta}(s,s')D_{\theta}^{\rho}(s,s')\chi\wedge i_{\rho}\tau_{\beta}^{M}, \tag{2.2}$$

where χ is a differential 1-form built by means of the variables ξ^{k}.

We consider an arbitrary line $t\to s'_t$ in the space $\mathscr{S}$ parametrized by a variable t. The velocity of the point s'_t can be written in the form

$$\frac{ds'_t}{dt} = v^{\theta}(t)A_{\theta}(s'_t). \tag{2.3}$$

It has been shown in Ref. 1 that the velocity of a point

$$s = s(t,\xi^{k}) = \exp(\xi^{k}A_{k})s'_t \tag{2.4}$$

of the surface $\mathscr{R}(s'_t)$ which corresponds to constant values of ξ^{k} is given by

$$\frac{\partial s}{\partial t} = v^{\theta}(t)D_{\theta}^{\rho}(s,s'_t)A_{\rho}(s). \tag{2.5}$$

We consider also the set

$$\mathscr{W} = \bigcup_{0<t<t_0}\mathscr{R}(s'_t) \tag{2.6}$$

and we assume that it is a 4-dimensional submanifold of $\mathscr{S}$ parametrized by the coordinates ξ^{k} and t. This implies that the vector field $\partial/\partial t$ is never tangent to the surfaces $\mathscr{R}(s'_t)$, in such a way that the vectors $\partial/\partial\xi^{k}$ and $\partial/\partial t$ form a basis in the tangent spaces of $\mathscr{W}$. Under these assumptions, we show that:

Proposition 1: The functionals

$$p_{\alpha}[\varphi,s'_t], \quad v^{\theta}(t)j_{\theta\alpha}[\chi,s'_t] \tag{2.7}$$

determine completely the restrictions $\tau_{\beta}^{M}|_{\mathscr{W}}$ of the differential forms τ_{β}^{M} to the submanifold $\mathscr{W}$.

Proof: It is sufficient to show that $\tau_{\beta}^{M}|_{\mathscr{W}}$ vanishes when the functionals (2.7) vanish. In fact, in this case from Eq. (2.1), since φ is arbitrary, we have

$$\tau_{\beta}^{M}\left(\frac{\partial}{\partial\xi^{1}},\frac{\partial}{\partial\xi^{2}},\frac{\partial}{\partial\xi^{3}}\right) = 0, \tag{2.8}$$

and from Eqs. (2.2) and (2.5), since the differential 1-form χ is also arbitrary, we have

$$\tau_{\beta}^{M}\left(\frac{\partial}{\partial t},\frac{\partial}{\partial\xi^{i}},\frac{\partial}{\partial\xi^{k}}\right) = 0, \quad i,k = 1,2,3. \tag{2.9}$$

It follows that the differential forms τ_{β}^{M} vanish when they are applied to three arbitrary vectors tangent to $\mathscr{W}$. We call the functionals (2.7) the "generalized moments."

Now we consider the derivative of the quantity (2.1) along the line $t\to s'_t$:

$$\frac{d}{dt}p_{\alpha}[\varphi,s'_t] = \int_{\mathscr{R}(s'_t)} v^{\theta}(t)L'_{\theta}[D_{\alpha}^{\beta}(s,s'_t)\varphi]\,\tau_{\beta}^{M}$$
$$+\int_{\mathscr{R}(s'_t)} v^{\theta}(t)D_{\theta}^{\rho}(s,s'_t)i_{\rho}\,d\,[D_{\alpha}^{\beta}(s,s'_t)\,\varphi\tau_{\beta}^{M}]. \tag{2.10}$$

The last contribution, due to the displacement of $\mathscr{R}(s'_t)$, has been obtained by means of Stokes' theorem and of Eq. (2.5). We have indicated by L'_{θ} the derivatives with respect to the point s'. From now on, we omit the arguments s and s'_t when there is no danger of confusion and we indicate by $\overline{D}_{\alpha}^{\beta}$ the inverse of the matrix D_{α}^{β}.

Since the function φ depends only on the variables ξ^{k}, from Eqs. (2.3) and (2.5) we have

$$\frac{\partial\varphi}{\partial t} = v^{\theta}L'_{\theta}\,\varphi + v^{\theta}D_{\theta}^{\rho}L_{\rho}\,\varphi = 0. \tag{2.11}$$

If we use this formula, we see after some calculation that the condition

$$\int_{\mathscr{R}(s'_t)} v^{\theta}D_{\theta}^{\rho}D_{\alpha}^{\beta}\varphi i_{\rho}\,\eta_{\beta} = 0, \tag{2.12}$$

where the differential 4-forms η_{β} are defined by Eq. (1.7), is equivalent to the equation

$$\frac{d}{dt}p_{\alpha}[\varphi] = v^{\theta}\{-F_{\theta\alpha}^{\beta}\,p_{\beta}[\varphi] + p_{\beta}[\varphi\hat{\psi}_{\theta\alpha}^{\beta}]$$
$$- j_{\theta\beta}[\varphi\hat{\chi}_{\alpha}^{\beta}] - j_{\theta\alpha}[d\varphi]\}, \tag{2.13}$$

where the functions $\hat{\psi}_{\theta\alpha}^{\beta}$ and the differential 1-forms $\hat{\chi}_{\alpha}^{\beta}$ are the restrictions to $\mathscr{R}(s'_t)$ of

$$\psi_{\theta\alpha}^{\beta} = F_{\theta\alpha}^{\beta}(s') + \overline{D}_{\gamma}^{\beta}L'_{\theta}D_{\alpha}^{\gamma}$$
$$+ D_{\theta}^{\rho}\overline{D}_{\gamma}^{\beta}(L_{\rho}D_{\alpha}^{\gamma} - F_{\rho\delta}^{\gamma}(s)D_{\alpha}^{\delta}), \tag{2.14}$$
$$\chi_{\alpha}^{\beta} = \overline{D}_{\gamma}^{\beta}(L_{\mu}D_{\alpha}^{\gamma} - F_{\mu\delta}^{\gamma}(s)D_{\alpha}^{\delta})\omega^{\mu}. \tag{2.15}$$

As a consequence, we have the following result:

Proposition 2: The equation (2.13) is equivalent to the condition

$$\eta_{\beta}|_{\mathscr{W}} = 0. \tag{2.16}$$

Proof: From Eqs. (2.12) and (2.5) since the function φ is arbitrary, we have

$$\eta_\beta\left(\frac{\partial}{\partial t},\frac{\partial}{\partial \xi^1},\frac{\partial}{\partial \xi^2},\frac{\partial}{\partial \xi^3}\right)=0 \qquad (2.17)$$

and Eq. (2.16) follows. It is clear that Eq. (2.16) implies Eq. (2.12).

In order to analyze with more detail Eq. (2.13), we need the following result:

Lemma: A differential 1-form χ in the variables ξ^k can be decomposed uniquely in the following way:

$$\chi = d\varphi + \theta, \qquad (2.18)$$

where the differential form θ is "transversal," namely it has the form

$$\theta = a_k d\xi^k \quad \text{with } a_k\,\xi^k = 0. \qquad (2.19)$$

Proof: if we put

$$\chi = b_k d\xi^k, \qquad (2.20)$$

we have

$$a_k\,\xi^k = \left(b_k - \frac{\partial\varphi}{\partial\xi^k}\right)\xi^k = 0. \qquad (2.21)$$

This differential equation has the general solution

$$\varphi(\xi) = \int_0^1 b_k(\alpha\xi)\xi^k d\alpha + \varphi(0) \qquad (2.22)$$

and $d\varphi$ is uniquely determined.

If we interpret the variables ξ^k as normal coordinates on the surface $\mathscr{R}(s_t')$, from Eq. (1.4) we have

$$\xi^k\frac{\partial}{\partial\xi^k} = \xi^k \mathrm{A}_k \qquad (2.23)$$

and the transversality condition (2.19) takes the form

$$\xi^k i_k \theta = 0. \qquad (2.24)$$

As a consequence of the preceding lemma, we can split the functional (2.2) into two independent functionals which vanish when their argument χ is, respectively, transversal or exact:

$$j_{\theta\alpha}[\chi] = \hat{j}_{\theta\alpha}[\chi] + \tilde{j}_{\theta\alpha}[\chi], \qquad (2.25)$$

$$\hat{j}_{\theta\alpha}[\chi] = j_{\theta\alpha}[d\varphi], \qquad (2.26)$$

$$\tilde{j}_{\theta\alpha}[\chi] = j_{\theta\alpha}[\theta]. \qquad (2.27)$$

With this notation we have

Proposition 3: If Eq. (2.16) holds, the functionals

$$p_\alpha[\varphi, s_t'],\ v^\theta(t)\tilde{j}_{\theta\alpha}(\chi, s_t'),\quad 0 \leqslant t \leqslant t_0, \qquad (2.28)$$

determine completely the differential forms $\tau_\alpha^M|_{\mathscr{W}}$ and satisfy the equation

$$\frac{d}{dt}p_\alpha = v^\theta(-F^{\beta}_{\theta\alpha}\,p_\beta + f_{\theta\alpha}), \qquad (2.29)$$

where

$$f_{\theta\alpha} = p_\beta[\hat{\psi}^\beta_{\theta\alpha}] - \tilde{j}_{\theta\beta}[\hat{\chi}^\beta_\alpha]. \qquad (2.30)$$

Proof: The differential form $\hat{\chi}^\beta_\alpha$ is transversal. In fact, from Eqs. (2.15) and (1.5) we have

$$\xi^k i_k\,\hat{\chi}^\beta_\alpha = \xi^k \overline{D}{}^\beta_\gamma(L_k D^\gamma_\alpha - F^\gamma_{k\delta}D^\delta_\alpha)|_{\mathscr{R}(s_t')} = 0. \qquad (2.31)$$

As a consequence, we can write Eq. (2.13) in the form

$$v^\theta \hat{j}_{\theta\alpha}[d\varphi] = v^\theta\{-F^\beta_{\theta\alpha}\,p_\beta(\varphi) + p_\beta[\varphi\hat{\psi}^\beta_{\theta\alpha}] - \tilde{j}_{\theta\beta}[\varphi\hat{\chi}^\beta_\alpha]\} - \frac{d}{dt}p_\alpha[\varphi]. \qquad (2.32)$$

This equation, together with Eq. (2.25), shows that the functionals (2.28) determine the functionals (2.7) and therefore the differential forms $\tau_\alpha^M|_{\mathscr{W}}$. If we put $\varphi = 1$, we get Eq. (2.29). We call the functionals (2.28) the "reduced generalized moments."

3. THE LOCAL CASE

In the preceding section we have used a minimal set of assumptions, in order to get a simpler treatment. However, in all the physically interesting problems one deals with a more specific situation, in which our results can be made stronger. First, we take into account the fact that $\mathscr{S}$ is a principal fiber bundle over the space-time manifold and therefore the structure coefficients have the property[5,6]

$$F^\gamma_{a\beta} = -F^\gamma_{\beta a} = \hat{F}^\gamma_{a\beta},\ a \geqslant 4, \qquad (3.1)$$

where we have indicated by $\hat{F}^\gamma_{\alpha\beta}$ the structure constants of the product of the Poincaré group and the internal gauge group.

Then we assume that the differential forms τ_α^M are "local," namely that

$$\tau_\alpha^M = \tfrac{1}{6}\,T^\mu_\alpha\,e_{urst}\,\omega^r\wedge\omega^s\wedge\omega^t, \qquad (3.2)$$

where the indices r,s,t,u take the values 0,1,2,3 and e_{urst} is the completely antisymmetric Levi-Civita symbol. The locality property can also be written in the form

$$i_a\,\tau_\alpha^M = 0,\ a \geqslant 4. \qquad (3.3)$$

If we indicate by

$$L_a = i_\alpha d + d i_\alpha \qquad (3.4)$$

the Lie derivative corresponding to the vector field A_α, from the balance equations (1.7) and from Eqs. (3.1) and (3.3) we have

$$L_a\tau_\alpha^M = -\hat{F}^\beta_{a\alpha}\,\tau_\beta^M,\ a \geqslant 4. \qquad (3.5)$$

As a consequence of this equation, the quantities $T^u_\alpha(s)$ which appear in Eq. (3.2) transform tensorially under Lorentz and internal gauge transformations of the reference frame s.

If the differential forms τ_α^M are known on a given set $\mathscr{W}$, Eq. (3.5) permits one to compute them on the larger set $\widetilde{\mathscr{W}}$, which is the union of all the fibers which intersect $\mathscr{W}$. Equations (3.3) and (3.5) show that the differential forms τ_α^M are the components of a tensorial form.[7] More briefly, we shall say that the differential forms τ_α^M are tensorial.

If we assume that τ_α^M are the components of a tensorial form, even if they do not satisfy the balance equations, after some calculation we see that the differential forms η_α defined by Eq. (1.7) satisfy the conditions

$$i_a\,\eta_\alpha = 0, \qquad (3.6)$$

$$L_a\,\eta_\alpha = -\hat{F}^\beta_{a\alpha}\,\eta_\beta,\quad a \geqslant 4, \qquad (3.7)$$

namely they are tensorial too.

It is natural to assume that the line $t \to s_t'$ has the property

$$v^0(t) > 0. \tag{3.8}$$

Since at the points s_t' we have

$$\frac{\partial}{\partial \xi^k} = A_k, \tag{3.9}$$

from Eq. (2.3) we see that the n vectors

$$\frac{\partial}{\partial t}, \frac{\partial}{\partial \xi^k}, A_a, \quad k = 1,2,3, \quad a \geqslant 4 \tag{3.10}$$

are linearly independent at these points. If the external fields are sufficiently weak and the parameter l which appears in Eq. (1.9) and depends on the size of the particle is sufficiently small, we can assume that the vector fields (3.10) are linearly independent at all the points of the set $\mathscr{W}$ defined by Eq. (2.6). Under this assumption, the propositions of the preceding section can be stated in the following stronger form.

Proposition 1': If the differential forms τ_α^M are local (respectively tensorial), they are completely determined on $\mathscr{W}$ (respectively on $\overline{\mathscr{W}}$) by the generalized moments (2.7).

Proof: if the generalized moments (2.7) vanish, from the preceding assumption and from Eqs. (2.8), (2.9), and (3.3) we see that $\tau_\alpha^M = 0$ on $\mathscr{W}$. If moreover Eq. (3.5) holds, these forms vanish also on $\overline{\mathscr{W}}$.

Proposition 2': If the differential forms τ_α^M are tensorial, Eq. (2.13) is equivalent to the condition

$$\eta_\alpha = 0 \quad \text{on } \overline{\mathscr{W}}. \tag{3.11}$$

Proof: From Eqs. (2.17) and (3.6) we see that $\eta_\alpha = 0$ on $\mathscr{W}$. Then, using Eq. (3.7) we obtain Eq. (3.11).

Finally, reasoning as in the proof of Proposition 1' we have the following.

Proposition 3': If the differential forms τ_α^M are tensorial and satisfy the balance equations, they are completely determined on $\overline{\mathscr{W}}$ by the reduced generalized moments (2.28).

4. POWER EXPANSION OF THE GEOMETRIC QUANTITIES

In order to find the explicit multipole expansion of Eq. (2.30), we have to find the expansion of the quantities (2.14) and (2.15) into powers of the normal coordinates ξ^α, for a given fixed choice of s'.

If we consider a function $f(s)$ and we put

$$s_h = \exp(h\xi^\mu A_\mu)s', \tag{4.1}$$

we have

$$\frac{\partial}{\partial h} f(s_h) = [\xi^\mu L_\mu f(s)]_{s=s_h}. \tag{4.2}$$

Note that the normal coordinates ξ^μ of the point s_1 have to be considered as constant and the operators L_μ do not act on them. Therefore, if we consider the Taylor series with respect to the parameter h and we put $h = 1$, we get

$$f(s) = \sum_{n=0}^{\infty} \frac{1}{n!} \xi^{\mu_1} \cdots \xi^{\mu_n} L'_{\mu_1} \cdots L'_{\mu_n} f(s'). \tag{4.3}$$

This series converges if f is an analytic function of the coordinates ξ^μ; if f is only a C^∞ function, Eq. (4.3) has to be considered as an asymptotic expansion.

It is useful to introduce the vector field

$$R = \xi^\alpha \frac{\partial}{\partial \xi^\alpha} = \xi^\alpha A_\alpha, \tag{4.4}$$

where the last equality is an immediate consequence of Eq. (1.4). The corresponding Lie derivative L_R has the properties

$$L_R \xi^\alpha = \xi^\alpha, \tag{4.5}$$

$$L_R d\xi^\alpha = d\xi^\alpha, \tag{4.6}$$

$$L_R \omega^\alpha = d\xi^\alpha + \xi^\beta F^\alpha_{\beta\gamma} \omega^\gamma. \tag{4.7}$$

From Eqs. (4.5) and (4.6), we see that the operator L_R when applied to a product of the quantities ξ^α and $d\xi^\alpha$ multiplies it by the number of factors. One can easily see that it has an inverse L_R^{-1} in the space of the differential 1-forms and in the space of the differentiable functions which vanish at the origin.

If we put

$$F_\alpha^\beta(s,s') = \xi^\gamma(s,s') F^\beta_{\gamma\alpha}(s), \tag{4.8}$$

Eq. (1.5) can be written in matrix form as

$$L_R D = FD. \tag{4.9}$$

Its solution satisfying the condition (1.6) is given by the series

$$D = \sum_{n=0}^{\infty} (L_R^{-1} F)^n. \tag{4.10}$$

It is convenient to introduce the notations

$$\{q\} = \mu_1, \mu_2, \ldots, \mu_q, \quad \xi^{\{q\}} = \xi^{\mu_1} \xi^{\mu_2} \ldots \xi^{\mu_q}. \tag{4.11}$$

Then, from Eq. (4.8) and the general expansion formula (4.3) we obtain

$$F_\alpha^\beta(s,s') = \sum_{q=1}^{\infty} \xi^{\{q\}} \Phi^\beta_{\{q\}\alpha}(s'), \tag{4.12}$$

where

$$\Phi^\beta_{\{q\}\alpha}(s') = \frac{1}{(q-1)!} L'_{(\mu_1} \ldots L'_{\mu_{q-1}} F^\beta_{\mu_q)\alpha}(s'). \tag{4.13}$$

The round brackets indicate the symmetrization with respect to the indices enclosed by them. If we substitute Eq. (4.12) into Eq. (4.10) we get the power expansion

$$D = \sum_{n=0}^{\infty} \sum_{q_1 \ldots q_n > 1} C(q_1, \ldots, q_n) \times \Phi_{\{q_1\}} \cdots \Phi_{\{q_n\}} \xi^{\{q_1\}} \ldots \xi^{\{q_n\}}, \tag{4.14}$$

where

$$C(q_1, \ldots, q_n) = [(q_1 + \cdots + q_n)(q_2 + \cdots + q_n) \cdots q_n]^{-1}. \tag{4.15}$$

In a similar way, from the differential equation

$$L_R \overline{D}^T = -F^T \overline{D}^T \tag{4.16}$$

we obtain the expansion

$$\overline{D} = \sum_{n=0}^{\infty} \sum_{q_1 \ldots q_n > 1} (-1)^n C(q_1, \ldots, q_n) \times \Phi_{\{q_n\}} \cdots \Phi_{\{q_1\}} \xi^{\{q_1\}} \ldots \xi^{\{q_n\}} \tag{4.17}$$

and from Eq. (4.7), which can be written in matrix form as

$$L_R \omega = d\xi + F\omega, \tag{4.18}$$

we obtain

$$\omega = \sum_{n=0}^{\infty} (L_R^{-1} F)^n d\xi, \tag{4.19}$$

namely

$$\omega^\alpha = \sum_{n=0}^{\infty} \sum_{q_1 \ldots q_n \geqslant 1} C(q_1,\ldots,q_n,1) \times [\Phi_{\{q_1\}} \cdots \Phi_{\{q_n\}}]^\alpha_\beta \xi^{\{q_1\}} \ldots \xi^{\{q_n\}} d\xi^\beta. \tag{4.20}$$

From Eqs. (2.14) and (2.15) after some calculations we get

$$L_R \psi^\beta_{\theta\alpha} = \bar{D}^\beta_\sigma \xi^\mu D^\gamma_\alpha L_\gamma F^\sigma_{\mu\rho}(s) D^\rho_\theta, \tag{4.21}$$

$$L_R \chi^\beta_\alpha = \bar{D}^\beta_\sigma \xi^\mu D^\gamma_\alpha L_\gamma F^\sigma_{\mu\rho}(s) \omega^\rho. \tag{4.22}$$

If we substitute Eq. (4.3) into the identity

$$L'_\alpha f(s) = 0 \tag{4.23}$$

and we use Eq. (4.8) of Ref. 1, we obtain the formula

$$D^\gamma_\alpha(s,s') L_\gamma f(s) = \sum_{n=0}^{\infty} \frac{1}{n!} \xi^{\mu_1} \ldots \xi^{\mu_n} \times L'_\alpha L'_{\mu_1} \ldots L'_{\mu_n} f(s'), \tag{4.24}$$

which, in particular, gives

$$\xi^\mu D^\gamma_\alpha L_\gamma F^\sigma_{\mu\rho}(s) = \sum_{l=1}^{\infty} \xi^{\{l\}} L'_\alpha \Phi^\sigma_{\{l\}\rho}(s'). \tag{4.25}$$

Finally, from Eqs. (4.21) and (4.22), using Eqs. (4.14), (4.17), (4.20), and (4.25), we obtain the desired power expansions

$$\begin{aligned} \psi^\beta_{\theta\alpha} &= \sum_{r=1}^{\infty} \psi^\beta_{\theta\alpha\{r\}} \xi^{\{r\}} \\ &= \sum_{m,n \geqslant 0} \sum_{l \geqslant 1} \sum_{q_1 \ldots q_m \geqslant 1} \sum_{p_1 \ldots p_n \geqslant 1} [(-1)^m \\ &\quad \times (q_1 + \cdots + q_m + l + p_1 + \cdots + p_n)^{-1}] \\ &\quad \times C(q_1,\ldots,q_m) C(p_1,\ldots,p_n) [\Phi_{\{q_m\}} \cdots \Phi_{\{q_1\}} L'_\alpha \Phi_{\{l\}} \\ &\quad \times \Phi_{\{p_1\}} \cdots \Phi_{\{p_n\}}]^\beta_\theta \xi^{\{q_1\}} \ldots \xi^{\{q_m\}} \xi^{\{l\}} \xi^{\{p_1\}} \ldots \xi^{\{p_n\}}, \end{aligned} \tag{4.26}$$

$$\begin{aligned} \chi^\beta_\alpha &= \sum_{r=1}^{\infty} \chi^\beta_{\alpha\{r\}\theta} \xi^{\{r\}} d\xi^\theta \\ &= \sum_{m,n \geqslant 0} \sum_{l \geqslant 1} \sum_{q_1 \ldots q_m \geqslant 1} \sum_{p_1 \ldots p_n \geqslant 1} [(-1)^m \\ &\quad \times (q_1 + \cdots + q_m + l + p_1 + \cdots + p_n + 1)^{-1}] \\ &\quad \times C(q_1,\ldots,q_m) C(p_1,\ldots,p_n,1) [\Phi_{\{q_m\}} \cdots \Phi_{\{q_1\}} L'_\alpha \Phi_{\{l\}} \\ &\quad \times \Phi_{\{p_1\}} \cdots \Phi_{\{p_n\}}]^\beta_\theta \xi^{\{q_1\}} \ldots \xi^{\{q_m\}} \xi^{\{l\}} \xi^{\{p_1\}} \ldots \xi^{\{p_n\}} d\xi^\theta. \end{aligned} \tag{4.27}$$

5. THE MULTIPOLE EXPANSION

We define the multipole moments in terms of the generalized moments introduced in Sec. 2 in the following way

$$p^{\{r\}}_\alpha = p_\alpha [\xi^{\{r\}}], \tag{5.1}$$

$$v^\theta j^{\{r\}i}_{\theta\alpha} = v^\theta j_{\theta\alpha} [\xi^{\{r\}} d\xi^i], \quad r \geqslant 0. \tag{5.2}$$

The indices represented by $\{r\}$ take the values 1,2,3. The reduced multipole moments are given by the quantities (5.1) together with the quantities

$$\begin{aligned} v^\theta \tilde{j}^{\{r\}i}_{\theta\alpha} &= v^\theta \tilde{j}_{\theta\alpha} [\xi^{\{r\}} d\xi^i] \\ &= v^\theta j_{\theta\alpha} \left[\xi^{\{r\}} d\xi^i - \frac{1}{r+1} d(\xi^{\{r\}} \xi^i) \right] \\ &= v^\theta (j^{\{r\}i}_{\theta\alpha} - j^{\{r|i\}}_{\theta\alpha}). \end{aligned} \tag{5.3}$$

Since the polynomials form a dense set in the space of the test functions defined in the cubes (1.9), the multipole moments determine uniquely the corresponding generalized moments. It follows that one can easily restate the results of Secs. 2 and 3 in terms of multipole moments.

If we substitute Eqs. (4.26) and (4.27) into Eq. (2.30) we obtain the multipole expansion

$$f_{\theta\alpha} = \sum_{r=1}^{\infty} (\psi^\beta_{\theta\alpha\{r\}} p^{\{r\}}_\beta - \chi^\beta_{\alpha\{r\}i} \tilde{j}^{\{r\}i}_{\theta\beta}). \tag{5.4}$$

Complete multipole expansions of this kind have been obtained by Yasskin and Stoeger.[8] Due to the particular definitions we have adopted, the formalism presented here has two features whose importance has been emphasized by Dixon.[2,3] One of them is the introduction of the reduced multipole moments, which are not constrained by the balance equations.

The other feature appears when the space $\mathscr{S}$ has a one-parameter symmetry group generated by the vector field

$$B = b^\alpha A_\alpha, \tag{5.5}$$

which has the property

$$[B, A_\alpha] = 0. \tag{5.6}$$

From this condition we obtain[1]

$$L_\theta b^\alpha = F^\alpha_{\theta\beta} b^\beta, \tag{5.7}$$

$$b^\alpha L_\alpha \Phi_{\{l\}} = 0 \tag{5.8}$$

and from Eqs. (4.26), (4.27), and (5.4) we obtain

$$b^\alpha f_{\theta\alpha} = 0. \tag{5.9}$$

Note that all the terms of the multipole expansion of this quantity vanish separately. From Eq. (2.29), (5.7), and (5.9) we obtain immediately the conservation equation

$$\frac{d}{dt}(b^\alpha p_\alpha) = 0, \tag{5.10}$$

corresponding to the symmetry group we have considered. We see that the conserved quantity is a linear combination of the components of the n-momentum.

If we use Eq. (3.1) and we assume that there is no gravitational field, namely that

$$F^\alpha_{rs} = 0, \quad \alpha \leqslant 9, \quad r,s \leqslant 3, \tag{5.11}$$

Eq. (5.4) takes the much simpler form

$$f_{\theta\alpha} = \sum_{l \geqslant 1} \left(\frac{1}{l} p^{k_1 \ldots k_l}_b \delta^r_\theta + \frac{1}{l+1} p^{k_1 \ldots k_l i}_b \widehat{F}^r_{i\theta} - \frac{1}{l+1} \tilde{j}^{k_1 \ldots k_l r}_{\theta b} \right) \frac{1}{(l-1)!} L_\alpha L_{k_1} \ldots L_{k_{l-1}} F^b_{k_l r}, \tag{5.12}$$

where the index b takes the value $10,\ldots,(n-1)$ and the quantities F^{b}_{rs} are the strengths of the gauge fields. Note that the expression (5.12) is linear in these fields.

For a general external field, considering only dipole and quadrupole terms, Eq. (5.4) can be written in the more explicit form

$$\begin{aligned} f_{\theta\alpha} = {} & p^{k}_{\beta} L_{\alpha} F^{\beta}_{k\theta} - \tfrac{1}{2} \tilde{j}^{ik}_{\theta\beta} L_{\alpha} F^{\beta}_{ik} \\ & + \tfrac{1}{2} p^{ik}_{\beta} (L_{\alpha} L_{i} F^{\beta}_{k\theta} + L_{\alpha} F^{\beta}_{i\sigma} F^{\sigma}_{k\theta} - F^{\beta}_{i\sigma} L_{\alpha} F^{\sigma}_{k\theta}) \\ & - \tfrac{1}{3} \tilde{j}^{ijk}_{\theta\beta} (L_{\alpha} L_{i} F^{\beta}_{jk} + \tfrac{1}{2} L_{\alpha} F^{\beta}_{i\sigma} F^{\sigma}_{jk} - F^{\beta}_{i\sigma} L_{\alpha} F^{\sigma}_{jk}). \end{aligned} \tag{5.13}$$

Note that this expression is not linear in the gravitational fields.

These formulas, as well as the more general ones given above, are written with an unusual compact notation, but it is easy to translate them into the familiar tensor language based on anholonomic components. It is sufficient to take Eq. (3.1) into account and to remember that the quantities F^{α}_{rs} for $r,s = 0,\ldots,3$ and for various values of the index α are the components with respect to the orthogonal frames s'_t of the torsion tensor, the Riemann tensor [see Eq. (2.7) of Ref. 1], and the strength tensors of the gauge fields. The operators L_r $(r = 0,\ldots,3)$ acting on these components can be interpreted directly as covariant derivatives, while the operators L_a $(a \geqslant 4,\ldots,n)$ are linear transformations of the components which represent the infinitesimal generators of the Lorentz and the internal gauge groups [see Eq. (2.8) of Ref. 1].

[1] M. Toller, J. Math Phys. **24**, 613 (1983).
[2] W. G. Dixon, Proc. R. Soc. London, Ser. A **314**, 499 (1970); **319**, 509 (1970).
[3] W. G. Dixon, Philos. Trans. R. Soc. London, Ser. A **277**, 59 (1974).
[4] R. Vaia, thesis (University of Firenze, Italy, 1983) (unpublished).
[5] M. Toller, Nuovo Cimento B **44**, 67 (1978).
[6] G. Cognola, R. Soldati, L. Vanzo, and S. Zerbini, J. Math. Phys. **20**, 2613 (1979).
[7] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969).
[8] P. B. Yasskin and W. R. Stoeger, Phys. Rev. D **21**, 2081 (1980).

Material sources in the nonsymmetric Kaluza–Klein theory

M. W. Kalinowski[a)]

Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A7, Canada

(Received 4 February 1983; accepted for publication 29 April 1983)

In this paper, we deal with material sources in the nonsymmetric Kaluza–Klein theory. We consider phenomenological sources with nonzero fermion current, nonzero electric current, and nonzero spin–density tensor. From the Palatini variational principle we find equations for the gravitational and electromagnetic fields. We also consider the geodesic equation in the theory and the equation of motion for charged test particles.

PACS numbers: 04.50. + h

INTRODUCTION

In this paper, we consider material sources in the nonsymmetric Kaluza–Klein theory, i.e., the energy momentum tensor of external sources, the fermion current, the electric current, and the spin–density tensor of external sources.

The nonsymmetric Kaluza–Klein theory[1] unifies Moffat's theory of gravitation[2–4] and electromagnetism. In Ref. 1, we find vacuum equations in this theory. Here, we work with the phenomenological Lagrangian for such a theory. The paper is organized as follows.

In Sec. I, we briefly describe elements of the nonsymmetric Kaluza–Klein theory. In the second section, we introduce material sources and find equations for the gravitational and electromagnetic fields in the presence of matter with nonzero fermion current and nonzero electric current. In the third section, we define a new geometrical degree of freedom (a generalized contortion tensor) in a similar method as in the Einstein–Cartan extension of Moffat's theory (Einstein–Cartan–Moffat theory, see Ref. 5). Simultaneously, we introduce spin sources. We find equations for the gravitational and electromagnetic fields and the Cartan equation in this case. In Sec. IV we deal with the geodesic equation in the nonsymmetric Kaluza–Klein theory and consider the equation of motion for a charged test particle without spin in the theory.

I. ELEMENTS OF THE NONSYMMETRIC KALUZA–KLEIN THEORY

Let $\underset{\sim}{P}$ be an electromagnetic fiber bundle over space-time E with a projection π and a typical fiber U(1), and let α be an electromagnetic connection defined on $\underset{\sim}{P}$.

We introduce on $\underset{\sim}{P}$ a natural frame

$$\theta^A = (\pi^A(\bar{\theta}^\alpha), \theta^5 = \lambda\alpha), \tag{1.1}$$

where $\lambda = 2\sqrt{G}/c^2$ (G is a gravitational constant and c is the velocity of light in a vacuum). For every local section of P, e; we have

$$e^*\alpha = A_\mu \bar{\theta}^\mu, \tag{1.2}$$

where e: $P \rightarrow E$ and A_μ is a four-potential of electromagnetic field in a gauge e. For a curvature of α we have the following:

$$\Omega = d\alpha = \tfrac{1}{2}\pi^*(F_{\mu\nu}\bar{\theta}^\mu{}_\Lambda\bar{\theta}^\nu), \tag{1.3}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$

and we have Bianchi's identity

$$d\Omega = 0. \tag{1.4}$$

Due to (1.4), the four-potential exists. It is, of course, the first Maxwell equation. On the space-time E we define a nonsymmetric metric tensor $g_{\alpha\beta}$ such that

$$g_{\alpha\beta} = g_{(\alpha\beta)} + g_{[\alpha\beta]},$$
$$g_{\alpha\beta}\, g^{\gamma\beta} = g_{\beta\alpha} g^{\beta\gamma} = \delta^\gamma_\alpha, \tag{1.5}$$

where the order of indices is important. We define on E two connections $\bar{\tilde{\omega}}^\alpha_\beta$ and $\bar{\bar{\tilde{W}}}{}^\alpha_\beta$, where

$$\bar{\tilde{\omega}}^\alpha_\beta = (\bar{\bar{\tilde{\Gamma}}}{}^\alpha_{\beta\gamma})\bar{\theta}^\gamma$$

and

$$\bar{\bar{\tilde{W}}}{}^\alpha_\beta = \bar{\bar{\tilde{W}}}{}^\alpha_{\beta\gamma}\bar{\theta}^\gamma, \tag{1.6}$$

$$\bar{\bar{\tilde{W}}}{}^\alpha_\beta = \bar{\tilde{\omega}}^\alpha_\beta - \tfrac{2}{3}\delta^\alpha_\beta \bar{\bar{\tilde{W}}}, \tag{1.7}$$

where $\bar{\bar{\tilde{W}}} = \bar{\bar{\tilde{W}}}_\gamma \bar{\theta}^\gamma = \frac{1}{2}[\bar{\bar{\tilde{W}}}{}^\sigma_{\gamma\sigma} - \bar{\bar{\tilde{W}}}{}^\sigma{}_{\sigma\gamma}]\bar{\theta}^\gamma$. For the connection $\bar{\tilde{\omega}}^\alpha{}_\beta$, we suppose the following condition:

$$\bar{\bar{\tilde{Q}}}{}^\alpha_{\beta\alpha}(\bar{\bar{\tilde{\Gamma}}}) = 0, \tag{1.8}$$

where $\bar{\bar{\tilde{Q}}}{}^\alpha_{\beta\gamma}(\bar{\bar{\tilde{\Gamma}}})$ is the torsion of $\bar{\tilde{\omega}}^\alpha_\beta$. In Ref. 1, we also assume

$$\bar{\bar{\tilde{D}}}g_{\alpha+\beta-} = \bar{\bar{\tilde{D}}}g_{\alpha\beta} - g_{\alpha\delta}\bar{\bar{\tilde{Q}}}{}^\delta_{\beta\gamma}(\bar{\bar{\tilde{\Gamma}}})\bar{\theta}^\gamma = 0, \tag{1.9}$$

where $\bar{\bar{\tilde{D}}}$ is an exterior covariant derivative with respect to $\bar{\tilde{\omega}}^\alpha{}_\beta$.

Condition (1.9) is satisfied in Moffat's theory of gravitation if the fermion current is zero (see Ref. 4), and (1.9) is satisfied by the other connection called $\bar{\bar{\tilde{\Lambda}}}{}^\alpha_\beta = \bar{\bar{\tilde{\Lambda}}}{}^\alpha_{\beta\gamma}\bar{\theta}^\gamma$ if the fermion current is not zero. In Refs. 1 and 6, we introduce a natural nonsymmetric metrization of the fiber bundle:

$$\bar{\gamma} = \pi^*\bar{g} - \theta^5 \otimes \theta^5 = \pi^*(g_{(\alpha\beta)}\bar{\theta}^\alpha \otimes \bar{\theta}^\beta) - \theta^5 \otimes \theta^5,$$
$$\underset{\sim}{\gamma} = \pi^* g = \pi^*(g_{[\alpha\beta]}\bar{\theta}^\alpha \Lambda \bar{\theta}^\beta), \tag{1.10}$$

or

$$\gamma_{AB} = \left(\begin{array}{c|c} g_{\alpha\beta} & 0 \\ \hline 0 & -1 \end{array}\right). \tag{1.11}$$

We introduce, in Ref. 1, two connections on P, $\tilde{\omega}^A{}_B$, and $\tilde{W}^A_B$, invariant with respect to the action of the group U(1). Now the connection $\tilde{W}^A_B$ does not satisfy the compatibility condition for γ_{AB}, but a different connection $\tilde{\Lambda}^A_B = \tilde{\Lambda}^A_{BC}\theta^C$ satisfies the condition

---

[a)]On leave of absence from the Institute of Philosophy and Sociology of Polish Academy of Sciences, 00-330 Warsaw, Nowy Świat 72, Poland.

 0022-2488/84/041045-05$02.50

$$\widetilde{D}_{\gamma_{A+B}} = \widetilde{D}_{\gamma_{AB}} - \gamma_{AD}\widetilde{Q}^{D}{}_{BC}(\widetilde{\Lambda})\theta^{C} = 0, \qquad (1.12)$$

$$\mathop{\mathscr{L}}_{\xi_5} \widetilde{\Lambda}^{A}_{B} = 0, \qquad (1.12a)$$

where $\widetilde{D}$ is the exterior covariant derivative with respect to the connection $\widetilde{\Lambda}^{A}_{B}$, and $\widetilde{Q}^{D}_{BC}(\widetilde{\Lambda})$ is the tensor of torsion for the connection $\widetilde{\Lambda}^{A}_{B}$. One easily finds

$$\widetilde{\Lambda}^{A}_{B} = \left(\begin{array}{c|c} \pi^*(\bar{\bar{\Lambda}}{}^{\alpha}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^{5} & H_{\beta\gamma}\theta^{\gamma} \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right), \qquad (1.13)$$

where we have, for $\bar{\bar{\Lambda}}{}^{\alpha}_{\beta\gamma}$,

$$g_{\mu\nu,\sigma} - g_{\rho\nu}\bar{\bar{\Lambda}}{}^{\rho}_{\mu\sigma} - g_{\mu\rho}\bar{\bar{\Lambda}}{}^{\rho}_{\sigma\nu} = 0. \qquad (1.14)$$

For $\widetilde{W}^{A}_{B}$ we have, in a similar manner as in Ref. 1,

$$\widetilde{W}^{A}_{B} = \left(\begin{array}{c|c} \pi^*(\bar{\bar{W}}{}^{\alpha}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^{5} & H_{\beta\gamma}\theta^{\gamma} \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right). \qquad (1.15)$$

$H_{\beta\gamma} = -H_{\gamma\beta}$ is a tensor on E and satisfies the following condition:

$$g_{\delta\beta}g^{\gamma\delta}H_{\gamma\alpha} + g_{\alpha\delta}g^{\delta\gamma}H_{\beta\gamma} = 2g_{\alpha\delta}g^{\delta\gamma}F_{\beta\gamma} \qquad (1.16)$$

(see Ref. 1).

For the connection $\widetilde{\omega}^{A}_{B}$, we have the following:

$$\widetilde{\omega}^{A}_{B} = \left(\begin{array}{c|c} \pi^*(\bar{\bar{\omega}}{}^{\alpha}{}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^{5} & H_{\beta\gamma}\theta^{\alpha} \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right). \qquad (1.17)$$

Thus we have on $\underset{\sim}{P}$ all five-dimensional analogs of quantities from Moffat's theory of gravitation (see Refs. 2–4), i.e., $\widetilde{W}^{A}_{B}$, $\widetilde{\omega}^{A}_{B}$, $\widetilde{\Lambda}^{A}_{B}$, and γ_{AB}.

In Ref. 1 we calculate the Moffat–Ricci curvature scalar for $\widetilde{W}^{A}_{B}$:

$$R(\widetilde{W}) = \gamma^{AB}(R^{C}{}_{ABC}(\widetilde{W}) + \tfrac{1}{2}R^{C}{}_{CAB}(\widetilde{W})), \qquad (1.18)$$

where $R^{A}_{BCD}(\widetilde{W})$ is the tensor of curvature for the connection $\widetilde{W}^{A}_{B}$, and we get

$$R(\widetilde{W}) = \bar{\bar{R}}(\bar{\bar{W}}) + \frac{8\pi G}{c^4}\left(\frac{1}{8\pi}(2(g^{[\mu\nu]}F_{\mu\nu})^2 - H^{\mu\alpha}F_{\mu\alpha})\right). \qquad (1.19)$$

In (1.19), we come back to a normal system of physical units and we put in the place of $\lambda = 2$, $\lambda = 2\sqrt{G}/c^2$. We use, in Ref. 1, the system in which $G = c = 1$. $\bar{\bar{R}}(\bar{\bar{W}})$ is the Moffat–Ricci curvature scalar for the connection $\bar{\bar{W}}{}^{\alpha}_{\beta}$ (see Refs. 2 and 4) and

$$\begin{aligned}\bar{\bar{R}} &= g^{\alpha\beta}(\bar{\bar{R}}{}^{\gamma}{}_{\alpha\beta\gamma}(\bar{\bar{W}}) + \tfrac{1}{2}\bar{\bar{R}}{}^{\gamma}{}_{\gamma\alpha\beta}(\bar{\bar{W}})) \\ &= \bar{\bar{R}}(\bar{\bar{\Gamma}}) + \tfrac{2}{3}g^{[\alpha\beta]}\bar{\bar{W}}_{[\alpha,\beta]},\end{aligned} \qquad (1.20)$$

where $\bar{\bar{R}}{}^{\alpha}{}_{\beta\gamma\delta}(\bar{\bar{W}})$ is the tensor of curvature for the connection $\bar{\bar{W}}{}^{\alpha}_{\beta}$, $\bar{\bar{W}}_{[\alpha,\beta]} = \tfrac{1}{2}(\bar{\bar{W}}_{\alpha,\beta} - \bar{\bar{W}}_{\beta,\alpha})$, and

$$\bar{\bar{R}}(\bar{\bar{\Gamma}}) = g_{\alpha\beta}(\bar{\bar{R}}{}^{\gamma}{}_{\alpha\beta\gamma}(\bar{\bar{\Gamma}}) + \tfrac{1}{2}\bar{\bar{R}}{}^{\gamma}{}_{\gamma\alpha\beta}(\bar{\bar{\Gamma}}))$$

is the Moffat–Ricci curvature scalar for the connection $\bar{\bar{\omega}}{}^{\alpha}_{\beta}$ [$\bar{\bar{R}}{}^{\alpha}{}_{\beta\gamma\delta}(\bar{\bar{\Gamma}})$ is the curvature tensor for the connection $\bar{\bar{\omega}}{}^{\alpha}_{\beta}$].

II. MATERIAL SOURCES. PALATINI VARIATIONAL PRINCIPLE AND FIELD EQUATIONS

Let us introduce material sources: a tensor of energy–momentum $\underset{\sim}{T}^{\mu\nu}$, a fermion current $\underset{\sim}{S}^{\mu}$, an electric current $\underset{\sim}{j}^{\alpha}$, and a phenomenological Lagrangian of material sources

$$L_m = -\left(\frac{8\pi G}{c^4}\right)g^{\mu\nu}\underset{\sim}{T}_{\mu\nu} + \left(\frac{8\pi a^2}{3}\right)\bar{\bar{W}}_{\mu}\underset{\sim}{S}^{\mu} + \frac{4\pi}{c}\underset{\sim}{j}^{\mu}A_{\mu}, \qquad (2.1)$$

where

$$\begin{aligned}\underset{\sim}{T}_{\mu\nu} &= -\left(\frac{c^4}{8\pi G}\right)\frac{\delta L_m}{\delta g^{\mu\nu}}, \\ \underset{\sim}{S}^{\mu} &= \left(\frac{3}{8\pi a^2}\right)\frac{\delta L_m}{\delta \bar{\bar{W}}_{\mu}}, \\ \underset{\sim}{j}^{\mu} &= \left(\frac{c}{4\pi}\right)\frac{\delta L_m}{\delta A_{\mu}},\end{aligned} \qquad (2.2)$$

and A_{μ} is a four-potential of an electromagnetic field. We assume that L_m is gauge invariant. This means that $\underset{\sim}{j}^{\mu}$ is conserved:

$$\partial^{\mu}\underset{\sim}{j}^{\mu} = 0. \qquad (2.3)$$

Let us define the Palatini variational principle on the manifold $\underset{\sim}{P}$ for the density $(\sqrt{\gamma}R(\bar{W}) + L_m)$ as

$$\delta\int_v(\sqrt{\gamma}R(\bar{W}) - L_m)d^5x = 0, \quad v \subset \underset{\sim}{P}, \qquad (2.4)$$

where $\gamma = \det(\gamma_{AB}) = -\det(g_{\alpha\beta}) = -\underset{\sim}{g}$. We vary with respect to the independent quantities $g_{\alpha\beta}$, $\bar{\bar{W}}{}^{\alpha}_{\beta\alpha}$, and A_{μ}. After simple calculations one gets

$$\bar{\bar{R}}_{\alpha\beta}(\bar{\bar{W}}) - \tfrac{1}{2}g_{\alpha\beta}\bar{\bar{R}}(\bar{\bar{W}}) = \frac{8\pi G}{c^4}\left(\overset{\text{em}}{T}_{\alpha\beta} + \underset{\sim}{T}_{\alpha\beta}\right), \qquad (2.5)$$

$$\underset{\sim}{g}^{[\mu\nu]}{}_{,\nu} = 4\pi a^2\underset{\sim}{S}^{\mu}, \qquad (2.6)$$

$$g_{\nu\mu,\sigma} - g_{\rho\nu}\bar{\bar{\Lambda}}{}^{\rho}_{\mu\sigma} - g_{\mu\rho}\bar{\bar{\Lambda}}{}^{\rho}_{\sigma\nu} = 0, \qquad (2.7)$$

$$\begin{aligned}\partial_{\mu}\underset{\sim}{H}^{\alpha\mu} = \frac{4\pi}{c}\Big(&\underset{\sim}{j}^{\alpha} + 4a^2c\underset{\sim}{S}^{\alpha}(g^{[\mu\nu]}F_{\mu\nu}) \\ &+ \frac{c}{\pi}\underset{\sim}{g}^{[\alpha\beta]}\partial_{\beta}(g^{[\mu\nu]}F_{\mu\nu})\Big),\end{aligned} \qquad (2.8)$$

where

$$\begin{aligned}\overset{\text{em}}{T}_{\alpha\beta} = \frac{1}{4\pi}&(g^{\gamma\mu}H_{\gamma\alpha}F_{\mu\beta} - 2g^{[\mu\nu]}F_{\mu\nu}F_{\alpha\beta} \\ &(-\tfrac{1}{4}g_{\alpha\beta}(H^{\mu\nu}F_{\mu\nu} - 2(g^{[\mu\nu]}F_{\mu\nu})^2)),\end{aligned} \qquad (2.9)$$

$$\underset{\sim}{g}^{[\mu\nu]} = \sqrt{-g}\,g^{[\mu\nu]}, \quad \underset{\sim}{H}^{\mu\alpha} = \sqrt{-g}\,g^{\beta\mu}g^{\gamma\alpha}H_{\beta\gamma}, \qquad (2.10)$$

and

$$g_{\delta\beta}g^{\gamma\delta}H_{\gamma\alpha} + g_{\alpha\delta}g^{\delta\gamma}H_{\beta\gamma} = 2g_{\alpha\delta}g^{\delta\gamma}F_{\beta\gamma}. \qquad (2.11)$$

$\bar{\bar{\Lambda}}{}^{\rho}_{\mu\sigma}$ is the connection from Moffat's theory of gravitation (see Ref. 4) and

$$\bar{\bar{\Lambda}}{}^{\rho}_{\mu\sigma} = \bar{\bar{\Gamma}}{}^{\rho}_{\mu\sigma} + D^{\rho}_{\mu\sigma}(S), \qquad (2.12)$$

where

$$\begin{aligned}&g_{\rho\nu}D^{\rho}_{\mu\sigma} + g_{\mu\rho}D^{\rho}_{\sigma\nu} \\ &= (\tfrac{4}{3}\pi a^2)S^{\rho}(g_{\mu\sigma}g_{\rho\nu} - g_{\mu\rho}g_{\sigma\nu} + g_{\mu\nu}g_{[\sigma\rho]}).\end{aligned} \qquad (2.13)$$

Equations (2.5) and (2.6) are equations for the gravitational field in the presence of material and electromagnetic sources. $\overset{\text{em}}{T}_{\alpha\beta}$ plays the role of an energy–momentum tensor for the electromagnetic field. Equation (2.7) is a compatibility condition for the metric on space-time, and it is usually satisfied in Moffat's theory of gravitation (see Ref. 4) if the fermion current is not zero. Equation (2.8) plays the role of the second Maxwell equation. Now we have, on the right-hand side of (2.8), a sum of three currents: $\underset{\sim}{j}^{\alpha}$, $(c/\pi)g^{[\alpha\beta]}\partial_\beta(g^{[\mu\nu]}F_{\mu\nu})$, and $4\pi a^2\underset{\sim}{S}^{\alpha}(g^{[\mu\nu]}F_{\mu\nu})$. The first is a current of external sources, the second is that known from the nonsymmetric Kaluza–Klein theory (see Ref. 1), and the third is induced by the fermion current. The total electric current

$$\overset{\text{tot}}{\underset{\sim}{j}^{\alpha}} = \underset{\sim}{j}^{\alpha} + (c/\pi)g^{[\alpha\beta]}\partial_\beta(g^{[\mu\nu]}F_{\mu\nu}) + 4\pi a^2\underset{\sim}{S}^{\alpha}(g^{[\mu\nu]}F_{\mu\nu}) \tag{2.14}$$

is conserved:

$$\partial_\alpha \overset{\text{tot}}{\underset{\sim}{j}^{\alpha}} = 0. \tag{2.15}$$

In a similar manner as in Ref. 1, we have

$$g^{\alpha\beta}\overset{\text{em}}{T}_{\alpha\beta} = 0. \tag{2.16}$$

$H_{\alpha\beta}$ plays the role of the second tensor of the electromagnetic strength, and Eq. (2.11) expresses the relationship between both tensors $F_{\alpha\beta}$ and $H_{\alpha\beta}$ (see Ref. 1 for more details).

Let us define the tensor of the electromagnetic polarization $M_{\alpha\beta}$,

$$H_{\alpha\beta} = F_{\alpha\beta} - (4\pi/c)M_{\alpha\beta}. \tag{2.17}$$

It is easy to see that

$$Q^{5}_{\ \alpha\beta}(\widetilde{\Gamma}) = Q^{5}_{\ \alpha\beta}(\widetilde{\Lambda}) = (8\pi/c)M_{\alpha\beta}, \tag{2.18}$$

where $Q^{5}_{\ \alpha\beta}(\widetilde{\Gamma})$ is a tensor of torsion in the fifth dimension for the connection $\widetilde{\omega}^A_B$ and $Q^{5}_{\ \alpha\beta}(\widetilde{\Lambda})$ is a tensor of torsion in the fifth dimension for the connection $\widetilde{\Lambda}^A_B$. For the connection $\widetilde{\Lambda}^A_B$ we have the compatibility condition (1.12). Thus we get (as in Ref. 1) the compatibility condition for $\widetilde{\Lambda}^A_B$ and an interpretation of the electromagnetic polarization as the torsion in the fifth dimension for the connection $\widetilde{\Lambda}^A_B$. If $\underset{\sim}{S}^{\alpha} = 0$, we get $\widetilde{\omega}^A_B = \widetilde{\Lambda}^A_B$.

III. SPIN SOURCES

Let us introduce spin sources to the phenomenological Lagrangian (2.1). To do this, we define on E (as in Ref. 5) two connections $\overline{W}$ and $\overline{\Gamma}$:

$$\overline{W}^{\alpha}_{\beta\gamma} = \overline{\overline{W}}^{\alpha}_{\beta\gamma} + \kappa^{\alpha}_{\beta\gamma}, \tag{3.1}$$

$$\overline{\Gamma}^{\alpha}_{\beta\alpha} = \overline{\overline{\Gamma}}^{\alpha}_{\beta\gamma} + \kappa^{\alpha}_{\beta\gamma}, \tag{3.2}$$

where $\kappa^{\alpha}_{\beta\gamma}$ is a tensor field such as

$$\kappa^{\alpha}_{\beta\alpha} = 0, \quad \kappa^{\alpha}_{\beta\gamma} = -\kappa^{\alpha}_{\gamma\beta}. \tag{3.3}$$

$\kappa^{\alpha}_{\beta\gamma}$ plays the role of a generalized contortion tensor from Einstein–Cartan theory.

It is easy to see that

$$\overline{W}^{\lambda}_{\mu\nu} = \overline{\Gamma}^{\lambda}_{\mu\nu} - \tfrac{2}{3}\delta^{\lambda}_{\mu}\overline{W}_{\nu}, \tag{3.4}$$

where $\overline{W}_\nu = \frac{1}{2}(\overline{W}^{\sigma}{}_{\nu\sigma} - \overline{W}^{\sigma}_{\sigma\nu}) = \frac{1}{2}(\overline{\overline{W}}^{\sigma}{}_{\nu\sigma} - \overline{\overline{W}}^{\sigma}{}_{\sigma\nu}) = \overline{\overline{W}}_\nu$. We have

$$\overline{Q}^{\lambda}{}_{\mu\lambda}(\overline{\Gamma}) = \overline{\overline{Q}}^{\lambda}{}_{\mu\lambda}(\overline{\overline{\Gamma}}) = 0, \tag{3.5}$$

where $\overline{Q}^{\lambda}{}_{\mu\nu}(\overline{\Gamma})$ is a tensor of torsion for the connection $\overline{\Gamma}^{\lambda}_{\mu\nu}$ and $\overline{\overline{Q}}^{\lambda}_{\mu\nu}(\overline{\overline{\Gamma}})$ is a tensor of torsion for the connection $\overline{\overline{\Gamma}}^{\lambda}_{\mu\nu}$. (See Ref. 5 for more details.)

Let us define connections W^A_B and ω^A_B on P such that

$$W^A_B = \left(\begin{array}{c|c} \pi^*(\overline{W}^{\alpha}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^5 & H_{\beta\gamma}\theta^{\gamma} \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right), \tag{3.6}$$

$$\omega^A_B = \left(\begin{array}{c|c} \pi^*(\overline{\omega}^{\alpha}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^5 & H_{\beta\gamma}\theta^{\gamma} \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right), \tag{3.7}$$

where

$$\overline{W}^{\alpha}_{\beta} = \overline{W}^{\alpha}_{\beta\gamma}\theta^{\gamma}, \quad \overline{\omega}^{\alpha}_{\beta} = \overline{\Gamma}^{\alpha}_{\beta\gamma}\theta^{\gamma}. \tag{3.8}$$

We also define the third connection

$$\Lambda^A_B = \left(\begin{array}{c|c} \pi^*(\Omega^{\alpha}_{\beta}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^5 & H_{\beta\gamma}\theta^{\gamma} \\ \hline g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right), \tag{3.9}$$

where $\Omega^{\alpha}_{\beta} = \Omega^{\alpha}_{\beta\gamma}\theta^{\gamma}$ is a connection on space-time E such that

$$g_{\mu\nu,\sigma} - g_{\rho\nu}\Omega^{\rho}_{\mu\sigma} - g_{\mu\rho}\Omega^{\rho}_{\sigma\nu} = 0. \tag{3.10}$$

It is easy to see that Λ^A_B satisfies the compatibility condition (1.12). Using formulas from Refs. 1 and 5, one easily finds the Moffat–Ricci curvature scalar for the connection W^A_B.

$$\begin{aligned} R(W) &= \overline{R}(\overline{W}) + \frac{8\pi G}{c^4}\left[\frac{1}{8\pi}\left(2(g^{[\mu\nu]}F_{\mu\nu})^2 - H^{\mu\alpha}F_{\mu\alpha}\right)\right] \\ &= \overline{\overline{R}}(\overline{\Gamma}) + g^{\mu\alpha}\kappa^{\beta}_{\ \mu}{}^{+}{}_{+\alpha-;\beta} - g^{\mu\alpha}\kappa^{\beta}{}_{\gamma\alpha}\kappa^{\gamma}{}_{\mu\beta} \\ &\quad + \frac{2}{3}g^{[\mu\alpha]}W_{[\mu,\alpha]} \\ &\quad + \frac{8\pi G}{c^4}\left[\frac{1}{8\pi}(2(g^{[\mu\nu]}F_{\mu\nu})^2 - H^{\mu\alpha}F_{\mu\alpha})\right], \end{aligned} \tag{3.11}$$

where $\overline{R}(\overline{W})$ is the Moffat–Ricci curvature scalar for the connection $\overline{W}^{\alpha}_{\beta\gamma}$ and $\overline{\overline{R}}(\overline{\overline{\Gamma}})$ is the Moffat–Ricci curvature scalar for the connection $\overline{\overline{\Gamma}}^{\alpha}_{\beta\gamma}$. Let us define the Lagrangian for material sources such that

$$L'_m = L_m + \left(\frac{8\pi G}{c^3}\right)\overline{W}^{\sigma}_{\mu\nu}\underset{\sim}{S}^{\mu\nu}_{\sigma} \tag{3.12}$$

[see (2.1)], where we put in place of $\overline{\overline{W}}_\mu$, a vector $\overline{W}_\mu$ which is really equal to $\overline{\overline{W}}_\mu$. We have (2.2) for L'_m and

$$\underset{\sim}{S}^{\mu\nu}_{\sigma} = \left(\frac{c^2}{8\pi G}\right)\frac{\delta L'_m}{\delta\overline{W}^{\sigma}_{\mu\nu}}, \quad \underset{\sim}{S}^{\mu\nu}_{\sigma} = -\underset{\sim}{S}^{\mu\nu}_{\sigma}. \tag{3.13}$$

For $\underset{\sim}{S}^{\mu}$, we have

$$\underset{\sim}{S}^{\mu} = \left(\frac{3}{8\pi a^2}\right)\frac{\delta L'_m}{\delta\overline{W}_\mu} = \left(\frac{3}{8\pi a^2}\right)\frac{\delta L_m}{\delta\overline{\overline{W}}_\mu}. \tag{3.14}$$

Let us define the Palatini-variational principle on the manifold $\underset{\sim}{P}$:

$$\delta\int_{\upsilon}(L'_m - \sqrt{\gamma}R(W))d^5x = 0, \quad \upsilon\subset P. \tag{3.15}$$

We vary with respect to the independent quantities $g_{\mu\nu}$, $\overline{W}^{\lambda}_{\mu\nu}$, and A_μ. After some calculations one gets

$$\overline{R}_{\mu\nu}(\overline{W}) - \frac{1}{2}g_{\mu\nu}\overline{R}(\overline{W}) = \frac{8\pi G}{c^4}\left(\overset{\text{em}}{T}_{\mu\nu} + T_{\mu\nu}\right), \tag{3.16}$$

$$g^{[\mu\nu]}{}_{,\nu} = 4\pi\Big(a^2 \underset{\sim}{S}{}^{\mu} - \frac{2G}{c^3} \underset{\sim}{S}{}^{\mu\nu}_{\ \ \nu}\Big) = 4\pi a^2 \underset{\sim}{K}{}^{\mu}, \tag{3.17}$$

$$g_{\nu\mu,\sigma} - g_{\rho\nu}\widetilde{\widetilde{\Lambda}}{}^{\rho}_{\mu\sigma} - g_{\mu\rho}\widetilde{\widetilde{\Lambda}}{}^{\rho}_{\sigma\nu} = \Big(g_{\rho\nu}\,\kappa^{\rho}_{\mu\sigma} + g_{\mu\rho}\,\kappa^{\rho}_{\sigma\nu} + \frac{8\pi G}{c^3} g_{\rho\nu} g_{\mu\gamma}\, S^{\rho\gamma}_{\ \ \sigma}\Big), \tag{3.18}$$

$$\partial_\mu \underset{\sim}{H}{}^{\alpha\mu} = \frac{4\pi}{c}\Big(\underset{\sim}{j}{}^{\alpha} + 4a^2 c \underset{\sim}{K}{}^{\alpha}(g^{[\mu\nu]}F_{\mu\nu}) + \frac{c}{\pi} g^{[\alpha\beta]}\partial_\beta (g^{[\mu\nu]}F_{\mu\nu})\Big), \tag{3.19}$$

where $\overset{\text{em}}{T}_{\mu\nu}$ is defined by (2.9), $\underset{\sim}{H}{}^{\alpha\mu}$ by (2.10), and $\widetilde{\Lambda}{}^{\lambda}_{\mu\nu}$ by (2.12) and (2.13). To be in line with the usual interpretation of the Moffat compatibility condition, we suppose that

$$g_{\nu\mu,\sigma} - g_{\rho\nu}\widetilde{\widetilde{\Lambda}}{}^{\rho}_{\mu\sigma} - g_{\mu\rho}\widetilde{\widetilde{\Lambda}}{}^{\rho}_{\sigma\nu} = 0, \tag{3.20}$$

and we get

$$g_{\rho\nu}\,\kappa^{\rho}_{\mu\sigma} + g_{\mu\rho}\,\kappa^{\rho}_{\sigma\nu} = -(8\pi G/c^3) g_{\rho\nu} g_{\mu\gamma}\, S^{\rho\gamma}_{\ \ \sigma}, \tag{3.21}$$

i.e., a generalization of the Cartan equation from the Einstein–Cartan–Moffat theory (see Ref. 5).

Equations (3.17) and (3.19) differ from the analogous equation of Sec. II [(2.6) and (2.8)]. The tensorial density $\underset{\sim}{S}{}^{\mu\nu}_{\ \ \sigma}$ is a spin density, and for a microscopic spin density (of a Dirac field or Rarita–Schwinger field), we have

$$\underset{\sim}{S}{}^{\mu\nu}_{\ \ \nu} = 0. \tag{3.22}$$

In the case of Mathisson spin (hydrodynamic macroscopic spin) one easily checks that

$$\underset{\sim}{S}{}^{\mu\nu}_{\ \ \nu} = 0, \tag{3.23}$$

where $\underset{\sim}{S}{}^{\mu\nu}_{\ \ \sigma} = u_\sigma \underset{\sim}{S}{}^{\mu\nu}$, $u_\nu \underset{\sim}{S}{}^{\mu\nu} = 0$, $\underset{\sim}{S}{}^{\mu\nu} = -\underset{\sim}{S}{}^{\nu\mu}$, u_σ is a four-velocity of fluid, and $\underset{\sim}{S}{}^{\mu\nu}$ is a spin–density tensor in the rest frame. Thus we get

$$g^{[\mu\nu]}{}_{,\nu} = 4\pi a^2 \underset{\sim}{S}{}^{\mu}, \tag{3.24}$$

$$\partial_\mu \underset{\sim}{H}{}^{\alpha\mu} = \frac{4\pi}{c}\overset{\text{tot}}{\underset{\sim}{j}}{}^{\alpha}, \tag{3.25}$$

where $\overset{\text{tot}}{\underset{\sim}{j}}{}^{\alpha}$ is defined by (2.14).

Using Eqs. (1.10) and (1.13) from Ref. 5, one transforms (3.16) into

$$\widetilde{\widetilde{R}}_{\mu\alpha}(\widetilde{\widetilde{W}}) - \frac{1}{2} g_{\mu\alpha}\widetilde{\widetilde{R}}(\widetilde{\widetilde{W}}) = \frac{8\pi G}{c^4}\Big(\overset{\text{em}}{T}_{\mu\alpha} + \overset{\text{eff}}{T}_{\mu\alpha}\Big), \tag{3.26}$$

where

$$\overset{\text{eff}}{T}_{\mu\alpha} = T_{\mu\alpha} - \Big(\frac{c^4}{8\pi G}\Big)\big(\kappa^{\beta}{}_{\alpha}{}^{+}{}_{-\mu+;\beta} - \kappa^{\beta}{}_{\alpha\gamma}\,\kappa^{\gamma}{}_{\mu\beta} - \tfrac{1}{2} g_{\mu\alpha}\, g^{\gamma\nu}(\kappa^{\beta}{}_{\nu}{}^{+}{}_{-\gamma+;\beta} - \kappa^{\beta}{}_{\nu\delta}\,\kappa^{\delta}{}_{\gamma\beta})\big), \tag{3.27}$$

and $\kappa^{\alpha}{}_{\beta\gamma}$ is defined by (3.21). Thus we get spin–spin interaction corrections from Einstein–Cartan–Moffat theory.

Now it is easy to see that

$$\Omega^{\lambda}_{\ \mu\nu} = \widetilde{\widetilde{\Lambda}}{}^{\lambda}_{\mu\nu} \tag{3.28}$$

and

$$\Lambda^{A}_{\ B} = \widetilde{\Lambda}{}^{A}_{\ B}, \tag{3.29}$$

and the connection $\widetilde{\Lambda}{}^{A}_{\ B}$ satisfies the compatibility condition for γ_{AB} on the manifold $\underset{\sim}{P}$. For the polarization tensor $M_{\alpha\beta}$, we have the same geometrical interpretation as in Sec. II [see Eq. (2.19)].

IV. GEODESICS

In the nonsymmetric Kaluza–Klein theory[1] we have the following equation for geodesics:

$$\frac{\widetilde{\widetilde{D}}u^{\alpha}}{d\tau} + 2u^5(g^{\alpha\gamma}F_{\gamma\beta} + g^{[\alpha\gamma]}H_{\gamma\beta})u^{\beta} = 0, \tag{4.1}$$

$$u^5 = \text{const} \quad (2u^5 = q/m_0),$$

where q is a charge and m_0 is the rest mass of a test particle. $\widetilde{\widetilde{D}}/d\tau$ means covariant derivative with respect to $\widetilde{\widetilde{\omega}}{}^{\alpha}_{\beta}$ along a curve to which $u^{\alpha}(\tau)$ is tangent.

The usual interpretation of the geodesic equation in the Kaluza–Klein theory is that (4.1) is an equation of motion for a test particle in the gravitational and electromagnetic fields.

If we have nonzero fermion current $\underset{\sim}{S}{}^{\alpha} \neq 0$, it is necessary to put in the place of $\widetilde{\widetilde{\omega}}{}^{\alpha}_{\beta}$ the connection $\widetilde{\Lambda}{}^{\alpha}_{\ \beta}$ and we get, in the holonomic system of coordinates,

$$m_0\Big(\frac{d^2x^{\alpha}}{d\tau^2} + \widetilde{\widetilde{\Lambda}}{}^{\alpha}_{(\beta\gamma)}\frac{dx^{\beta}}{d\tau}\frac{dx^{\gamma}}{d\tau}\Big) + q(g^{\alpha\gamma}F_{\gamma\beta} + g^{[\alpha\gamma]}H_{\gamma\beta})\frac{dx^{\beta}}{d\tau} = 0. \tag{4.2}$$

Connection $\widetilde{\Lambda}{}^{A}_{\ B}$ is compatible with the metric γ_{AB}. In Moffat's theory, this kind of geodesic is called a nonextremal geodesic. However, in Moffat's theory, particles move along different geodesics (see Ref. 4), i.e.,

$$\frac{d^2x^{\alpha}}{dt^2} + \left\{{\alpha \atop \beta\gamma}\right\}\frac{dx^{\beta}}{dt}\frac{dx^{\gamma}}{dt} = 0. \tag{4.3}$$

Thus we should put in place of $\widetilde{\widetilde{\Lambda}}{}^{\alpha}{}_{(\beta\gamma)}$ the Christoffel symbol $\left\{{\alpha \atop \beta\gamma}\right\}$ for $g_{(\alpha\beta)}$. Finally we get

$$m_0\Big(\frac{d^2x^{\alpha}}{dt^2} + \left\{{\alpha \atop \beta\gamma}\right\}\frac{dx^{\beta}}{dt}\frac{dx^{\gamma}}{dt}\Big) + q(g^{\alpha\gamma}F_{\gamma\beta} + g^{[\alpha\gamma]}H_{\gamma\beta})\frac{dx^{\beta}}{d\tau} = 0. \tag{4.4}$$

We will consider Eq. (4.4) the equation of motion for a test particle in the nonsymmetric Kaluza–Klein theory. The connection

$$\hat{\omega}^{A}_{B} = \left(\begin{array}{c|c} \pi^*(\left\{{\alpha \atop \beta\gamma}\right\}\theta^{\gamma}) + g^{\gamma\alpha}H_{\gamma\beta}\theta^5 & H_{\beta\gamma}\theta^{\gamma} \\ g^{\alpha\beta}(H_{\gamma\beta} + 2F_{\beta\gamma})\theta^{\gamma} & 0 \end{array}\right) \tag{4.5}$$

is not compatible with the metric γ_{AB} on $\underset{\sim}{P}$, just as connection $\widetilde{\widetilde{\omega}}{}^{\alpha}_{\beta} = \left\{{\alpha \atop \beta\gamma}\right\}\theta^{\gamma}$ is not compatible with $g_{\alpha\beta}$ on E. In the theory with spin sources, particles without spin and fermion charge move along geodesics in $\hat{\omega}^{A}_{B}$ (as was supposed in Ref. 5). The problem of motion for spinning particles with fermion charge demands more investigation.

ACKNOWLEDGMENTS

I would like to thank Professor J. W. Moffat, Dr. G. Kunstatter, and Dr. R. B. Mann for their kind hospitality and numerous extremely valuable discussions and advice during my stay at the Physics Department of the University of Toronto.

[1]M. W. Kalinowski, "The nonsymmetric Kaluza–Klein theory," J. Math. Phys. **24**, 1834 (1983).
[2]J. W. Moffat, "New theory of Gravitation," Phys. Rev. D **19**, 3557 (1979).
[3]J. W. Moffat, "Gauge invariance and string interactions in a generalized theory of gravitation," Phys Rev. D **23**, 2870 (1981).
[4]J. W. Moffat, "Generalized theory of gravitation and its physical consequences," in *Proceedings of the VII International School of Gravitation and Cosmology*, Erice, Sicily, edited by de Sabatta (World Scientific Publishing, Singapore, 1982), p. 127.
[5]M. W. Kalinowski, "An Einstein–Cartan–Moffat theory," Phys. Rev. D **26**, 3419 (1983).
[6]M. W. Kalinowski, "The nonsymmetric-nonabelian Kaluza–Klein theory," J. Phys. A **16**, 1669 (1983).

Diffusion processes, quantum dynamical semigroups, and the classical KMS condition

Alberto Frigerio and Vittorio Gorini
Dipartimento di Fisica, Milano, Italy, and INFN, Sezione di Milano, Milano, Italy[a)]

(Received 18 January 1983; accepted for publication 22 July 1983)

We study some properties of the generator of a diffusion process describing the reduced dynamics of a classical system weakly coupled to an external reservoir. We show that such processes may be equivalently obtained as classical limits of quantum dynamical semigroups derived in the weak coupling limit. We prove that the stationary state is independent of the coupling if and only if the reservoir is in a state of thermal equilibrium.

PACS numbers: 05.20.Gg, 02.50.Ga

1. INTRODUCTION

In a preceding paper,[1] a definition of detailed balance for a dynamical semigroup was investigated. For the case of a quantum dynamical semigroup describing the reduced evolution of a spatially confined open quantum system S in the weak coupling limit, it was shown that the property of detailed balance is closely related to the condition that the reservoir R to which S is coupled be in thermal equilibrium (the KMS condition). Precisely, if the reference state $\rho_{\rm ref}$ of the reservoir R is such that the reduced dynamics of any spatially confined quantum system S weakly coupled to R satisfies detailed balance with respect to some faithful normal state ρ (depending, *a priori*, on the coupling of S to R) for any admissible coupling, then $\rho_{\rm ref}$ satisfies the KMS condition at some inverse temperature β, and ρ is the canonical state $e^{-\beta H^s}/\mathrm{tr}\, e^{-\beta H^s}$. It was also shown that, if detailed balance is not assumed, but the stationary state ρ is required to be independent of the coupling of S to R, then detailed balance follows, $\rho_{\rm ref}$ is KMS, and ρ is canonical (see also Ref. 2).

In this paper we are concerned with the problem of the extent to which analogous results hold for classical systems weakly coupled to classical reservoirs. "The idea of the weak coupling limit is that, by some central limit effect, very many, but weak collisions should lead to a diffusion type evolution" (quoted from Spohn,[3] p. 574); i.e., to what is called a Fokker–Planck equation in the physical literature (see, e.g., Ref. 4). However, a diffusion semigroup always satisfies detailed balance (in the sense of Ref. 1) with respect to any of its stationary states (up to technicalities; see Sec. 2); this circumstance has been also noted by Graham.[5] Hence detailed balance in the weak coupling limit cannot discriminate between equilibrium and nonequilibrium stationary states.

A more restrictive definition of detailed balance has been given in the literature with reference to transformation properties of the Fokker–Planck equation under time reversal[6,7]; this is indeed a condition, and not a general property, but it may be also satisfied in certain stationary states far from thermal equilibrium (Refs. 6, 7 and Sec. 5 below).

On the other hand, we still expect on physical grounds that the requirement of independence of the stationary state ρ of the coupling of S to R will imply the KMS condition for R also in the classical case. This requirement can be regarded as the condition that R behaves as a *heat bath* with respect to S (compare Ref. 2), and is indeed a stability condition for the state $\rho\rho_{\rm ref}$ of $S+R$; stability (plus some technical assumptions) is known to imply the KMS condition both in the quantum and in the classical case.[8–10]

In order to substantiate our claim that the heat bath property should imply the KMS condition in the classical case, we ought to give a rigorous treatment of the weak coupling limit for classical system + reservoir models. The difficulty of this problem can be appreciated by comparison with the amount of technical apparatus which was needed to deal with the simpler problem of a classical particle in a weak stochastic force field (Kesten and Papanicolaou[11]). So we have confined ourselves (Sec. 3) to a formal application of the weak coupling limit theory of Davies,[12] which is rigorous in the case of bounded perturbations. This not being the case for classical systems, we have overlooked all the problems of convergence of series and interchange of limits; we believe, notwithstanding, that our result is essentially correct. We feel comforted in this by the fact that the same result is obtained by taking the formal classical limit $\hbar \to 0$ of the generator of a quantum dynamical semigroup obtained in the weak coupling limit (Sec. 4).

We should warn the reader that our results hold only for classical systems for which the Liouville operator $\{\cdot, H^s\}$ has a pure point spectrum; we suspect that this property may be true only for those systems which can be identified with a system of independent harmonic oscillators by a canonical transformation (possibly up to physically pathological cases).

With all these qualifications, we compute the stationary state ρ of the reduced dynamics of a one-dimensional harmonic oscillator weakly coupled to a reservoir, and we prove that it is independent of the coupling if and only if the reference state $\rho_{\rm ref}$ satisfies the classical KMS condition (Sec. 6).

2. DETAILED BALANCE FOR DIFFUSION PROCESSES

We begin by recalling the definition of detailed balance for dynamical semigroups on W^*-algebras, which was given in Ref. 1.

Let $\mathscr{M}$ be a W^*-algebra, $\{\phi_t : t \in \mathbb{R}^+\}$ a dynamical se-

a) Postal address: Dipartimento di Fisica, Sezione di Fisica Teorica, via Celoria 16, 20133 Milano, Italy.

 0022-2488/84/041050-16$02.50

migroup on $\mathscr{M}$, ρ a faithful normal state on $\mathscr{M}$, which is invariant under $\{\phi_t\}$. Let $\mathscr{K}$ be the Hilbert space completion of $\mathscr{M}$ equipped with the inner product $\langle A,B\rangle = \rho(A^*B)$, $A, B \in \mathscr{M}$ (the GNS space); then $\{\phi_t: t\in\mathbb{R}^+\}$ extends to a strongly continuous contraction semigroup $\{\hat{\phi}_t = \exp \hat{L}t: t\in\mathbb{R}^+\}$ on $\mathscr{K}$.

We say that $\{\phi_t: t\in\mathbb{R}^+\}$ satisfies *detailed balance* with respect to ρ[1] (cf. Ref. 13) if there exists a self-adjoint operator $\hat{L}_s$ and a skew-adjoint operator $\hat{L}_h$ in $\mathscr{K}$ such that

(i) $\hat{L}\psi = \hat{L}_s\psi + \hat{L}_h\psi$ for all ψ in a common core for $\hat{L}$, $\hat{L}_s$, $\hat{L}_h$;

(ii) $\{\exp \hat{L}_h t: t\in\mathbb{R}\}$ maps $\mathscr{M}\subseteq\mathscr{K}$ into itself and defines a weakly $*$ continuous group $\{\alpha_t: t\in\mathbb{R}\}$ of $*$-automorphisms of $\mathscr{M}$;

(iii) $\{\exp \hat{L}_s t: t\in\mathbb{R}^+\}$ maps $\mathscr{M}\subseteq\mathscr{K}$ into itself and defines a dynamical semigroup $\{\Gamma_t: t\in\mathbb{R}^+\}$ on $\mathscr{M}$.

If a decomposition (i) exists, with $\hat{L}_s = \hat{L}_s^*$, $\hat{L}_h = -\hat{L}_h^*$, then it is unique, and it can be shown that (iii) is actually a consequence of (i) and (ii).[1] The state ρ is stationary under $\{\alpha_t\}$ and $\{\Gamma_t\}$, and satisfies the symmetry condition $\rho(\Gamma_t(A)^*B) = \rho(A^*\Gamma_t(B))$ for all A, B in $\mathscr{M}$, t in $\mathbb{R}^+$.

Now we consider diffusion in $\mathbb{R}^d$. We do not aim at full generality; in particular, we avoid the consideration of bounded regions and boundary conditions.

We denote $\mathbb{R}^d$ by X and the Lebesgue measure on X by dx. We let $\mathscr{A}$ be the $*$-algebra $C_0^\infty(X) + \mathbb{C}1$ and by $\mathscr{M}$ the W^*-algebra $L^\infty(X)$; for any strictly positive function ρ in $L^1(X)$, $L^\infty(X)$ coincides with $L^\infty(X,\rho)$; both $\mathscr{A}$ and $\mathscr{M}$ are dense in the Hilbert space $\mathscr{K} = L^2(X,\rho)$.

We shall be concerned with diffusion processes, i.e., Markov processes with almost everywhere continuous paths,[14] whose associated Markov semigroups $\{\phi_t: t\geqslant 0\}$ on $\mathscr{M}$ are dynamical semigroups [i.e., satisfy also $\phi_t(1) = 1$ and not just $\phi_t(1)\leqslant 1$ for all t]. Then typically, the weak $*$ derivative

$$\frac{d}{dt}\phi_t(A)\Big|_{t=0} = L(A) \tag{2.1}$$

exists for all A in $\mathscr{A}$, and one has

$$L(A) = \sum_{i,j=1}^d a_{ij}\,\partial_i\partial_j A + \sum_{j=1}^d b_j\,\partial_j A, \tag{2.2}$$

where a_{ij}, b_j, $i,j = 1,\dots,d$, are real-valued functions on X, $(a_{ij}(x))$ is a positive matrix for almost all x in X.

The adjoint equation to (2.1), (2.2), governing the time evolution of states $\rho(t)$ in $L^1(X)$, is given by

$$\frac{\partial}{\partial t}\rho(t) = \sum_{i,j=1}^d \partial_i\partial_j(a_{ij}\,\rho(t)) - \sum_{j=1}^d \partial_j(b_j\,\rho(t)), \tag{2.2$'$}$$

which is called the Fokker–Planck equation in the physical literature.[4–7]

We cannot summarize here the enormous literature concerning the characterization of the generators of the Markov semigroups associated with diffusion processes.[14] Given a map L of the form (2.2), with some continuity conditions on the coefficients a_{ij}, b_j, we want to find conditions under which there exists a dynamical semigroup $\{\phi_t: t\in\mathbb{R}^+\}$ satisfying (2.1), and a state ρ which is invariant under $\{\phi_t\}$, and show that $\{\phi_t\}$ satisfies detailed balance with respect to ρ.

Lemma 2.1: Let L be of the form (2.2), and assume $a_{ij}\in C^2(X)$, $b_j\in C^1(X)$, $i,j = 1,\dots,d$. Let ρ be a strictly positive function in $L^1(X)\cap C^2(X)$, with $\int_X \rho\,dx = 1$. Then there is a unique decomposition

$$L = L_s + L_h, \tag{2.3}$$

where L_s, L_h are linear maps from $\mathscr{A}$ into $\mathscr{M}$, satisfying

$$\int_X L_s(A)^*B\rho\,dx = \int_X A^*L_s(B)\rho\,dx \quad \text{for all } A, B \text{ in } \mathscr{A}, \tag{2.4}$$

$$L_h(A) = \sum_{j=1}^d v_j\partial_j A \quad \text{for all } A \text{ in } \mathscr{A}, \tag{2.5}$$

where v_j, $j = 1,\dots,d$, are real-valued functions in $C^1(X)$.

Proof: L_s must have the form

$$L_s(A) = \sum_{i,j=1}^d a_{ij}\,\partial_i\partial_j A + \sum_{j=1}^d (b_j - v_j)\partial_j A$$

for all A in $\mathscr{A}$;

then

$$\begin{aligned}\int_X L_s(A)^*B\rho\,dx = &-\sum_{i,j=1}^d \int_X (\partial_i A)^* a_{ij}(\partial_j B)\rho\,dx \\ &+ \sum_{j=1}^d \int_X \Big\{-\sum_{i=1}^d [(\partial_i a_{ij}) + (\rho^{-1}\partial_i\rho)a_{ij}] \\ &+ b_j - v_j\Big\}(\partial_j A)^*B\rho\,dx,\end{aligned} \tag{2.6}$$

where we have used integration by parts and the symmetry of the matrix (a_{ij}) (when A and B are multiples of 1, we might not be allowed to use integration by parts, but then the above equality reduces to $0 = 0$). Similarly,

$$\begin{aligned}\int_X A^*L_s(B)\rho\,dx = &-\sum_{i,j=1}^d \int_X (\partial_i A)^* a_{ij}(\partial_j B)\rho\,dx \\ &+ \sum_{j=1}^d \int_X \Big\{-\sum_{i=1}^d [(\partial_i a_{ij}) + (\rho^{-1}\partial_i\rho)a_{ij}] \\ &+ b_j - v_j\Big\}A^*(\partial_j B)\rho\,dx.\end{aligned}$$

Then condition (2.4) determines v_j uniquely to be

$$v_j = -\sum_{i=1}^d [(\partial_i a_{ij}) + (\rho^{-1}\partial_i\rho)a_{ij}] + b_j, \tag{2.7}$$

which is indeed a C^1 function for all $j = 1,\dots,d$; then

$$L_s(A) = \sum_{i,j=1}^d \{a_{ij}\partial_i\partial_j A + [(\partial_i a_{ij}) + (\rho^{-1}\partial_i\rho)a_{ij}]\partial_j A\} \quad \text{for all } A \text{ in } \mathscr{A}. \tag{2.8}$$

By the continuity of the coefficients, L_s and L_h map $\mathscr{A}$ into $\mathscr{M}$. ■

Lemma 2.2: Let L and ρ be as in Lemma 2.1. Then the following conditions are equivalent:

(i) $\int_X L(A)\rho\, dx = 0$ for all A in $\mathscr{A}$; (2.9)

(ii) $\sum_{i,j=1}^{d} \partial_i\, \partial_j(a_{ij}\,\rho) - \sum_{j=1}^{d} \partial_j(b_j\,\rho) = 0$;

(iii) $\sum_{j=1}^{d} \partial_j(v_j\,\rho) = 0$;

(iv) $\int_X L_h(A)\rho\, dx = 0$ for all A in $\mathscr{A}$;

(v) $\int_X L_h(A)^*B\rho\, dx = -\int_X A^*L_h(B)\rho\, dx$

for all A, B, in $\mathscr{A}$. (2.10)

Proof: (i) $\Leftrightarrow$ (ii) and (iii) $\Leftrightarrow$ (iv) are proved by using integration by parts for $A \in C_0^\infty(X)$ (for $A = \lambda\,1$, there is nothing to prove). Taking into account (2.7), we see that (ii) is just the explicit expression of (iii). From (2.5), we get

$L_h(A)^*B + A^*L_h(B) = L_h(A^*B)$ for all A, B in $\mathscr{A}$;

so it is clear that (iv) and (v) are equivalent. ■

Corollary 2.3: Let $\{\phi_t : t \in \mathbb{R}^+\}$ be a dynamical semigroup on $\mathscr{M}$ satisfying (2.1) and leaving some state ρ invariant; let L and ρ satisfy the conditions of Lemma 2.1. Then

$L = L_s + L_h$,

where L, L_s, L_h, defined on $\mathscr{A}$, are dissipative, negative symmetric, and skew-symmetric, respectively, as operators in $\mathscr{K} = L^2(X, \rho)$, and L_h is a $*$-derivation (a first-order differential operator with real coefficients).

Proof: Invariance of ρ under $\{\phi_t\}$ implies (2.9), which in turn implies (2.10), by Lemma 2.2. Then the statement follows, by Lemma 2.1. In particular, L_s is negative by (2.6) and (2.7), and L is dissipative since, by (2.10), Re $\int_X L(A)^*A\rho\, dx = \int_X L_s(A)^*A\rho\, dx$ for all A in $\mathscr{A}$.■

The conclusion of Corollary 2.3 amounts "almost" to detailed balance. A precise statement is:

Theorem 2.4: Let L, ρ be as in Lemma 2.1 and satisfy the equivalent conditions of Lemma 2.2. Denote by $\hat{L}, \hat{L}_s, \hat{L}_h$ the closures of L, L_s, L_h as operators in $\mathscr{K} = L^2(X, \rho)$. If $\hat{L}$, $\hat{L}_s$, and $\hat{L}_h$ are maximal dissipative, self-adjoint, and skew-adjoint, respectively, then there is a dynamical semigroup $\{\phi_t ; t \in \mathbb{R}^+\}$ on $\mathscr{M}$ such that (2.1) holds, and $\{\phi_t\}$ satisfies the detailed balance condition with respect to ρ.

Proof: The closures of L, L_s, and L_h are dissipative, negative symmetric, and skew-symmetric, respectively. Under the assumptions of the theorem, there exist contraction semigroups $\{T_t : t \in \mathbb{R}^+\}$, $\{V_t : t \in \mathbb{R}^+\}$ and a group of unitaries $\{U_t : t \in \mathbb{R}^+\}$, whose generators are $\hat{L}, \hat{L}_s, \hat{L}_h$, respectively, and T_t is the Lie–Trotter product of V_t and U_t.

By assumption, $\hat{L}_s$ is essentially self-adjoint on $\mathscr{A}$; hence $\hat{L}_s$ coincides with the Friedrichs extension of L_s, which is the generator of a positivity-preserving semigroup of self-adjoint contractions, by the general theory of symmetric Markov semigroups.[14] Then $V_t = \exp \hat{L}_s t$ preserves positivity, $V_t 1 = 1$ [since $1 \in \mathrm{Dom}(\hat{L}_s)$ and $\hat{L}_s(1) = 0$], and V_t maps $\mathscr{M}$ into itself for all $t \geqslant 0$; the restriction of $\{V_t : t \in \mathbb{R}^+\}$ to $\mathscr{M}$ is a dynamical semigroup.

Similarly, $\hat{L}_h$ is essentially skew-adjoint on $\mathscr{A}$ by assumption, and is a $*$-derivation by (2.5). Then it follows from Ref. 15 that $U_t = \exp L_h\, t$ maps $\mathscr{M}$ into itself for all t in $\mathbb{R}$, and the restriction of $\{U_t : t \in \mathbb{R}\}$ to $\mathscr{M}$ is a group of $*$-automorphisms of $\mathscr{M}$.

Finally, also $T_t = \lim_{n\to\infty} (U_{t/n} V_{t/n})^n$ maps $\mathscr{M}$ into itself for all $t \geqslant 0$, and the restriction of $\{T_t : t \in \mathbb{R}^+\}$ to $\mathscr{M}$ is a dynamical semigroup $\{\phi_t : t \in \mathbb{R}^+\}$ on $\mathscr{M}$ (cf. Ref. 1, Appendix B). The existence of the weak $*$ derivative $(d/dt)\, \phi_t(A)|_{t=0}$ for all A in $\mathscr{A}$ follows from the existence of the derivative $(d/dt)T_t(A)|_{t=0} = \hat{L}(A) = L(A)$ in $\mathscr{K}$; then also (2.1) holds.

By collecting the above results, we see that $\{\phi_t\}$ satisfies detailed balance with respect to ρ.■

We have seen that a dynamical semigroup on $\mathscr{M} = L^\infty(X)$ which is the Markov semigroup of a diffusion process in X satisfies detailed balance with respect to a faithful normal state ρ "if and only if" (up to technicalities) it leaves ρ invariant; in other words, detailed balance (in the sense of Ref. 1) is not a condition for diffusion processes, but a property. On the contrary, it is well known that detailed balance is indeed a condition for jump processes. Indeed, if L, defined by

$$L(A)(x) = \int_X W(y,x)[A(y) - A(x)]\, dy \quad [W(y,x) \geqslant 0] \tag{2.11}$$

is the (formal) generator of the Markov semigroup $\{\phi_t : t \in \mathbb{R}^+\}$ of a jump process, a state ρ is invariant under $\{\phi_t\}$ if and only if

$$\int_X [W(x,y)\rho(y) - W(y,x)\rho(x)]\, dy = 0 \quad \text{for all } x \text{ in } X, \tag{2.12}$$

in which case the adjoint L^+ of L as an operator in $L^2(X, \rho)$ is (formally) given by

$$L^+(A)(x) = \int_X W(x,y)\rho(y)\rho(x)^{-1}[A(y) - A(x)]\, dy. \tag{2.13}$$

Then

$$(L - L^+)(A)(x) = \int_X [W(y,x) - W(x,y)\rho(y)\rho(x)^{-1}] \times [A(y) - A(x)]\, dy, \tag{2.14}$$

and $L - L^+$ is not the generator of a dynamical semigroup, since the sign of $W(y,x) - W(x,y)\rho(y)\rho(x)^{-1}$ is not definite, unless, of course, $L = L^+$, or, equivalently,

$$W(x,y)\rho(y) = W(y,x)\rho(x) \quad \text{for all } x,y \text{ in } X, \tag{2.15}$$

which is the familiar form of the detailed balance condition for a jump process.

If $W(x,y)$ is vanishingly small for large $|x - y|$, so that $\int\int W(x,y)(x-y)^n\, dx\, dy$ may be neglected for $n > 2$, then L is approximated by a second-order differential operator and $L - L^+$ is approximated by a first-order differential operator:

$$\frac{1}{2}(L - L^+)(A) \approx \sum_{j=1}^{d} v_j\, \partial_j\, A, \tag{2.16}$$

where

$$v_j(x) = \tfrac{1}{2}\int_X \Big\{ [W(x+\tfrac{1}{2}y, x-\tfrac{1}{2}y) - W(x-\tfrac{1}{2}y, x+\tfrac{1}{2}y)] - W(x+\tfrac{1}{2}y, x-\tfrac{1}{2}y)\Big[\sum_{i=1}^{d}(\rho^{-1}\partial_i\rho)(x)y_i\Big]\Big\} y_j\, dy, \tag{2.17}$$

and

$$\tfrac{1}{2}(L+L^+)(A) \approx \sum_{i,j=1}^{d} \{a_{ij}\,\partial_i\partial_j A + [(\partial_i a_{ij}) + (\rho^{-1}\partial_i\rho)a_{ij}]\partial_j A\}, \tag{2.18}$$

where

$$a_{ij}(x) = \tfrac{1}{2}\int_X W(x+\tfrac{1}{2}y, x-\tfrac{1}{2}y)\, y_i\, y_j\, dy. \tag{2.19}$$

Then detailed balance holds, in this approximation, because $L - L^+$ becomes a $*$-derivation up to this order of approximation.

If L, given by (2.11), already satisfies detailed balance with respect to ρ, then its approximate expression is just (2.18), and contains no antisymmetric part. Detailed balance for a jump process means that, in the stationary state ρ, for any pair E, F of measurable subsets of X, there are as many transitions (jumps) per unit time from E to F as from F to E: this is the property of "balancing in detail." For a diffusion process, there can be transitions in an infinitesimal time interval only between adjacent regions: hence, in any stationary state, there are as many transitions from E to F as from F to E, in an infinitesimal time interval, if the boundary between E and F is a closed $(d-1)$-dimensional surface in X; one has "balancing in detail" only for transitions across a closed surface. It would be too restrictive to give a definition of detailed balance for a diffusion process, which would involve balancing in detail for transitions between any pair of regions, because this would rule out evolutions induced by flows; however, such a condition would be satisfied for a diffusion process without antisymmetric part.[5]

Given L and ρ which satisfy the equivalent conditions of Lemma 2.2, the mathematically hard part of the detailed balance condition is to check whether L_s is essentially self-adjoint, L_h is essentially skew-adjoint, and the closure of L is the generator of a contraction semigroup on $\mathscr{K}$. For applications in Sec. 5, we shall just suppose that this is the case, without going into the problem of proving these properties.

We remark here that this is the reason why we have excluded the consideration of bounded regions. For instance, let I be an open bounded interval of $\mathbb{R}$, and $L_s(A) = \partial^2 A$ for all A in $C_0^\infty(I) + \mathbb{C}1$. Then L_s has at least two self-adjoint extensions (with Neumann and with periodic boundary conditions). In order to have a unique self-adjoint extension, it would be necessary to choose a domain $\mathscr{A}$ for L_s which is strictly larger than $C_0^\infty(I) + \mathbb{C}1$, and this would force us to go into details which we wish to avoid here.

On the other hand, when X is the whole of $\mathbb{R}^d$, there are cases in which it is possible to prove that L_s and L_h are essentially self-adjoint and essentially skew-adjoint, respectively, with the aid of known theorems. Under certain conditions of differentiability and boundedness for the coefficients a_{ij}, b_j, the diffusion equation may be integrated and the domain of the generator of ϕ_t may be found[16]; in some cases it might be possible to check the core condition.

Suppose instead that a stationary state ρ can be found explicitly, and consider the unitary mapping $A \mapsto A\rho^{1/2}$ of $\mathscr{K}$ onto $L^2(X)$. Then L_s, L_h are unitarily equivalent to operators S, H, respectively, given by

$$Sf = \sum_{i,j=1}^{d} [(\partial_i a_{ij}\partial_j f) - (\rho^{-1/2}\,\partial_i a_{ij}\partial_j\,\rho^{1/2})f], \tag{2.20}$$

$$Hf = \sum_{j=1}^{d} [v_j\partial_j f + \tfrac{1}{2}(\partial_j v_j) f] \tag{2.21}$$

[we have used $\Sigma_{j=1}^{d}\partial_j(v_j\,\rho) = 0$] for all f in $\mathrm{Dom}(S) = \mathrm{Dom}(H) = [C_0^\infty(X) + \mathbb{C}]\,\rho^{1/2}$. If $\rho \in C^\infty(X)$, it is sufficient to prove that S, H are essentially self- (skew-) adjoint on $C_0^\infty(X)$. Now $-S$ looks like a Schrödinger operator and is a Schrödinger operator if $a_{ij}(x)$ is constant, whereas H looks like the generator of a classical Hamiltonian flow and is so if $d = 2n$, $v_j = -\partial h/\partial x_{j-n}$ for $j = n+1, \ldots 2n$, $v_j = \partial h/\partial x_{j+n}$ for $j = 1, \ldots, n$, where h is the Hamiltonian. Theorems giving sufficient conditions for the essential self- (skew-) adjointness of such operators on $C_0^\infty(\mathbb{R}^d)$ may be found, for instance, in Ref. 17.

In the physical literature, the detailed balance condition is often related to the behavior under time reversal. Suppose that there is a time reversal operation $\mathscr{T}$ on $\mathscr{M}$, defined by

$$(\mathscr{T}f)(x_1, \ldots, x_d) = f(\epsilon_1 x_1, \ldots, \epsilon_d\, x_d), \tag{2.22}$$

where $\epsilon_j = \pm 1$, $j = 1, \ldots, d$. Then the detailed balance condition as formulated by Graham and Haken[6] amounts, in our notation, to the condition that $\int_X L(A)\,\rho\, dx = 0$ for all A and

$$\mathscr{T}L_s\mathscr{T} = L_s, \quad \mathscr{T}L_h\mathscr{T} = -L_h, \tag{2.23}$$

i.e., that the symmetric and antisymmetric parts of L are, respectively, the even and odd part of L under time reversal; this is indeed a condition. When $\{\phi_t\}$ describes the reduced dynamics of a subsystem of a Hamiltonian system, detailed balance in the form (2.23) is generally "derived" as a consequence of time reversal invariance of the global dynamics[6]; we shall come back to this point in Sec. 5.

3. THE WEAK COUPLING LIMIT FOR CLASSICAL SYSTEMS

In this section we are concerned with the derivation of a diffusion equation in phase space as describing the reduced dynamics of a classical Hamiltonian system weakly coupled to classical Hamiltonian reservoirs. To our knowledge, there are no rigorous results concerning this problem. The only related results that we are aware of are the proof of the convergence of the weak coupling limit for a particle in a stochastic force field (Kesten and Papanicolaou[11]) and of the Brownian motion limit for a heavy particle in a gas of mutually noninteracting light particles (Holley,[18] and Dürr, Goldstein, and Lebowitz[19]). Hence we shall confine ourselves to a formal application of the weak coupling limit theory of Davies.[12] The price we have to pay is that we have to restrict our consideration to classical systems whose Liouville operator has pure point spectrum; we suspect that the only physically interesting systems enjoying this property are canonically equivalent to systems of uncoupled harmonic oscillators.

To be specific, we consider a particle with position q and momentum p, which interacts with particles with positions q_j and momenta p_j. We assume a total (formal) Hamiltonian

$$H_\lambda = \frac{1}{2m} p^2 + U(q) + \lambda \sum_j V(q - q_j) + \sum_j \frac{1}{2m_j} p_j{}^2 + \frac{1}{2} \sum_{i,j} W(q_i - q_j), \tag{3.1}$$

which is of the form $H_\lambda = H^S + \lambda H^{SR} + H^R$, the superscripts S and R denoting "system" and "reservoir," respectively. Let $x = (q, p)$ and $y = (\{q_1,...,q_n,...\}, \{p_1,..., p_n,...\})$. The equations of motion for the system variables, with initial conditions $x(0) = x_0$, $y(0) = y_0$, are

$$\begin{aligned} \frac{d}{dt} q(x_0, y_0, t) &= \frac{1}{m} p(x_0, y_0, t), \\ \frac{d}{dt} p(x_0, y_0, t) &= -U'(q(x_0, y_0, t)) \\ &\quad - \lambda \sum_j V'(q(x_0, y_0, t) - q_j(x_0, y_0, t)) \end{aligned} \tag{3.2}$$

(we are using a one-dimensional notation). These equations are insufficient for the determination of q,p, since also $q_j(x_0, y_0, t)$ are dynamical variables. Only in the limit of infinitely heavy reservoir particles do Eq. (3.2) determine q and p as a function of the initial conditions x_0, y_0 and of time. When y_0 is treated as a random variable, we are led back to the problem of a particle in a random force field.[11]

Here we shall consider the same kind of weak coupling limit as was studied by Davies in Ref. 12 for the quantum case. We let $\lambda \to 0$, with constant rescaled time $\tau = \lambda^2 t$, and we pass to the interaction picture in order to remove the effects of the uncoupled motion. So we study the limit as $\lambda \to 0$ of a probability distribution

$$\rho^\lambda(q, p, \tau) = \mathscr{U}^0_{-\tau/\lambda^2} \iint \mathscr{U}^\lambda_{\tau/\lambda^2} \times [\rho(q, p, 0) \rho_{\text{ref}}(\{q_j\}, \{p_j\}, 0)] \prod_j dq_j \, dp_j, \tag{3.3}$$

where $\mathscr{U}^\lambda_t$ is the solution of $(d/dt)\, \mathscr{U}^\lambda_t f = \{H_\lambda, \mathscr{U}^\lambda_t f\}$, $\mathscr{U}^\lambda_0 f = f$, and ρ_{ref} is chosen to be invariant under $\mathscr{U}^0_t$. Since $\lambda\{H^{SR},\cdot\}$ is not a bounded perturbation of $\{H_0,\cdot\}$, the estimates of Ref. 12 do not apply. For harmonic oscillators coupled to harmonic reservoirs by a bilinear interaction, the problem can be reduced to a problem of bounded linear operators on an underlying "test function space," and precise estimates can be given (see Ref. 20.).

A formal application of the techniques of Ref. 12 leads to the following statement. For any function $A = A(q, p)$ in $L^\infty(\mathbb{R}^{2f})$ and any $\rho = \rho(q, p, 0)$ in $L^1(\mathbb{R}^{2f})$, we have, for all $\tau \geqslant 0$,

$$\lim_{\lambda \to 0} \iint \rho^\lambda(q, p, \tau) A(q, p) \, dq \, dp = \iint \rho(q, p, 0) \, \phi_\tau(A)(q, p) \, dq \, dp, \tag{3.4}$$

and $\{\phi_\tau : \tau \in \mathbb{R}^+\}$ is a dynamical semigroup, whose formal generator L is given by

$$L(A) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \mathscr{U}^0_{-s} \times \left[\int_0^\infty \left(\iint \{H^{SR}, \{H^{SR}(t), \mathscr{U}^0_s A\}\} \times \rho_{\text{ref}} \prod_j dq_j \, dp_j \right) dt \right] ds, \tag{3.5}$$

where $H^{SR}(t) = \mathscr{U}^0_{-t} H^{SR}$ (cf. Refs. 12 and 21). In order to ensure the existence of the time average, we assume that $\{H^S,\cdot\}$ has a pure point spectrum. Specifically, we assume that there are smooth functions $V_k = V_k(q, p)$, $k \in \mathbb{Z}$, such that

$$\{H^S, V_k\} = i\omega_k V_k, \quad k \in \mathbb{Z}, \tag{3.6}$$

with $\omega_{-k} = -\omega_k$, $V_{-k} = V_k^*$, and that H^{SR} may be expanded as

$$H^{SR} = \sum_{k \in Z} V_k V_k^R, \tag{3.7}$$

where V_k^R are functions of the reservoir coordinates (and momenta), satisfying $V^R_{-k} = V_k^{R*}$; then

$$H^{SR}(t) = \sum_{k \in Z} e^{-i\omega_k t} V_k V_k^R(t). \tag{3.8}$$

We put

$$\psi_{kk'}(t) = \iint V^R_{-k'} V_k^R(t) \rho_{\text{ref}} \prod_j dq_j \, dp_j \equiv \langle V^R_{-k'} V_k^R(t) \rangle, \tag{3.9}$$

$$\varphi_{kk'}(t) = \iint \{V^R_{-k'}, V_k^R(t)\} \rho_{\text{ref}} \prod_j dq_j \, dp_j \equiv \langle \{V^R_{k'}, V_k^R(t)\} \rangle, \tag{3.10}$$

and we assume $\langle V_k^R \rangle = 0$; then we obtain

$$L(A) \equiv \int_0^\infty \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} e^{-i\omega_k t} (\psi_{kk'}(t) \{V_{-k'}, \{V_k, A\}\} + \varphi_{kk'}(t) V_{-k'} \{V_k, A\}) \, dt. \tag{3.11}$$

Remark: Although we have used a notation appropriate for reservoirs of finitely many particles, we should really think of the reservoir as being infinitely extended. This does not constitute a problem for such model reservoirs as infinite harmonic crystals or gases of particles interacting through suitable potentials, for which the time evolution of local observables relative to certain stationary states, can be defined.[20,22] Note that the definition of the Poisson bracket in (3.10) actually involves only a finite sum, if V_k^R is assumed to depend only on a finite number of reservoir coordinates and momenta.

Proposition 3.1: Suppose that there is only a finite number of functions $\psi_{kk'}$, $\varphi_{kk'}$, which do not vanish identically, and that these functions are in $L^1(\mathbb{R})$. Then $L(A)$, given by (3.11), exists for all A in $\mathscr{A} = C_0^\infty(\mathbb{R}^{2f}) + \mathbb{C}1$, and

$$L(A) = \{A,\delta H\} + \frac{1}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} (\widehat{\psi}_{kk'}(\omega_k)\{V_{-k'},\{V_k,A\}\} + \widehat{\varphi}_{kk'}(\omega_k) V_{-k'}\{V_k,A\}) \tag{3.12}$$

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

$$\widehat{\psi}_{kk'}(\omega) = \int_{-\infty}^{\infty} e^{-i\omega t}\psi_{kk'}(t)\,dt = \widehat{\psi}_{-k',-k}(-\omega), \tag{3.13}$$

$$\widehat{\varphi}_{kk'}(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t}\varphi_{kk'}(t)\,dt = -\varphi_{-k',-k}(-\omega), \tag{3.14}$$

and

$$\delta H = \frac{1}{2}\int_0^\infty \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} e^{-i\omega_k t} \times (\psi_{kk'}(t)\{V_k,V_{-k'}\} - \varphi_{kk'}(t) V_k V_{-k'})\,dt. \tag{3.15}$$

Proof: We have, using the invariance of ρ_{ref} under $\mathscr{U}_t^0$,

$$\psi_{kk'}(t) = \langle V^R_{-k'} V^R_k(t)\rangle = \langle V^R_{-k'}(-t) V^R_k\rangle = \langle V^R_k V^R_{-k'}(-t)\rangle = \psi_{-k',-k}(-t), \tag{3.13'}$$

$$\varphi_{kk'}(t) = \langle\{V^R_{-k'},V^R_k(t)\}\rangle = \langle\{V^R_{-k'}(-t),V^R_k\}\rangle = -\langle\{V^R_k,V^R_{-k'}(-t)\}\rangle = -\varphi_{-k',-k}(-t), \tag{3.14'}$$

from which (3.13), (3.14) follow. Then we introduce the formal generator $\widetilde{L}$ of the reduced dynamics for negative times, given by

$$\widetilde{L}(A) = \int_{-\infty}^0 \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} e^{-i\omega_k t}(\psi_{kk'}(t)\{V_{-k'},\{V_k,A\}\} + \varphi_{kk'}(t)V_{-k'}\{V_k,A\})\,dt$$

and compute L as $\frac{1}{2}(L + \widetilde{L}) + \frac{1}{2}(L - \widetilde{L})$. Clearly, $\frac{1}{2}(L + \widetilde{L})$ yields the second term on the right-hand side of (3.12). We also have

$$\frac{1}{2}(L - \widetilde{L})(A) = \frac{1}{2}\int_0^\infty \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [(e^{-i\omega_k t}\psi_{kk'}(t) - e^{i\omega_k t}\psi_{kk'}(-t)) \times \{V_{-k'},\{V_k,A\}\} + (e^{-i\omega_k t}\varphi_{kk'}(t) - e^{i\omega_k t}\varphi_{kk'}(-t)) \times V_{-k'}\{V_k,A\}]\,dt.$$

In the part containing $\widetilde{L}$ we operate the replacements $k \to -k$, $k' \to -k'$, we use $\omega_{-k} = -\omega_k$ and relations (3.13'), (3.14'), to get

$$\frac{1}{2}(L - \widetilde{L})(A) = \frac{1}{2}\int_0^\infty \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} e^{-i\omega_k t}(\psi_{kk'}(t)[\{V_{-k'},\{V_k,A\}\} - \{V_k,\{V_{-k'},A\}\}] + \varphi_{kk'}(t) \times [V_{-k'}\{V_k,A\} + V_k\{V_{-k'},A\}])\,dt.$$

Using the formal properties of the Poisson bracket and the definition (3.15) of δH, we finally obtain

$$\frac{1}{2}(L - \widetilde{L})(A) = \{A,\delta H\}. \quad \blacksquare$$

Remark: Of course, the expression (3.12) of L is formally valid also when there are infinitely many functions $\psi_{kk'}$, $\varphi_{kk'}$ which do not vanish identically.

Remark: L commutes with $\{\cdot,H\}$ and $\{\delta H,H\} = 0$.

Proposition 3.2: L has the form (2.2)

$$L(A) = \sum_{i,j=1}^{2f} a_{ij}\,\partial_i\,\partial_j\,A + \sum_{j=1}^{2f} b_j\,\partial_j\,A$$

for all A in $\mathscr{A}$, (2.2)

where the positive symmetric matrix (a_{ij}) is given by

$$a_{ij} = \frac{1}{2}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \widehat{\psi}_{kk'}(\omega_k)\bar{u}_{k'i}u_{kj}, \quad i,j = 1,\ldots,2f, \tag{3.16}$$

and the real vector b_j is given by

$$b_j = \sum_{i=1}^{2f} \partial_i a_{ij} + \frac{1}{2}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \widehat{\varphi}_{kk'}(\omega_k)V_{-k'}u_{kj} + w_j,$$

$$j = 1,\ldots,1f, \tag{3.17}$$

and where

$$u_{kj} = \begin{cases} \partial_{j-f}V_k, & j = f+1,\ldots,2f, \\ -\partial_{j+f}V_k, & j = 1,\ldots,f, \end{cases} \tag{3.18}$$

$$w_j = \begin{cases} -\partial_{j-f}\,\delta H, & j = f+1,\ldots,2f, \\ \partial_{j+f}\,\delta H, & j = 1,\ldots,f. \end{cases} \tag{3.19}$$

Proof: By (3.18), (3.19), we have for all A in $\mathscr{A}$

$$\{V_k,A\} = \sum_{j=1}^{2f} u_{kj}\,\partial_j\,A,$$

$$\{A,\delta H\} = \sum_{j=1}^{2f} w_j\,\partial_j\,A,$$

hence (3.12) can be rewritten as

$$L(A) = \sum_{j=1}^{2f} w_j\partial_j A + \frac{1}{2}\sum_{i,j=1}^{2f}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [\widehat{\psi}_{kk'}(\omega_k)\bar{u}_{k'i}\partial_i(u_{kj}\partial_j A) + \widehat{\varphi}_{kk'}(\omega_k)V_{-k'}u_{kj}\partial_j A].$$

Now

$$\bar{u}_{k'i}\,\partial_i(u_{kj}\,\partial_j\,A) = \bar{u}_{k'i}u_{kj}\,\partial_i\partial_j\,A + \partial_i(\bar{u}_{k'i}u_{kj})\partial_j\,A - (\partial_i\bar{u}_{k'i})u_{kj}\,\partial_j\,A,$$

and

$$\sum_{i=1}^{2f}\partial_i\bar{u}_{k'i} = \sum_{i=1}^{f}(-\partial_i\partial_{i+f}\,V_{-k'}) + \sum_{i=f+1}^{2f}\partial_i\partial_{j-f}V_{-k'} = 0;$$

hence (2.2) follows, upon defining a_{ij} and b_j by (3.16) and (3.17), respectively.

By construction, L maps real functions into real functions; hence a_{ij} and b_j are real. Moreover, $(a_{ij}(x))$ is a positive matrix for all x, since

$$\hat{\psi}_{kk'}(\omega_k)\delta_{\omega_k,\omega_{k'}} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} e^{-i\omega_k s}\left(\int_{-\infty}^{+\infty} e^{-i\omega_k t}\langle V_{-k'}\, V_k(t)\rangle\, dt\right)e^{i\omega_k s}ds$$

is a positive matrix, by the stationarity of $\rho_{\rm ref}$.■

Now we introduce time reversal invariance. For a classical Hamiltonian system, the time reversal operation $\mathscr{T}$ consists in leaving all coordinates q unchanged and reversing the signs of all momenta p.

Proposition 3.3: Suppose that the global dynamics $\mathscr{U}_t^\lambda$ is time reversal invariant, i.e., $\mathscr{T}\mathscr{U}_t^\lambda\mathscr{T} = \mathscr{U}_{-t}^\lambda$, for all λ, and that also $\mathscr{T}\rho_{\rm ref} = \rho_{\rm ref}$. Then $\{\cdot,\delta H\}$ is precisely the odd part of L under time reversal.

Proof: We have seen that $\{\cdot,\delta H\} = \frac{1}{2}(L - \widetilde{L})$. When $\mathscr{U}_t^\lambda$ and $\rho_{\rm ref}$ are time reversal invariant, we have $\widetilde{L} = \mathscr{T}L\mathscr{T}$; hence the result follows (see Ref. 21).■

4. THE CLASSICAL LIMIT OF QUANTUM DYNAMICAL SEMIGROUPS

As an independent check of the results of the previous section, we consider here the classical limit of the generator of a quantum dynamical semigroup obtained in the weak coupling limit. We perform the classical limit in a formal sense, according to the prescriptions

$$(i\hbar)^{-1}[A,B] \underset{\hbar\to 0}{\longrightarrow} \{A,B\},\quad \tfrac{1}{2}[A,B]_+ \underset{\hbar\to 0}{\longrightarrow} AB = BA, \qquad (4.1)$$

where A, B are linear operators on the Hilbert space $\mathscr{H} = L^2(\mathbb{R}^f)$, expressed as $\hbar$-independent, sufficiently smooth functions of the operators $Q_j = x_j\cdot$, $P_j = -i\hbar\,\partial/\partial x_j$, $j = 1,\dots,f$.

We shall not be concerned here with the difficult problem of the precise meaning of the limits (4.1) for large classes of operators. A method to tackle this problem involves the use of the Weyl correspondence $\mathbb{Q}$:

$$\mathbb{Q}f = \frac{1}{2\pi}\iint \hat{f}(u,v)\exp[i(uQ + vP)]\,du\,dv, \qquad (4.2)$$

where $\hat{f}$ is the Fourier transform of f. See Ref. 23 for some classes of operators which are of the form $\mathbb{Q}f$ for suitable functions or distributions f. A situation in which the limits (4.1) can be shown to hold rigorously with the help of (4.2) is $A = \mathbb{Q}f$, $B = \mathbb{Q}g$, where f,g are in $\mathscr{S}(\mathbb{R}^{2f})$: then A and B are trace class operators, and

$$(i\hbar)^{-1}[\mathbb{Q}f,\mathbb{Q}g] = \mathbb{Q}h_\hbar,\quad \tfrac{1}{2}[\mathbb{Q}f,\mathbb{Q}g]_+ = \mathbb{Q}k_\hbar, \qquad (4.3)$$

where

$$h_\hbar(q,p) = \frac{1}{(2\pi)^2}\iiiint \frac{2}{\hbar}\sin\left[\frac{\hbar}{2}(u'v - uv')\right] \times e^{i(uq+vp)}\hat{f}(u,v)e^{i(u'q+v'p)}\,\hat{g}(u',v')\,du\,dv\,du'\,dv', \qquad (4.4)$$

and $k_\hbar(q,p)$ has a similar expression, with sin replaced by cos. When $\hbar\to 0$, $h_\hbar(q,p)$, and $k_\hbar(q,p)$ tend to $\{f,g\}(q,p)$ and to $(fg)(q,p)$, respectively.

We point out that this is *not* the kind of classical limit that was studied by Hepp in Ref. 24 and applied to the generators of dynamical semigroups by Pulé and Verbeure[25]: the latter limit consists of letting $\hbar\to 0$ in the expectation values of functions of $\hbar^{1/2}x_j$, $-i\hbar^{1/2}\partial/\partial x_j$ in coherent states centered around positions $\hbar^{-1/2}q_j$ and wave numbers $\hbar^{-1/2}p_j$, respectively.

Let $\mathscr{H}$ be the Hilbert space $L^2(\mathbb{R}^f)$. We consider the limit as $\hbar\to 0$ of an expression

$$L_\hbar(A) = \frac{i}{\hbar}[H,A] + \frac{1}{\hbar^2}\sum_r f_r(\hbar) \times(V_r^* A V_r - \frac{1}{2}[V_r^* V_r, A]_+), \qquad (4.5)$$

where A is a bounded operator on $\mathscr{H}$, H is a self-adjoint operator on $\mathscr{H}$, V_r are linear operators on $\mathscr{H}$, and $f_r(\hbar)$ are positive numbers, depending continuously on $\hbar$. When H, V_r are bounded operators and the series converges ultraweakly, (4.5) in the general form of the generator of a norm-continuous dynamical semigroup on $\mathscr{B}(\mathscr{H})$ (Lindblad[26]; see Davies[27] for what is known for unbounded operators H and V_r). Here we shall disregard all domain problems and proceed formally. The explicit $\hbar$ dependence appearing in (4.5) is what occurs in applications to open systems. We assume that H and V_r are $\hbar$-independent smooth functions of $Q_1,\dots,Q_f$, $P_1,\dots,P_f$ (for example, that they are obtained from $\hbar$-independent smooth functions by means of the Weyl correspondence). We put

$$V_r = X_r + iY_r,\quad X_r = X_r^*,\ Y_r = Y_r^*,$$

and we get

$$L_\hbar(A) = \frac{1}{i\hbar}[A,H] + \frac{1}{2(i\hbar)^2}\sum_r f_r(\hbar) \times([X_r,[X_r,A]] + [Y_r,[Y_r,A]]) + \frac{1}{2i\hbar}\sum_r \frac{1}{\hbar}f_r(\hbar)([X_r,[Y_r,A]]_+ - [Y_r,[X_r,A]]_+). \qquad (4.6)$$

The formal classical limit of (4.6), according to the prescriptions (4.1), is

$$L_{\rm cl}(A) = \{A,H\} + \frac{1}{2}\sum_r f_r(0) \times(\{X_r,\{X_r,A\}\} + \{Y_r,\{Y_r,A\}\}) + \lim_{\hbar\to 0}\sum_r \frac{1}{\hbar}f_r(\hbar)(X_r\{Y_r,A\} - Y_r\{X_r,A\}). \qquad (4.7)$$

When all V_r are self-adjoint (i.e., all Y_r vanish identically), we have simply

$$L_{cl}(A) = \{A,H\} + \frac{1}{2}\sum_r f_r(0)\{X_r,\{X_r,A\}\} \quad (Y_r = 0). \quad (4.8)$$

In general, when the Y_r and $f_r(0)$ do not vanish identically, the third term on the right-hand side of (4.7) is divergent. We discuss here a meaningful situation in which a finite limit is obtained.

Proposition 4.1: Let $L_\hbar$ be of the form (4.5)

$$L_\hbar(A) = \frac{i}{\hbar}[H,A] + \frac{1}{\hbar^2}\sum_{r\in\mathbb{Z}} f_r(\hbar) \times (V_r{}^*AV_r - \frac{1}{2}[V_r{}^*V_r,A]_+),$$

where

$$V_r{}^* = V_{-r}, \quad (4.9)$$

$$f_r(\hbar) - f_{-r}(\hbar) = O(\hbar), \quad r\in\mathbb{Z}, \quad (4.10)$$

and where H, V_r are smooth functions of $Q_1,...,Q_f$, $P_1,...,P_f$, not depending explicitly on $\hbar$. Then the formal classical limit of $L_\hbar$ is given by

$$L_\hbar(A) \underset{\hbar\to 0}{\to} L_{cl}(A),$$

where

$$L_{cl}(A) = \{A,H\} + \tfrac{1}{2}f_0(0)\{V_0,\{V_0,A\}\} + \sum_{r=1}^{\infty} f_r(0)(\{X_r,\{X_r,A\}\} + \{Y_r,\{Y_r,A\}\}) + \sum_{r=1}^{\infty}\lim_{\hbar\to 0}[\hbar^{-1}(f_r(\hbar) - f_{-r}(\hbar))] \times (X_r\{Y_r,A\} - Y_r\{X_r,A\}), \quad (4.11)$$

and L_{cl} is the formal generator of the Markov semigroup of a diffusion process.

Proof: By assumption (4.9), the expression (4.7) becomes (4.11), where $\lim_{\hbar\to 0}\hbar^{-1}(f_r(\hbar) - f_{-r}(\hbar))$ is finite, by assumption (4.10). Now we put L_{cl} in the form (2.2). Define

$$\xi_{rj} = \begin{cases} \partial_{j-f}X_r, & j = f+1,...,2f \\ -\partial_{j+f}X_r, & j = 1,...,f, \end{cases} \quad (4.12)$$

$$\eta_{rj} = \begin{cases} \partial_{j-f}Y_r, & j = f+1,...,2f \\ -\partial_{j+f}Y_r, & j = 1,...,f, \end{cases} \quad (4.13)$$

$$\zeta_j = \begin{cases} -\partial_{j-f}H, & j = f+1,...,2f \\ \partial_{j+f}H, & j = 1,...,f. \end{cases} \quad (4.14)$$

Then, we get

$$\begin{aligned} L_{cl}(A) &= \sum_{j=1}^{2f}\zeta_j\,\partial_j A + \frac{1}{2}f_0(0)\sum_{i,j=1}^{2f}\xi_{0i}\,\partial_i(\xi_{0j}\,\partial_j A) \\ &\quad + \sum_{r=1}^{\infty}\sum_{i,j=1}^{2f} f_r(0)[\xi_{ri}\partial_i(\xi_{rj}\,\partial_j A) \\ &\quad + \eta_{ri}\,\partial_i(\eta_{rj}\,\partial_j A)] + \sum_{r=1}^{\infty}\sum_{j=1}^{2f}\lim_{\hbar\to 0}[\hbar^{-1}(f_r(\hbar) \\ &\quad - f_{-r}(\hbar))][X_r\,\eta_{rj} - Y_r\xi_{rj}]\partial_j A \\ &= \sum_{i,j=1}^{2f}a_{ij}\,\partial_i\partial_j A + \sum_{j=1}^{2f}b_j\,\partial_j A, \end{aligned}$$

where

$$a_{ij} = \frac{1}{2}f_0(0)\xi_{0i}\xi_{0j} + \sum_{r=1}^{\infty}f_r(0)(\xi_{ri}\xi_{rj} + \eta_{ri}\,\eta_{rj}) \quad (4.15)$$

and

$$b_j = \sum_{i=1}^{2f}\partial_i a_{ij} + \zeta_j + \lim_{\hbar\to 0}[\hbar^{-1}(f_r(\hbar) - f_{-r}(\hbar))](X_r\eta_{rj} - Y_r\,\xi_{rj}) \quad (4.16)$$

(we have used $\Sigma_{i=1}^{2f}\,\partial_i\xi_{ri} = \Sigma_{i=1}^{2f}\,\partial_i\,\eta_{ri} = 0$ as in the proof of Proposition 3.2). Then (a_{ij}) is a positive matrix and b_j in a real vector, and L_{cl} has the form (2.2); hence it is the formal generator of the Markov semigroup of a diffusion process. ■

The generator of a dynamical semigroup describing the reduced dynamics of a quantum open system in the weak coupling limit[12] is indeed of the form described in Proposition 4.1. Let S be a quantum system coupled to a reservoir R, let $H_\lambda = H^S + H^R + \lambda H^{SR}$ be the total Hamiltonian, and assume that the Hamiltonian H^S of the uncoupled system S has a pure point spectrum. Expand the interaction Hamiltonian λH^{SR} as

$$\lambda H^{SR} = \lambda\sum_{k\in\mathbb{Z}} V_k \otimes V_k^R$$

(possibly, only a finite number of V_k may be different from zero), where

$$[H^S,V_k] = -\hbar\omega_k\,V_k, \quad k\in\mathbb{Z},$$

and where

$$\omega_{-k} = -\omega_k, \quad V_{-k} = V_k{}^*, V_{-k}^R = V_k^{R*}, \quad k\in\mathbb{Z}.$$

Let $\langle V_k^R\rangle = 0$, and let

$$\hat{h}_{kk'}(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t}\langle V_{-k'}^R V_k^R(t)\rangle\,dt, \quad (4.17)$$

where $V_k^R(t)$ is the time evolution of V_k^R according to the dynamics of the uncoupled reservoir, and $\langle ...\rangle$ denotes expectation value in a stationary reference state, and put

$$\begin{aligned} \hat{\psi}_{kk'}^\hbar(\omega) &= \frac{1}{2}(\hat{h}_{kk'}(\omega) + \hat{h}_{-k',-k}(-\omega)) \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} e^{-i\omega t}\langle[V_{-k'}^R,V_k^R(t)]_+\rangle\,dt, \end{aligned} \quad (4.18)$$

$$\begin{aligned} \hat{\varphi}_{kk'}^\hbar(\omega) &= \frac{1}{i\hbar}(\hat{h}_{kk'}(\omega) - \hat{h}_{-k',-k}(-\omega)) \\ &= \frac{1}{i\hbar}\int_{-\infty}^{+\infty} e^{-i\omega t}\langle[V_{-k'}^R,V_k^R(t)]\rangle\,dt. \end{aligned} \quad (4.19)$$

We have to assume that the functions $t \mapsto |\langle V^R_{-k'} V^R_k(t)\rangle|$ are integrable; this requires the reservoir R to be infinitely extended.

Then the reduced dynamics of the system S, in the interaction picture, is approximated, in the limit $\lambda \to 0$, $\tau = \lambda^2 t$ constant, by a dynamical semigroup $\{\phi_\tau = \exp L\tau : \tau \in \mathbb{R}^+\}$ (see Ref. 12 for the proof when the V_k, V^R_k are finitely many and bounded, and the reservoir correlation functions satisfy appropriate conditions).

Proposition 4.2: The generator L of the reduced dynamics in the weak coupling limit is given by

$$L(A) = \frac{i}{\hbar}[\delta H, A] + \frac{1}{\hbar^2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{h}_{kk'}(\omega_k) \times (V_{-k'} A V_k - \frac{1}{2}[V_{-k'} V_k, A]_+), \tag{4.20}$$

where

$$\delta H = \frac{1}{2}\int_0^\infty \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} e^{-i\omega_k t}\left((i\hbar)^{-1}\psi^\hbar_{kk'}(t)[V_k, V_{-k'}] - \frac{1}{2}\varphi^\hbar_{kk'}(t)[V_k, V_{-k'}]_+\right) dt. \tag{4.21}$$

L satisfies the assumption of Proposition 4.1, and its classical limit is exactly (3.12).

Proof: The expression (4.20) is just a way of writing the well-known result of the weak coupling limit theory of Davies[12] (see also Ref. 21); it is of the form (4.5), as one can see by diagonalizing the positive matrix $(\hat{h}_{kk'}(\omega))$: the dependence on $\hbar$ is of the type (4.10), since

$$\hat{h}_{kk'}(\omega) - \hat{h}_{-k',-k}(-\omega) = i\hbar\hat{\varphi}^\hbar_{kk'}(\omega) = O(\hbar). \tag{4.22}$$

Then, by Proposition 4.1, the classical limit of (4.20) exists and is finite. In order to compute it, we rewrite (4.20) as

$$L(A) = \frac{i}{\hbar}[\delta H, A] - \frac{1}{2\hbar^2}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\psi}^\hbar_{kk'}(\omega_k) \times [V_{-k'}, [V_k, A]] - \frac{i}{4\hbar}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\varphi}^\hbar_{kk'}(\omega_k)[V_{-k'}, [V_k, A]]_+$$

and we notice that

$$\hat{\psi}^\hbar_{kk'}(\omega) \underset{\hbar \to 0}{\longrightarrow} \hat{\psi}_{kk'}(\omega), \quad \hat{\varphi}^\hbar_{kk'}(\omega) \underset{\hbar \to 0}{\longrightarrow} \hat{\varphi}_{kk'}(\omega),$$

where $\hat{\psi}_{kk'}(\omega)$, $\hat{\varphi}_{kk'}(\omega)$ are defined by (3.13), (3.14). The outcome is (3.12). ∎

Remark: In taking the classical limit of (4.20), we have tacitly assumed that the eigenvalues $\omega_k(\hbar)$ of the Liouville–von Neumann operator $(i\hbar)^{-1}[\cdot, H^S]$ have finite limits $\omega_k(0)$ as $\hbar \to 0$, and that the corresponding eigenvectors $V_k(\hbar)$ can be chosen to be such as to have limits $V_k(0)$ which are smooth functions of q and p. In the case of a harmonic oscillator, the eigenvalues ω_k are the integral multiples of the frequency ω, independently of $\hbar$, and the eigenvectors can be chosen to be the $\hbar$-independent smooth functions $(\omega^{1/2}Q + i\omega^{-1/2}P)^m \times (\omega^{1/2}Q - i\omega^{-1/2}P)^n$, by exploiting the infinite degeneracy of the spectrum of the Liouville–von Neumann operator. In the generic case, since a quantum system with potential $U(q)$ tending to $+\infty$ as $|q| \to \infty$ has pure point spectrum, whereas the corresponding classical system has, typically, continuous spectrum, some of the above tacit assumptions break down. For instance, for the case $U(q) = |q|$, the eigenvalues $\omega_k(\hbar)$ all collapse to zero as $\hbar \to 0$.

Finally, we spend a few words on the integrability of the correlation functions $\langle V^R_{-k'} V^R_k(t)\rangle$ and on their classical limits. One may take R to be a (finite collection of) infinitely extended harmonic crystal(s), and V^R_k to be suitable linear combinations, or series, of coordinates (and momenta) of the reservoir particles; the reference state may be taken to be a (tensor product of) thermal state(s). Then the multitime correlation functions can be explicitly computed, both in the quantum and in the classical case, and are expressed in terms of the two-time functions $h_{kk'}(t-s) = \langle V^R_{-k'}(s)V^R_k(t)\rangle$. The quantum correlation functions contain $\hbar$ explicitly, and their classical limit $\hbar \to 0$ gives exactly the classical correlation functions. The evolution of the uncoupled reservoir has a Lebesgue spectrum on the orthogonal complement of the cyclic vector, under very general conditions[20,22]; then it is possible to choose suitable combinations V_k of the coordinates (and momenta) which make the functions $\langle V^R_{-k'} V^R_k(t)\rangle$ integrable. For nearest-neighbor coupling and in ν space dimensions, the correlation functions $\langle q_{k'}\, q_k(t)\rangle$ decay as $|t|^{-\nu/2}$ (and are integrable for $\nu \geqslant 3$) both in the classical and in the quantum case.

5. STATIONARY STATES AND DETAILED BALANCE

In this section we deal with the problem of determining the stationary state ρ of a dynamical semigroup $\{\phi_t\}$ derived in the weak coupling limit for a classical system, or equivalently in the classical limit from a quantum dynamical semigroup obtained in the weak coupling limit. We also discuss briefly the detailed balance condition of Graham and Haken and its relationship to microscopic reversibility.

We rewrite the formal generator (3.12):

$$L(A) = \{A, \delta H\} + \frac{1}{2}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} (\hat{\psi}_{kk'}(\omega_k) \times \{V_{-k'}, \{V_k, A\}\} + \hat{\varphi}_{kk'}(\omega_k) V_{-k'}\{V_k, A\}) \tag{5.1}$$

for all A in $\mathscr{A} = C^\infty_0(X) + \mathbb{C}1$, $X = \mathbb{R}^{2f}$. Let ρ be a strictly positive function in $L^1(X) \cap C^2(X)$, with $\iint_X \rho\, dq\, dp = 1$. Then L can be uniquely split as $L_s + L_h$, where L_s is symmetric as an operator in $\mathscr{H} = L^2(X, \rho)$ and L_h is a first-order differential operator, by Lemma 2.1.

Lemma 5.1: We have explicitly

$$L_s(A) = \frac{1}{2}\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\psi}_{kk'}(\omega_k) \times (\{V_{-k'}, \{V_k, A\}\} + \rho^{-1}\{V_{-k'}, \rho\}\{V_k, A\}), \tag{5.2}$$

and

$$L_h(A) = \{A,\delta H\} + \frac{1}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [\hat{\varphi}_{kk'}(\omega_k) V_{-k'} - \hat{\psi}_{kk'}(\omega_k) \times \rho^{-1}\{V_{-k'},\rho\}]\{V_k,A\}. \tag{5.3}$$

for all A in $\mathscr{A}$.

Proof: We construct L_s through the quadratic form

$$E(A,B) = -\iint_X A^* L_s(B)\, dq\, dp, \quad A,B \in \mathscr{A}, \tag{5.4}$$

associated with it; E must be positive and symmetric. The only positive symmetric form which gives rise to a second-order differential operator whose second-order part is the same as for $-L$ is given by

$$E(A,B) = \frac{1}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\psi}_{kk'}(\omega_k) \iint_X \{V_{k'},A\}^* \{V_k,B\}\, \rho\, dq\, dp$$
$$= \sum_{i,j=1}^{2f} \iint_X (\partial_i A)^* a_{ij} (\partial_j A)\, \rho\, dq\, dp, \quad A,B \in \mathscr{A}, \tag{5.5}$$

where the positive matrix (a_{ij}) is given by (3.16); the operator L_s associated with E is indeed (5.2). Then L_h is obtained as $L - L_s$, and the result is (5.3). Note that L_h may be also written in the form (2.5), with

$$v_j = \frac{1}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\varphi}_{kk'}(\omega_k) V_{-k'}\, u_{kj} - \sum_{i=1}^{2f} (\rho^{-1}\partial_i \rho) a_{ij} + w_j, \quad j = 1,\ldots,2f, \tag{5.6}$$

where u_j, w_j are defined by (3.18), (3.19). ■

Taking into account Lemma 2.2, we have

Lemma 5.2: Let ρ be a strictly positive function in $L^1(X) \cap C^2(X)$, with $\iint_X \rho\, dq\, dp = 1$. Then the following conditions are equivalent:

(i) $\iint_X L(A)\rho\, dq\, dp = 0$ for all A in $\mathscr{A}$;

(ii) $\sum_{j=1}^{2f} \partial_j (v_j \rho) = 0$;

(iii) $$\{\delta H, \rho\} + \frac{1}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \{[\hat{\varphi}_{kk'}(\omega_k) \times V_{-k'}\rho - \hat{\psi}_{kk'}(\omega_k)\{V_{-k'},\rho\}], V_k\} = 0. \tag{5.7}$$

Proof: By Lemma 2.2, conditions (i) and (ii) are equivalent, and are also equivalent to $\iint_X L_h(A)\rho\, dq\, dp = 0$ for all A in $\mathscr{A}$. Upon inserting (5.3) in the above, and integrating by parts, we obtain (5.7). ■

As a special case, we consider the situation in which the reservoir is in thermal equilibrium at inverse temperature $\beta > 0$. This is expressed by the classical KMS condition[28,15]

$$\hat{\varphi}_{kk'}(\omega) = -i\beta\omega \hat{\psi}_{kk'}(\omega), \quad \omega \in \mathbb{R}, \; k,k' \in \mathbb{Z}. \tag{5.8}$$

Proposition 5.3: If the reservoir satisfies the classical KMS condition at inverse temperature $\beta > 0$, the canonical state $\rho = Ne^{-\beta H^S}$ ($N = [\iint_X e^{-\beta H^S} dq\, dp]^{-1}$) is stationary under the reduced dynamics.

Proof: If $\rho = Ne^{-\beta H^S}$, then $\rho^{-1}\{V_{-k'},\rho\} = -i\beta\omega_{k'} \times V_{-k'}$. Inserting this and (5.8) into (5.3), we find $L_h = \{\cdot,\delta H\}$. Now δH has vanishing Poisson bracket with all functions of H^S, and (5.7) follows. ■

It may not be easy, in general, to find the stationary state (or states) for arbitrary reservoirs. We shall solve this problem in general for the case of a one-dimensional harmonic oscillator, in the next section. Under certain conditions, it may be easier to find a stationary state such that $L_h = \{\cdot,\delta H\}$. The following construction is due to Graham and Haken.[6]

Proposition 5.4: Suppose that the matrix $(a_{ij}(q,p))$ is strictly positive for all (q,p) in $\mathbb{R}^{2f}$. Put

$$F_i = \sum_{j=1}^{2f} (a^{-1})_{ij} \frac{1}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\varphi}_{kk'}(\omega_k) V_{-k'} u_{kj},$$
$$i = 1,\ldots,2f. \tag{5.9}$$

Suppose that the F_i are C^1 functions satisfying

$$\partial_j F_i = \partial_i F_j \quad \text{for all } i,j = 1,\ldots,2f \tag{5.10}$$

and

$$\sum_{j=1}^{2f} (\partial_j w_j + w_j F_j) = 0. \tag{5.11}$$

Let ϕ be the C^2 function such that

$$F_i = -\partial_i \phi, \quad i = 1,\ldots 2f, \tag{5.12}$$

and suppose that $e^{-\phi}$ is in $L^1(X)$. Then

$$\rho = Ne^{-\phi}, \tag{5.13}$$

where N is a normalization factor, is a stationary state such that

$$L_h(A) = \{A,\delta H\} \quad \text{for all } A \text{ in } \mathscr{A}. \tag{5.14}$$

Proof: If (a_{ij}) is strictly positive, we can define F_i by (5.9). If (5.10) holds, there exists ϕ in $C^2(X)$ satisfying (5.12), unique up to an additive constant; then $e^{-\phi}$ is in $C^2(X)$, unique up to a multiplicative constant. If it is in $L^1(X)$ we may choose the constant N so as to have $\iint_X Ne^{-\phi}\, dq\, dp = 1$. Putting $\rho = Ne^{-\phi}$ and inserting it in (5.6) gives

$$v_j = w_j, \quad f = 1,\ldots,2f. \tag{5.15}$$

Then the stationarity condition (ii) of Lemma 5.2 becomes

$$0 = \sum_{j=1}^{2f} \partial_j (w_j e^{-\phi}) = \sum_{j=1}^{2f} (\partial_j w_j + w_j F_j) e^{-\phi}$$

which is (5.11). Finally, (5.14) follows from (5.15). ■

It may be the case that no state satisfying the conditions of Proposition 5.4 exists (see the end of the section). However, when such a state exists, it is the unique stationary state, and all normal states approach it under the action of $\phi_t = \exp Lt$, under some technical conditions. We have

Theorem 5.5: Let $(a_{ij}(q,p))$ be strictly positive for all (q,p), and ρ be a strictly positive function in $L^1(X) \cap C^2(X)$, with $\iint_X \rho\, dq\, dp = 1$, satisfying the equivalent conditions of Lemma 5.2; assume that the closures $\hat{L}$ and $\hat{L}_s$ of L and L_s as operators in $\mathscr{H} = L^2(X,\rho)$ are maximal dissipative and self-adjoint, respectively. Then

$$\lim_{t\to\infty} e^{\hat{L}t}A = \left(\iint_X A\rho\, dq\, dp\right)1$$

for all A in $\mathscr{M} = L^\infty(X)$. (5.16)

Proof: If L is maximal dissipative in $\mathscr{H}$, $\{e^{\hat{L}t}: t\in\mathbb{R}^+\}$ is a contraction semigroup on $\mathscr{H}$, and $\|A\|_2^2 - \|e^{\hat{L}t}A\|_2^2 \geqslant 0$ for all A in $\mathscr{M}$, t in $\mathbb{R}^+$. Let

$$\mathscr{N} = \{A\in\mathscr{M}: \|A\|_2^2 - \|e^{\hat{L}t}A\|_2^2 = 0 \quad \text{for all } t \text{ in } \mathbb{R}^+\}.$$

It can be shown as in Ref. 29, Theorem 3.1, that the conditions $\mathscr{N} = \mathbb{C}1$ implies (5.16). We show here that an element A of $\mathscr{N}$ must be a constant, under the assumptions made. Let $\hat{L}^+$ be the adjoint of $\hat{L}$; we have also $\hat{L}^+ = L^+$. For all A, B in $\mathscr{A}$ we have

$$(LA,B) = (L_s A + L_h A, B) = (A, L_s B - L_h B)$$

since L_s is symmetric and L_h is skew-symmetric with respect to the inner product given by ρ (see Lemma 2.2). Then $\mathscr{A}\subseteq \operatorname{Dom}\hat{L}^+$ and $\hat{L}^+B = L_s B - L_h B$ for all B in $\mathscr{A}$. For all A in $\mathscr{A}$ we have therefore

$$\hat{L}A + \hat{L}^+A = 2\hat{L}_sA$$

and $\mathscr{A}$ is a core for $\hat{L}_s$; then, by the Lie–Trotter product formula,

$$e^{2\hat{L}_st}A = \lim_{n\to\infty}(e^{\hat{L}^+t/n}e^{\hat{L}t/n})^nA$$

for all A in $\mathscr{H}$, t in $\mathbb{R}^+$. (5.17)

Let A be in $\mathscr{N}$. Then, by the Schwarz inequality,

$$(B,A) - (e^{\hat{L}t}B, e^{\hat{L}t}A) = 0 \quad \text{for all } B \text{ in } \mathscr{H};$$

hence $A = e^{\hat{L}^+t}e^{\hat{L}t}A$ for all $t\geqslant 0$. From (5.17), it follows that $e^{\hat{L}_st}A = A$ for all t; hence $A\in\operatorname{Dom}\hat{L}_s$ and $\hat{L}_s(A) = 0$. Then A is in the domain of the form closure $\hat{E}$ of E and

$$\hat{E}(A,A) = \iint_X (\partial_i A)^* a_{ij}(\partial_j A)\rho\, dq\, dp = 0,$$

where $(\partial_j A)_{j=1,\dots,2f}$ is the distributional gradient of A. Since $(a_{ij}(q,p))$ and $\rho(q,p)$ are strictly positive for all (q,p) in X, it follows that A is constant. ■

When the global dynamics $\mathscr{U}_t^\lambda$ and the reference state ρ_{ref} of the reservoir are time-reversal-invariant, $\{\cdot,\delta H\}$ is precisely the part of L which changes sign under time reversal, by Proposition 3.3. Then the condition that $L_h = \{\cdot,\delta H\}$ for the stationary state ρ amounts to the detailed balance condition of Graham and Haken.[6] If ρ is the only stationary state, it must be time-reversal-invariant: then the two terms in the left-hand side of (5.7), $\{\delta H,\rho\}$ and $\frac{1}{2}\sum_{k,k',\ \omega_k=\omega_{k'}}\{[\hat{\varphi}_{kk'}(\omega_k)V_{k'}\rho - \hat{\psi}_{kk'}(\omega_k)\{V_{-k'},\rho\}], V_k\}$ are odd and even, respectively, under time reversal; hence they must vanish separately if (5.7) holds. With the notation (5.6) stationarity of ρ becomes

$$\sum_{j=1}^{2f}\partial_j(w_j\rho) = 0$$

and

$$\sum_{j=1}^{2f}\partial_j\left(\left[\frac{1}{2}\sum_{\substack{k,k'\\ \omega_k=\omega_{k'}}}\hat{\varphi}_{kk'}(\omega_k)V_{-k'}u_{kj} - \sum_{i=1}^{2f}(\rho^{-1}\partial_i\rho)a_{ij}\right]\rho\right) = 0, \tag{5.18}$$

whereas the detailed balance condition of Graham and Haken is

$$\sum_{j=1}^{2f}\partial_j(w_j\rho) = 0$$

and

$$\frac{1}{2}\sum_{\substack{k,k'\\ \omega_k=\omega_{k'}}}\hat{\varphi}_{kk'}(\omega_k)V_{-k'}u_{kj} - \sum_{i=1}^{2f}(\rho^{-1}\partial_i\rho)a_{ij}\rho = 0 \quad \forall j. \tag{5.19}$$

Hence the detailed balance condition of Graham and Haken seems stronger than the condition of time reversal invariance of the global dynamics and of the reference state. A comparison of Eqs. (5.18) and (5.19) is not conclusive, since it has not been shown that (5.18) is equivalent to time-reversal invariance plus existence of a stationary state. However, the following counterexample proves that the latter conditions do *not* imply detailed balance in the sense of Graham and Haken.

We consider a system of two harmonic oscillators, with Hamiltonian

$$\begin{aligned} H^S = H &= \tfrac{1}{2}[(p_1^2+p_2^2)+\omega^2(q_1^2+q_2^2)] \\ &= \omega(a_1^*a_1 + a_2^*a_2), \end{aligned} \tag{5.20}$$

where

$$a_j = (1/\sqrt{2})(\omega^{1/2}q_j + i\omega^{-1/2}p_j), \quad j=1,2. \tag{5.21}$$

We choose a time-reversal-invariant coupling of the form

$$H^{SR} = a_1^*V_1^R + a_2^*V_2^R + a_1a_2^*V_3^R + \text{conj}, \tag{5.22}$$

where V_j^R are functions of the reservoir coordinates and momenta such that $\mathscr{T}V_j^R = V_j^{R*}$. We assume

$$\left.\begin{aligned} \langle V_i^{R*}V_j^R(t)\rangle &= \delta_{ij}\psi_j(t) \\ \langle\{V_i^{R*}, V_j^R(t)\}\rangle &= \delta_{ij}\varphi_j(t) \end{aligned}\right\}, \quad i,j=1,2,3 \tag{5.23}$$

(the remaining two-point correlations vanish),

$$\hat{\varphi}_j(\omega) = i\beta_j\omega\hat{\psi}_j(\omega), \quad j=1,2,3,\ \omega\in\mathbb{R}. \tag{5.24}$$

The above conditions express the fact that the system is coupled to three independent reservoirs at inverse temperatures β_j. From Proposition 3.1 we get

$$\begin{aligned} L(A) = \{A,\delta H\} &+ \tfrac{1}{2}\hat{\psi}_1(\omega)[\{a_1^*,\{a_1,A\}\} - i\beta_1\omega a_1^*\{a_1,A\} \\ &+ \{a_1,\{a_1^*,A\}\} + i\beta_1\omega a_1\{a_1^*,A\}] \\ &+ \tfrac{1}{2}\hat{\psi}_2(\omega)[\{a_2^*,\{a_2,A\}\} - i\beta_2\omega a_2^*\{a_2,A\} \\ &+ \{a_2,\{a_2^*,A\}\} + i\beta_2\omega a_2\{a_2^*,A\}] \\ &+ \tfrac{1}{2}\hat{\psi}_3(0)[\{a_1^*a_2,\{a_1a_2^*,A\}\} \\ &+ \{a_1a_2^*,\{a_1^*a_2,A\}\}], \end{aligned} \tag{5.25}$$

where

$$\delta H = \frac{1}{2}\int_0^\infty (- i[e^{-i\omega t}\psi_1(t) - e^{i\omega t}\psi_1(-t)]$$
$$+ [e^{-i\omega t}\varphi_1(t) - e^{i\omega t}\varphi_1(-t)]a_1^*a_1$$
$$- i[e^{-i\omega t}\psi_2(t) - e^{i\omega t}\psi_2(-t)]$$
$$+ [e^{-i\omega t}\varphi_2(t) - e^{i\omega t}\varphi_2(-t)]a_2^*a_2$$
$$- i[\psi_3(t) - \psi_3(-t)]\ [a_2^*a_2 - a_1^*a_1]$$
$$+ [\varphi_3(t) - \varphi_3(-t)]a_1^*a_1a_2^*a_2)\, dt. \tag{5.26}$$

We suppose $\widehat{\psi}_1(\omega) = \alpha_1$, $\widehat{\psi}_2(\omega) = \alpha_2$, $\widehat{\psi}_3(0) = \alpha_3$ to be strictly positive. Then the matrix (a_{ij}) (3.16) is strictly positive for all (q,p).

Proposition 5.6: ϕ_t has a unique stationary state ρ. In spite of the time-reversal invariance of $\mathscr{U}_t^\lambda$ and $\rho_{\rm ref}$, the detailed balance condition of Graham and Haken does not hold, unless $\beta_1 = \beta_2$.

Proof: The stationary state has been explicitly constructed in Ref. 30, as a convergent expansion in powers of $\beta_1 - \beta_2$. It is strictly positive, hence unique by Theorem 5.5, and is invariant under the gauge transformation

$$a_1 \mapsto e^{-i\vartheta_1}a_1, \quad a_2 \mapsto e^{-i\vartheta_2}a_2, \quad \vartheta_1,\vartheta_2 \in [0,2\pi).$$

Then, putting $\rho = Ne^{-\phi}$, we may write

$$\phi = \phi(J_1,J_2), \quad J_j = a_j^*a_j, \quad j = 1,2. \tag{5.27}$$

The explicit form of ρ is not needed here; it suffices to show that there is no function $\phi(J_1,J_2)$ such that $\{\delta H, \phi\} = 0$ and $L_h(A) = \{A,\delta H\}$. Since δH is a function of J_1,J_2, we have $\{\delta H, \phi\} = 0$ automatically. Taking into account (5.3), the condition $L_h(A) = \{A,\delta H\}$ becomes

$$\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [\widehat{\varphi}_{kk'}(\omega_k)V_{-k'} + \widehat{\psi}_{kk'}(\omega_k)\{V_{-k'}, \phi\}]$$
$$\times \frac{\partial}{\partial a_j^\#} V_k = 0, \quad j = 1,2, \tag{5.28}$$

where $a_j^\# = a_j$ or a_j^*. Using the explicit form of V_k, we obtain the two equations

$$\widehat{\varphi}_1(\omega)a_1^* + \widehat{\psi}_1(\omega)\{a_1^*, \phi\} + \widehat{\psi}_3(0)\{a_1^*a_2, \phi\}a_2^* = 0,$$
$$\widehat{\varphi}_2(\omega)a_2^* + \widehat{\psi}_2(\omega)\{a_2^*, \phi\} + \widehat{\psi}_3(0)\{a_1a_2^*, \phi\}a_1^* = 0,$$

and the complex conjugate equations. More explicitly, we find

$$\alpha_1\Big(\frac{\partial}{\partial a_1}\phi - \beta_1\omega a_1^*\Big) - \alpha_3\Big(a_1^*a_2^*\frac{\partial}{\partial a_2^*}\phi - a_2^*a_2\frac{\partial}{\partial a_1}\phi\Big) = 0,$$
$$\alpha_2\Big(\frac{\partial}{\partial a_2}\phi - \beta_2\omega a_2^*\Big) - \alpha_3\Big(a_1^*a_2^*\frac{\partial}{\partial a_1^*}\phi - a_1^*a_1\frac{\partial}{\partial a_2}\phi\Big) = 0.$$

Since $\phi = \phi(J_1,J_2)$, we have

$$\frac{\partial}{\partial a_j}\phi = a_j^*\frac{\partial}{\partial J_j}\phi,$$
$$\frac{\partial}{\partial a_j^*}\phi = a_j\frac{\partial}{\partial J_j}\phi, \quad j = 1,2;$$

hence

$$\Big(\frac{\partial}{\partial J_1}\phi - \beta_1\omega\Big) + \frac{\alpha_3}{\alpha_2}J_2\Big(\frac{\partial}{\partial J_1}\phi - \frac{\partial}{\partial J_2}\phi\Big) = 0,$$
$$\Big(\frac{\partial}{\partial J_2}\phi - \beta_2\omega\Big) + \frac{\alpha_3}{\alpha_2}J_1\Big(\frac{\partial}{\partial J_2}\phi - \frac{\partial}{\partial J_1}\phi\Big) = 0. \tag{5.29}$$

We add and subtract the two equations, obtaining

$$\Big(\frac{\partial}{\partial J_1} + \frac{\partial}{\partial J_2}\Big)\phi + \frac{\alpha_3}{\alpha_1\alpha_2}(\alpha_2J_1 - \alpha_1J_2) \times \Big(\frac{\partial}{\partial J_1} - \frac{\partial}{\partial J_2}\Big)\phi = (\beta_1 + \beta_2)\omega,$$
$$\Big[1 + \frac{\alpha_3}{\alpha_1\alpha_2}(\alpha_2J_1 + \alpha_1J_2)\Big]\Big(\frac{\partial}{\partial J_1} - \frac{\partial}{\partial J_2}\Big)\phi = (\beta_1 - \beta_2)\omega.$$

We introduce new variables $u = J_1 + J_2$, $v = J_1 - J_2$ and parameters $\lambda = (\alpha_3/2\alpha_1\alpha_2)(\alpha_1 + \alpha_2)$, $\mu = (\alpha_3/2\alpha_1\alpha_2)(\alpha_2 - \alpha_1)$ and get

$$\frac{\partial}{\partial u}\phi + (\lambda u + \mu v)\frac{\partial}{\partial v}\phi = \frac{1}{2}(\beta_1 + \beta_2)\omega,$$
$$(1 + \lambda v + \mu u)\frac{\partial}{\partial v}\phi = \frac{1}{2}(\beta_1 - \beta_2)\omega. \tag{5.30}$$

The solution of the linear system (5.30) is

$$\frac{\partial}{\partial u}\phi = \frac{1}{2}(\beta_1 + \beta_2)\omega - \frac{1}{2}(\beta_1 - \beta_2)\omega(\lambda u + \mu v)(1 + \lambda v + \mu u)^{-1},$$
$$\frac{\partial}{\partial v}\phi = \frac{1}{2}(\beta_1 - \beta_2)\omega(1 + \lambda v + \mu u)^{-1}. \tag{5.31}$$

If $\beta_1 = \beta_2 = \beta$, the solution is $\phi = \beta\omega u$, i.e., $\rho = Ne^{-\beta H}$. If $\beta_1 \neq \beta_2$, the compatibility condition

$$\frac{2}{\omega(\beta_1 - \beta_2)}\frac{\partial^2\phi}{\partial u\partial v} = \frac{\partial}{\partial u}\frac{1}{1 + \lambda v + \mu u} = -\frac{\partial}{\partial v}\frac{\lambda u + \mu v}{1 + \lambda v + \mu u} \tag{5.32}$$

is violated [the above equality is equivalent to $(\lambda^2 - \mu^2)u = \mu - 1$]. ■

We remark that the usual "proof" of detailed balance from time reversal[6] is reliable only in the case of a system coupled to a thermal reservoir, in that it tacitly assumes that $\rho\rho_{\rm ref}$ is at least approximately invariant under the coupled dynamics $\mathscr{U}_t^\lambda$ (cf. Ref. 21). In the case of a system coupled to a reservoir at a given inverse temperature β, this approximate invariance follows from the stability of KMS states,[31,10] but such an assumption need not be true for a system coupled to several reservoirs at different temperatures and is definitely false in the quantum case.[1,2] However, as we shall show in the following section, the detailed balance condition of Graham and Haken holds as a consequence of time-reversal invariance alone for a one-dimensional harmonic oscillator weakly coupled to several reservoirs, possibly at different temperatures. Hence the detailed balance condition of Graham and Haken, in contrast to the quantum detailed balance condition, does *not* have the role of singling out thermal reservoirs at the same temperature. This was to be expected in view of the variety of examples (see Refs. 6 and 7) of Fokker–Planck equations (however, not derived in the weak coupling limit *á la* Davies) which often satisfy the Graham–Haken detailed balance condition also far from thermal equilibrium.

6. STATIONARY STATE OF THE HARMONIC OSCILLATOR AND CLASSICAL KMS CONDITION

Here we specialize the results of the previous section to the case of a single one-dimensional harmonic oscillator, coupled to several reservoirs. Taking advantage of the fact that there exists only one constant of the uncoupled motion, the Hamiltonian H, we are able to compute the stationary state ρ explicitly as a function of H. If the global dynamics and the reference states of the reservoirs are time-reversal-invariant, the detailed balance condition of Graham and Haken holds. In conclusion, we prove that the stationary state is independent of the coupling if and only if all the reservoirs satisfy the classical KMS condition at the same inverse temperature (similarly to what occurs in the quantum case[1,2]). We let

$$H^S = H = \tfrac{1}{2}(p^2 + \omega^2 q^2) = \omega a^* a, \tag{6.1}$$

where

$$a = (1/\sqrt{2})(\omega^{1/2} q + \omega^{-1/2} p). \tag{6.2}$$

The eigenvectors V_k of $\{\cdot, H\}$ are of the form

$$V_k = \sum_{m > |n(k)|/2} C_{mk} a^{m + n(k)/2} a^{*m - n(k)/2}, \tag{6.3}$$

where $n(k) \in \mathbb{Z}$, $n(-k) = -n(k)$ and where $C_{m,-k} = \overline{C_{mk}}$, in order to have $V_{-k} = V_k^*$. The corresponding eigenvalues are

$$-i\omega_k = -in(k)\omega. \tag{6.4}$$

If $n(k) = n(k') = n$, we have

$$V_k V_{-k'} = \sum_{m,m' > |n|/2} C_{mk} \overline{C_{m'k'}} (\omega^{-1} H)^{m+m'}, \tag{6.5}$$

and

$$\begin{aligned}\{V_k, V_{-k'}\} &= -in \sum_{m,m' > |n|/2} C_{mk} \overline{C_{m'k'}} (m + m') \\ &\qquad \times (\omega^{-1} H)^{m+m'-1} \\ &= -i\omega_k \frac{d}{dH}(V_k V_{-k'}).\end{aligned} \tag{6.6}$$

We shall frequently need in the following the expression

$$\begin{aligned}&\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \omega_k^2 \hat{\psi}_{kk'}(\omega_k) V_k V_{-k'} \\ &= \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \hat{\psi}_{kk'}(\omega_k) \{V_k, H\}\{V_{-k'}, H\} \\ &= \sum_{i,j=1}^{2} a_{ij} (\partial_i H)(\partial_j H) \equiv G(H) \geqslant 0,\end{aligned} \tag{6.7}$$

where the matrix (a_{ij}) is given by (3.16).

Proposition 6.1: If $G(H)$ is strictly positive for all values of $H \neq 0$, then there exists at most one stationary state $\rho = \rho(H)$, which is given by

$$\begin{aligned}\rho(H) = N \exp\Bigg[\int_0^H G(H')^{-1} &\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} (-i\omega_k \hat{\varphi}_{kk'}(\omega_k) \\ &\times V_k V_{-k'})(H')\, dH'\Bigg]\end{aligned} \tag{6.8}$$

whenever the above expression defines a state.

Proof: If the stationary state is unique, it must be a constant of the uncoupled motion, since L commutes with $\{\cdot, H\}$. Hence we look for a stationary state of the form $\rho = \rho(H)$. Then $\{\delta H, \rho\} = 0$. Upon defining

$$b(H) = \rho(H)^{-1} \frac{d}{dH} \rho(H), \tag{6.9}$$

the stationarity condition (5.7) becomes

$$\begin{aligned}0 &= \sum_{\substack{k,k' \\ \omega_k \omega_{k'}}} \{[\hat{\varphi}_{kk'}(\omega_k) - i\omega_k \hat{\psi}_{kk'}(\omega_k) b(H)] \\ &\qquad \times V_{-k'} \rho(H), V_k\} \\ &= \rho(H) \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} (\omega_k^2 \hat{\psi}_{kk'}(\omega_k) b'(H) V_k V_{-k'} \\ &\qquad - [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k \hat{\psi}_{kk'}(\omega_k) b(H)] \\ &\qquad \times [\{V_k, V_{-k'}\} - i\omega_k b(H) V_k V_{-k'}]) \\ &= \rho(H) \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} \Big(\omega_k^2 \hat{\psi}_{kk'}(\omega_k) b'(H) V_k V_{-k'} \\ &\qquad + i\omega_k [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k \hat{\psi}_{kk'}(\omega_k) b(H)] \\ &\qquad \times \Big[b(H) V_k V_{-k'} + \frac{d}{dH}(V_k V_{-k'})\Big]\Big),\end{aligned} \tag{6.10}$$

where we have taken into account (6.6). Then, defining

$$\begin{aligned}F(H) = &\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} i\omega_k \\ &\times [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k \hat{\psi}_{kk'}(\omega_k) b(H)] V_k V_{-k'},\end{aligned} \tag{6.11}$$

we see that (6.10) is equivalent to

$$\frac{d}{dH} F(H) + b(H) F(H) = 0, \tag{6.12}$$

whose solution is

$$F(H) = F(0) \exp\Big[-\int_0^H b(H')\, dH'\Big] = 0, \tag{6.13}$$

since $(V_k V_{-k'})(0) = 0(\omega_k = \omega_{k'})$, by (6.5). Recalling the definition (6.7) of $G(H)$, and assuming $G(H)$ to be strictly positive for $H \neq 0$ (for $H = 0$ it certainly vanishes), we find from (6.11) and (6.13)

$$\begin{aligned}b(H) = G(H)^{-1} &\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} (-i\omega_k \hat{\varphi}_{kk'}(\omega_k) V_k V_{-k'}), \\ 0 < H < \infty,\end{aligned} \tag{6.14}$$

from which (6.8) follows. ■

As a special case, we consider the situation in which the reservoirs are thermal, at possibly different inverse temperatures β_r. This means that there are index sets I_r, $r = 1,2,\ldots$, such that

$$\hat{\varphi}_{kk'} = \hat{\psi}_{kk'} = 0 \quad \text{if } k \in I_r,\ k' \in I_s,\ r \neq s, \tag{6.15}$$

$$\hat{\varphi}_{kk'}(\omega) = -i\beta_r\omega\hat{\psi}_{kk'}(\omega) \quad \text{for all } \omega \text{ in } \mathbb{R} \text{ if } k,k' \in I_r. \tag{6.16}$$

Equation (6.16) is the classical KMS condition at inverse temperature β_r for the rth reservoir, Eq. (6.15) states that the reservoirs are mutually independent. Then we may define

$$G_r(H) = \sum_{\substack{k,k' \in I_r \\ \omega_k = \omega_{k'}}} \omega_k^2\, \hat{\psi}_{kk'}(\omega_k) V_k V_{-k'} \quad (\geqslant 0), \tag{6.17}$$

and we have:

Corollary 6.2: For a one-dimensional harmonic oscillator coupled to several reservoirs at inverse temperatures β_r, the stationary state exists and is given by

$$\rho(H) = N \exp\Big[-\sum_r p_r(H)\,\beta_r H\Big], \tag{6.18}$$

where

$$p_r(H) = \frac{1}{H}\int_0^H G(H')^{-1} G_r(H')\, dH', \quad r = 1,2,\ldots. \tag{6.19}$$

Remark: If $G(H)$ vanishes at some $H = H_0$, then also $G_r(H_0)$ vanishes for all r; then p_r is not uniquely defined and there exist more than one stationary state. For nonthermal reservoirs, it might happen that the denominator in (6.14) vanishes without the numerator vanishing; this is the reason why we have assumed $G(H) > 0$ for $H > 0$.

Now we consider the behavior under the time reversal operation $\mathscr{T}$. We choose the coefficients C_{mk} in (6.3) to be real, so that $\mathscr{T} V_k = V_{-k}$ for all k; then the time reversal invariance of H^{SR} is expressed by $\mathscr{T} V_k^R = V_{-k}^R$ for all k. If also the reference states of the reservoirs are time reversal invariant, this implies

$$\langle V_{-k'}^R\, V_k^R(t)\rangle = \langle V_{k'}^R\, V_{-k}^R(-t)\rangle = \langle V_{-k}^R V_{k'}^R(t)\rangle, \tag{6.20}$$

$$\begin{aligned}\langle\{V_{-k'}^R, V_k^R(t)\}\rangle &= \\ &-\langle\{V_{k'}^R, V_{-k}^R(-t)\}\rangle = \langle\{V_{-k'}^R, V_{k'}^R(t)\}\rangle\end{aligned}$$

(we have used $\{\mathscr{T} f, \mathscr{T} g\} = -\mathscr{T}\{f,g\}$), from which it follows

$$\hat{\psi}_{kk'}(\omega) = \hat{\psi}_{k'k}(\omega), \quad \hat{\varphi}_{kk'}(\omega) = \hat{\varphi}_{k'k}(\omega) \tag{6.21}$$

for all ω in $\mathbb{R}$ and for all indices k,k'.

Proposition 6.3: For a one-dimensional harmonic oscillator weakly coupled to several reservoirs, time-reversal invariance implies the detailed balance condition of Graham and Haken.

Proof: For all A in $\mathscr{A}$, we have from (5.3) and (6.9)

$$\begin{aligned} L_h(A) &= \{A,\delta H\} - \frac{i}{2} \sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k\, \psi_{kk'}(\omega_k) b(H)] \\ &\times V_{-k'}\Big[\Big(\frac{\partial}{\partial a} V_k\Big)\Big(\frac{\partial}{\partial a^*} A\Big) - \Big(\frac{\partial}{\partial a^*} V_k\Big)\Big(\frac{\partial}{\partial a} A\Big)\Big], \end{aligned}$$

hence the detailed balance condition of Graham and Haken amounts to

$$\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k\, \hat{\psi}_{kk'}(\omega_k) b(H)]\, V_{-k'} \frac{\partial}{\partial a^\#} V_k = 0, \tag{6.22}$$

where $a^\#$ means a or a^*. From (6.3) we get, for $n(k) = n(k') = n$,

$$\begin{aligned} V_{-k'} \frac{\partial}{\partial a} V_k &= a^* \sum_{m,m' \geqslant \frac{1}{2}|n|} C_{mk} C_{m'k'} (m + \tfrac{1}{2} n)(a^* a)^{m+m'-1}, \\ V_{-k'} \frac{\partial}{\partial a^*} V_k &= a \sum_{m,m' \geqslant \frac{1}{2}|n|} C_{mk} C_{m'k'} (m - \tfrac{1}{2} n)(a^* a)^{m+m'-1}, \end{aligned} \tag{6.23}$$

where we have used the assumed reality of the coefficients $C_{mk} = C_{m,-k}$. Inserting (6.23) into (6.22) and remembering $F(H) = 0$, with F given by (6.11), we see that the two equations (6.22) reduce to the single equation

$$\sum_{\substack{k,k' \\ \omega_k = \omega_{k'}}} [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k\, \hat{\psi}_{kk'}(\omega_k) b(H)] \times \sum_{m,m' \geqslant |n(k)|/2} m C_{mk} C_{m'k'} (\omega^{-1} H)^{m+m'} = 0. \tag{6.24}$$

We show that (6.24) indeed holds as a consequence of time reversal invariance. In the left-hand side of (6.24) we perform two relabellings of indices:

(i) $k \to -k'$, $k' \to -k$, $m \to m'$

(then $\omega_k \to -\omega_{k'} = -\omega_k$);

(ii) $k \to k'$, $k' \to k$, $m \to m'$.

Under (i) the left-hand side of (6.24) becomes

$$-\sum_{\substack{kk' \\ \omega_k = \omega_{k'}}} [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k\, \hat{\psi}_{kk'}(\omega_k) b(H)] \times \sum_{m,m' \geqslant |n(k)|/2} m' C_{mk} C_{m'k'} (\omega^{-1} H)^{m+m'},$$

where we have used (3.13), (3.14), and $C_{mk} = C_{m,-k}$.

Under (ii), it becomes

$$\sum_{\substack{kk' \\ \omega_k = \omega_{k'}}} [\hat{\varphi}_{kk'}(\omega_k) - i\omega_k\, \hat{\psi}_{kk'}(\omega_k) b(H)] \times \sum_{m,m' \geqslant |n(k)|/2} m' C_{mk} C_{m'k'} (\omega^{-1} H)^{m+m'},$$

where we have used (6.21). The above expressions are then equal and opposite, thus proving (6.24). ■

Remark: The result of Proposition 6.3 was to be expected, in consideration of a heuristic argument of Graham.[7]

Finally, we prove that, if the stationary state is independent of the coupling, then the reservoir must be KMS (the converse result was obtained in Proposition 5.3). We let $\mathscr{R}$

denote the linear space of those reservoir observables A which satisfy

$$\int_{-\infty}^{+\infty} |\langle A^* A(t)\rangle|\, dt < \infty, \int_{-\infty}^{+\infty} |\langle\{A^*, A(t)\}\rangle|\, dt < \infty. \tag{6.25}$$

We assume that for every ω in $\mathbb{R}$ there is an A_ω in $\mathscr{R}$ such that

$$\int_{-\infty}^{+\infty} e^{-i\omega t} \langle A_\omega^* A_\omega(t)\rangle\, dt \neq 0.$$

We fix the reservoir R and couple it subsequently to various one-dimensional harmonic oscillators with different frequencies, by an interaction of the form

$$\lambda H^{SR} = \lambda \sum_{k\in\mathbb{Z}} V_k V_k^R, \tag{6.26}$$

where the V_k are given by (6.3) and the V_k^R are in $\mathscr{R}$.

Proposition 6.4. If the stationary state of any one-dimensional harmonic oscillator weakly coupled to the reservoir R is independent of the coupling, then the observables in $\mathscr{R}$ satisfy the classical KMS condition[28,15] at some inverse temperature β:

$$\int_{-\infty}^{+\infty} e^{-i\omega t} \langle\{A^*, B(t)\}\rangle\, dt = -i\beta\omega \int_{-\infty}^{+\infty} e^{-i\omega t} \langle A^* B(t)\rangle\, dt \tag{6.27}$$

for all A, B in $\mathscr{R}$, ω in $\mathbb{R}$.

Proof: We fix A in $\mathscr{R}$, and we couple the reservoir to a harmonic oscillator with frequency ω_1, with an interaction

$$\lambda H^{SR} = \lambda (a^k A^* + a^{*k} A), \quad k \in \mathbb{N}. \tag{6.28}$$

We let

$$\hat{\psi}_A(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} \langle A^* A(t)\rangle\, dt,$$

$$\hat{\varphi}_A(\omega) = \int_{-\infty}^{+\infty} e^{-i\omega t} \langle\{A^*, A(t)\}\rangle\, dt,$$

and we choose ω_1 such that $\hat{\psi}_A(k\omega_1) \neq 0$. Then the stationary state $\rho(H)$ is unique and given by

$$b(H) = -i\hat{\varphi}_A(k\omega_1)/k\omega_1 \hat{\psi}_A(k\omega_1). \tag{6.29}$$

By assumption, the stationary state is independent of the coupling; hence the left-hand side of (6.29) is independent of A and of k. The right-hand side is independent of $H \in (0, \infty)$. If, for some A' in $\mathscr{R}$, k' in $\mathbb{N}$, $\hat{\psi}_{A'}(k'\omega_1) = 0$ then also $\hat{\varphi}_A(k'\omega_1) = 0$; otherwise $\rho(H)$ could not be a stationary state for that coupling. Then there exists a (necessarily positive) constant $\beta(\omega_1)$ such that

$$b(H) = -\beta(\omega_1),$$
$$\hat{\varphi}_A(k\omega_1) = -i\beta(\omega_1)k\omega_1\hat{\psi}_A(k\omega_1)$$
$$\text{for all } A \text{ in } \mathscr{R}, k \text{ in } \mathbb{N}. \tag{6.30}$$

Now we prove that $\beta(\omega_1)$ does not depend on ω_1 either. If we couple the reservoir to another harmonic oscillator with frequency ω_2, we find, in analogy to (6.30),

$$\hat{\varphi}_A(k\omega_2) = -i\beta(\omega_2)k\omega_2\hat{\psi}_A(k\omega_2)$$
$$\text{for all } A \text{ in } \mathscr{R}, k \text{ in } \mathbb{N}.$$

If ω_1 and ω_2 are mutually rational, there exist $k_1, k_2 \in \mathbb{N}$ such that $k_1\omega_1 = k_2\omega_2 = \omega \neq 0$. If we suppose that for all ω in $\mathbb{R}$ there exists some A in $\mathscr{R}$ such that $\hat{\psi}_A(\omega) \neq 0$, we conclude that $\beta(\omega_1) = \beta(\omega_2)$ for any pair of mutually rational frequencies ω_1, ω_2. By the continuity of $\hat{\psi}_A(\omega)$, $\hat{\varphi}_A(\omega)$, and by the density of the rationals in the real line, we find that $\beta(\omega)$ is actually independent of ω. So we have

$$\hat{\varphi}_A(\omega) = -i\beta\omega\,\hat{\psi}_A(\omega) \quad \text{for all } \omega \text{ in } \mathbb{R}, A \text{ in } \mathscr{R}, \tag{6.31}$$

where β is a positive constant. By polarization, we get (6.27). ■

ACKNOWLEDGMENT

We are grateful to G. Gallavotti for having raised our interest in the problem.

[1] A. Kossakowski, A. Frigerio, V. Gorini, and M. Verri, "Quantum Detailed Balance and KMS condition," Commun. Math. Phys. **57**, 97–110 (1977).

[2] A. Frigerio, V. Gorini, and M. Verri, "Stability, Detailed Balance and KMS Condition for Quantum Systems," in *Lecture Notes in Physics* (Springer-Verlag, Berlin, Heidelberg, New York, 1978), Vol. 80, pp. 420–425.

[3] H. Spohn, "Kinetic Equations from Hamiltonian Dynamics: Markovian Limits," Rev. Mod. Phys. **53**, 569–615 (1980).

[4] L. E. Reichl, *A Modern Course in Statistical Physics* (University of Texas Press, Austin, 1980).

[5] R. Graham, "Solution of Fokker Planck Equations with and without Manifest Detailed Balance," Z. Phys. B **40**, 149–55 (1980).

[6] R. Graham and H. Haken, "Generalized Thermodynamic Potential for Markoff Systems in Detailed Balance and far from Thermal Equilibrium," Z. Phys. **243**, 289–302 (1971).

[7] R. Graham, "Statistical Theory of Instabilities in Stationary Nonequilibrium Systems with Applications to Laser and Nonlinear Optics," in *Springer Tracts in Modern Physics* (Springer-Verlag, Berlin, Heidelberg, New York, 1973), Vol. 66, pp. 1–97.

[8] R. Haag, D. Kastler, and E. Trych-Pohlmeyer, "Stability and Equilibrium States," Commun. Math. Phys. **38**, 173–93 (1974).

[9] M. Aizenmann, G. Gallavotti, S. Goldstein, and J. L. Lebowitz, "Stability and Equilibrium States of Infinite Classical Systems," Commun. Math. Phys. **48**, 1–14 (1976).

[10] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics II* (Springer-Verlag, New York, Heidelberg, Berlin, 1981).

[11] H. Kesten and G. C. Papanicolaou, "A Limit Theorem for Stochastic Acceleration," Commun. Math. Phys. **78**, 19–63 (1980).

[12] E. B. Davies, "Markovian Master Equations," Commun. Math. Phys. **39**, 91–110 (1974).

[13] R. Alicki, "On the Detailed Balance Condition for Non-Hamiltonian Systems," Rep. Math. Phys. **10**, 249–58 (1976).

[14] K. Ito and H. P. McKean, *Diffusion Processes and Their Sample Paths* (Academic, New York, 1965); M. Fukushima, *Dirichlet Forms and Markov Processes* (North-Holland, Amsterdam, Oxford, New York, and Kodansha, Tokyo, 1980), and references quoted therein.

[15] G. Gallavotti and M. Pulvirenti, "Classical KMS Condition and Tomita–Takesaki Theory," Commun. Math. Phys. **46**, 1–9 (1976).

[16] K. Yosida, *Functional Analysis* (Springer-Verlag, Berlin, Heidelberg, New York, 1968), 2nd ed. pp. 413ff.

[17] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. II. Fourier Analysis, Self-Adjointness* (Academic, New York, San Francisco, London, 1975).

[18] R. Holley, "The Motion of a Heavy Particle in an Infinite One-Dimensional Gas of Hard Spheres," Z. Wahrsch. Verw. Gebiete **17**, 181–219 (1971).

[19] D. Dürr, S. Goldstein, and J. L. Lebowitz, "A Mechanical Model of Brownian Motion," Commun. Math. Phys. **78**, 509–30 (1981).

[20] H. Spohn and J. L. Lebowitz, "Stationary Non-Equilibrium States of Infinite Harmonic Systems," Commun. Math. Phys. **54**, 97–120 (1977).

[21]V. Gorini, A. Frigerio, M. Verri, A. Kossakowski, and E. C. G. Sudarshan, "Properties of Quantum Markovian Master Equations," Rep. Math. Phys. **13**, 149–73 (1978).
[22]O. E. Lanford, III and J. L. Lebowitz, "Time Evolution and Ergodic Properties of Harmonic Systems," in *Lecture Notes in Physics* (Springer-Verlag, Berlin, Heidelberg, New York, 1975), Vol. 38, pp. 144–77; O. E. Lanford, III *Time Evolution of Large Classical Systems* (Springer-Verlag, Berlin, Heidelberg, New York, 1975), pp. 1–111.
[23]I. E. Segal, "Transforms for Operators and Symplectic Automorphisms over a Locally Compact Abelian Group," Math. Scand. **13**, 31–43 (1963); I. Daubechies and A. Grossman, "An Integral Transform Related to Quantization," J. Math. Phys. **21**, 2080–90 (1980); I. Daubechies, "On the Distributions Corresponding to Bounded Operators in the Weyl Quantization," Commun. Math. Phys. **75**, 229–38 (1980).
[24]K. Hepp, "The Classical Limit for Quantum Mechanical Correlation Functions," Commun. Math. Phys. **35**, 265–77 (1974).
[25]J. V. Pulé and A. Verbeure, "The Classical Limit of Quantum Dissipative Generators," J. Math. Phys. **20**, 733–5 (1979).
[26]G. Lindblad, "On the Generators of Quantum Dynamical Semigroups," Commun. Math. Phys. **48**, 119–30 (1976).
[27]E. B. Davies, "Quantum Dynamical Semigroups and the Neutron Diffusion Equation," Rep. Math. Phys. **11**, 169–88 (1977); "Generators of Dynamical Semigroups," J. Funct. Anal. **34**, 421–32 (1979).
[28]R. Kubo, "Statistical Mechanical Theory of Irreversible Processes I," J. Phys. Soc. Jpn. **12**, 570–86 (1957).
[29]A. Frigerio, "Stationary States of Quantum Dynamical Semigroups," Commun. Math. Phys. **63**, 269–76 (1978).
[30](a) R. Artuso, "Studio dello stato stazionario per un modello rigoroso di conduzione del calore," thesis, University of Milan, 1983; (b) R. Artuso and V. Gorini, "On the stationary state for a model of heat conduction," in *Proceedings of the Workshop "Mathematical Problems of Statistical Physics," Rome, 17–25 February 1983,* edited by C. Boldrighini, G. F. Dell'Antonio, G. Jona-Lasinio, A. Pellegrinotti, and L. Triolo (to be published); (c) R. Artuso, V. Benza, A. Frigerio, V. Gorini, and E. Montaldi, "The stationary state for a variant of Davies' model of heat conduction," preprint 1983.
[31]H. Araki, "Relative Hamiltonian for Faithful Normal States of a Von Neumann Algebra," Publ. Res. Inst. Math. Sci. Kyoto University **9**, 165–209 (1973).

Dynamical properties of non-Markovian stochastic differential equations

A. Hernández-Machado and M. San Miguel
Departamento de Física Teórica, Universidad de Barcelona, Diagonal 647, Barcelona-28, Spain

(Received 16 November 1982; accepted for publication 17 June 1983)

We study nonstationary non-Markovian processes defined by Langevin-type stochastic differential equations with an Ornstein–Uhlenbeck driving force. We concentrate on the long time limit of the dynamical evolution. We derive an approximate equation for the correlation function of a nonlinear nonstationary non-Markovian process, and we discuss its consequences. Non-Markovicity can introduce a dependence on noise parameters in the dynamics of the correlation function in cases in which it becomes independent of these parameters in the Markovian limit. Several examples are discussed in which the relaxation time increases with respect to the Markovian limit. For a Brownian harmonic oscillator with fluctuating frequency, the non-Markovicity of the process decreases the domain of stability of the system, and it can change an infradamped evolution into an overdamped one.

PACS numbers: 05.40. + j, 02.50.Ey

1. INTRODUCTION

In this paper we study non-Markovian processes (NM) defined by Langevin-type stochastic differential equations. Several authors[1–8,37] have discussed the equation satisfied by the probability density of these processes. The knowledge of this quantity is not enough to decide about the Markovicity or non-Markovicity of a process. We focus our attention in quantities like the correlation functions and relaxation times, which give a characterization of the peculiarities of non-Markovian dynamics. These quantities cannot be obtained from the equation for the probability density. We also study non-Markovian effects in the time-dependent moments and in the stability properties of the process. We consider Langevin-type equations without memory kernels and driven by an Ornstein–Uhlenbeck process. The processes solution of this type of equation is nonstationary besides being non-Markovian (NMNS). In the white noise limit of the Ornstein–Uhlenbeck noise, the process becomes simultaneously Markovian and stationary. Nonstationarity is an additional complication which is not present in NM processes described by other equations of the Langevin type.[3] This paper aims to study the dynamical characterization of those NMNS processes.

The physical motivation of this work is that the type of equation mentioned above is the one used in the description of systems under the influence of external or parametric noise. This description is usually given in terms of phenomenological equations,[9] in which a parameter is substituted by a random process. This random processes is often modeled by an Ornstein–Uhlenbeck process or in a particular limit by a Gaussian white noise.[10,11] Such systems have been studied experimentally,[10] and, in particular, dynamical quantities like relaxation times have been measured.[12] Our results in this paper describe the dynamical response to an external source of noise coupled to the system. They show the main differences in the dynamical behavior of the system with respect to the Markovian limit in which the Ornstein–Uhlenbeck process is replaced by a Gaussian white noise. The dependence of the stationary distribution on the parameters of an Ornstein–Uhlenbeck noise has been determined experimentally.[13] In the same way it is possible to measure the dependence of a relaxation time on these parameters.

A relevant aspect of this paper is that it gives a practical method that allows an explicit calculation of the dependence of the correlation function and the relaxation time on the parameters of the Ornstein–Uhlenbeck noise. This method is based on ideas similar to the ones used to calculate the probability distribution of these processes[5,14,15]: It is based on an approximate calculation, by functional methods, of the response of the system to the stochastic driving force. This response function plays a central role in our development. We note that with this method we are able to deal with nonlinear processes, while the practical usefulness of a direct application of more standard approaches based on cumulant expansions seems to be restricted to linear processes.

A central result of this paper is the derivation of an approximate equation for the correlation function of a nonlinear NMNS process. This equation shows that the dynamics of the process depends on the noise parameters in situations in which it is independent of these parameters in the Markovian limit. This is, for example, the case for a process described by a single variable with "additive noise.[6]" In this case, the Markovian limit of the equation for the correlation function is noise-independent. This also happens for a Brownian harmonic oscillator with stochastic frequency. Another interesting feature of this equation is the existence of a term which cannot be obtained from the equation for the probability density of the process. This term becomes in some cases the dominant non-Markovian effect. We have found that the relaxation time is increased with respect to the Markovian limit for the most general one-variable model with "additive" or "multiplicative noise[6]" and for the Brownian harmonic oscillator with stochastic frequency. The origin of this common behavior is different in both cases. For the Brownian harmonic oscillator with stochastic frequency, we have also found important non-Markovian effects as the decrease in the range of values of the intensity of frequency fluctuations for which the system is energetically stable and the possibility of a change of dynamic regime from an infradamped to an overdamped mode. This possibility is due to an

 0022-2488/84/041066-10$02.50

effective reduction of the frequency of oscillation that also appears for a mechanic oscillator with stochastic frequency.

The outline of the paper is as follows. In Sec. 2 we first discuss the peculiarities of the class of nonstationarity non-Markovian processes that we consider. This is illustrated by means of a simple example. Next we derive approximate equations for the moments and correlation functions of that class of NMNS processes. Sections 3 and 5 are devoted to several examples. We compare our results with those of other approaches. We only consider linear processes (although with multiplicative noise) to concentrate on pure non-Markovian effects independently of nonlinear complications to be studied in future work. Section 3 is devoted to a one-variable model which corresponds to the point reactor kinetic equation.[16,17] Section 4 considers a general N-variable model used in quantum optics.[18,19] A particular case corresponds to the mechanic oscillator with stochastic frequency.[20,21] In Sec. 5 we study the harmonic Brownian oscillator with stochastic frequency.[22–24]

2. CORRELATION FUNCTIONS OF NON-MARKOVIAN-NONSTATIONARY PROCESSES

A. Non-Markovicity and nonstationarity

Time-convolutionless Fokker–Planck equations for the probability density of NM processes have been considered in Refs. 1–7. It has been pointed out[3,5–7] that these are not bonafide Fokker–Planck equations because the conditional probability of the process is not a fundamental solution of the Fokker–Planck equation for two arbitrary times. In general, for an NM process, the conditional probability $\alpha(q, t; q', t')$, $t > t'$, depends not only on t and t', but also on the previous history of the process. Only when $t' = 0$ is $\alpha(q, t; q', 0)$ a solution of the Fokker–Planck equation for the probability density $P(q, t)$ with initial condition $P(q', 0) = \delta(q - q')$. As a consequence, the Fokker–Planck equation can in general be used to calculate correlation functions like $\langle q(t)q(0)\rangle$ but not $\langle q(t)q(t')\rangle$ for arbitrary t'. Here $t = 0$ is taken as the time in which the initial conditions of the NM process are specified. In Ref. 3 an explicit discussion of some of these facts was given considering a Gaussian NM stationary (NMS) process for which exact expressions can be obtained for $\alpha(q, t; q', t')$ and $\langle q(t)q(t')\rangle$.

We are interested in the calculation of the correlation function and the relaxation time in the steady state of a process solution of a stochastic differential equation of the general form

$$\dot{q}(t) = v(q(t)) + g(q(t))\xi(t), \tag{2.1}$$

where $v(q)$ and $g(q)$ are in general nonlinear functions of q. The stochastic force $\xi(t)$ is an Ornstein–Uhlenbeck process: A Gaussian process with zero mean and correlation function

$$\langle \xi(t)\xi(t')\rangle = \gamma(t, t') = (D/\tau)\exp(-|t - t'|/\tau). \tag{2.2}$$

A particular process is defined by (2.1) and a given initial condition at the preparation time $t = 0$. Any solution of (2.1) is non-Markovian due to the fact that $\xi(t)$ is not a white noise.[25] For whatever initial condition, the solution of (2.1) is also a nonstationary process. Therefore, Eq. (2.1) defines a class of NMNS processes. These processes are characterized by being a solution of (2.1) and differ from each other in the choice of initial conditions. The nonstationarity gives rise to some difficulties which are not present in the example discussed in Ref. 3 for an NMS process. We note that in the limit $\tau \to 0$, (2.1) defines a Markovian process which is also stationary when the distribution of initial conditions is chosen as the stationary distribution. This distribution is the one reached for $t \to \infty$ and arbitrary initial conditions. Nevertheless, there do not exist two separate limits in which the process becomes NMS and MNS. Therefore, it is important to realize that, in the particular class of processes defined by (2.1), the effects of non-Markovicity and nonstationarity cannot be disentangled. Both have the same origin, which is the finite correlation time τ of the stochastic force $\xi(t)$. To clarify some features of this class of NMNS processes and the differences with NMS or MNS processes, we first consider a simple explicit example. This example is a Gaussian process defined by (2.1) with

$$v(q) = -aq, \quad a > 0, \quad g(q) = 1. \tag{2.3}$$

Following the methods in Refs. 5, 14, and 15 one can immediately obtain the equation satisfied by the probability density of the process $P(q, t)$. This is a time-convolutionless Fokker–Planck equation with time-dependent diffusion coefficient:

$$\frac{\partial}{\partial t} P(q, t) \equiv L_q(t) P(q, t), \tag{2.4}$$

$$L_q(t) = \frac{\partial}{\partial q} aq + D(t)\frac{\partial^2}{\partial q^2}, \tag{2.5}$$

$$D(t) = \frac{D}{1 + \tau a}\{1 - \exp[-(\tau^{-1} + a)t]\}. \tag{2.6}$$

The steady-state probability density $P_{st}(q)$ is obtained as the limit for $t \to \infty$ of the solution of (2.4) with an arbitrary initial condition: $P_{st}(q) = \lim_{t\to\infty} P(q, t)$. This distribution corresponds to the stationary solution of (2.4) when $D(t)$ is replaced by $D(\infty)$.

The intrinsic nonstationarity of (2.1) is explicitly seen in this example: Whatever the initial condition $P(q, t = 0)$, the solution of (2.4) depends on time during a transient. In particular, even choosing $P(q, t = 0) = P_{st}(q)$, we obtain a nonstationary process. In the example considered by Fox,[3] the equation for $P(q, t)$ also features a time-dependent operator $L_q(t)$. Nevertheless, the process is stationary when the stationary distribution is chosen as initial condition. Mathematically this happens because the time dependence of $L_q(t)$ is only given by a common factor $a(t)$, that is, $L_q(t) = a(t) l_q$. This does not happen in (2.5). The factorization of the time dependence of $L_q(t)$ follows from the assumption of a fluctuation–dissipation relation. In our case (2.1) has to be interpreted as a phenomenological modeling of a system coupled to an external source of noise. This parametric noise is modeled by $\xi(t)$. Therefore, the constant a in (2.3) and the noise parameters D and τ are assumed to be independent.

In passing, we note that there exists a Markovian nonstationary process (MNS) associated with the NMNS process (2.1), (2.2), and (2.3). It is defined by (2.1) and (2.3) and

$$\langle \xi(t)\xi(t')\rangle = 2D(t)\delta(t - t') \tag{2.7}$$

and $D(t)$ given by (2.6). The bonafide Fokker–Planck equation for this Markovian process is also (2.4). This explicitly shows that Markovicity cannot be decided in terms of the equation satisfied by the probability density. For $\tau \to 0$ and D fixed, (2.3) and (2.7) define an ordinary Markovian stationary process which is also the Markovian limit of (2.2)–(2.3).

The calculation of the steady-state correlation function of the example (2.2)–(2.3) is straightforward. Integrating (2.1)–(2.3) with arbitrary initial conditions and using (2.2), we have

$$\langle q(t)q(t')\rangle_{\rm st} = \lim_{\substack{t,t'\to\infty \\ t-t'=s}} \langle q(t)q(t')\rangle = \frac{D\tau}{(1+\tau a)(\tau a-1)} \times \left[\exp\left(\frac{-s}{\tau}\right) - (\tau a)^{-1}\exp(-as)\right]. \tag{2.8}$$

We can also consider the quantity $\langle q(s)\,q(0)\rangle_{\rm st\,IC}$. This is defined as the correlation function of the process (2.3) with initial conditions at $t=0$ given by $P_{st}(q)$. We have

$$\langle q(s)q(0)\rangle_{\rm st\,IC} = \frac{D}{a(1+\tau a)}\exp(-as). \tag{2.9}$$

The two quantities $\langle q(t)\,q(t')\rangle_{\rm st}$ and $\langle q(s)q(0)\rangle_{\rm st\,IC}$ are two correlation functions which can be considered for the process defined by (2.1)–(2.3) and $P(q, t=0) = P_{st}(q)$. They are different because the process is not stationary. Nonstationarity also implies that $\langle q(t+s)q(t)\rangle_{\rm st\,IC} \neq \langle q(s)q(0)\rangle_{\rm st\,IC}$ for a finite t.[26]

The steady-state relaxation time $T_{\rm st}$ is given by

$$T_{\rm st} = \int_0^\infty ds\, \frac{\langle q(t)q(t')\rangle_{\rm st} - \langle q\rangle_{\rm st}^2}{\langle q^2\rangle_{\rm st} - \langle q\rangle_{\rm st}^2} = a^{-1} + \tau. \tag{2.10}$$

Similarly, a relaxation time can be defined for $\langle q(s)\,q(0)\rangle_{\rm st\,IC}$

$$T_0 = \int_0^\infty ds\, \frac{\langle q(s)q(0)\rangle_{\rm st\,IC} - \langle q\rangle_{\rm st}^2}{\langle q^2\rangle_{\rm st} - \langle q\rangle_{\rm st}^2} = a^{-1}. \tag{2.11}$$

In the Markovian limit of the process ($\tau \to 0$), $T_{\rm st}$ and T_0 coincide. The dependence of $T_{\rm st}$ on τ is a dynamical effect which we study in this paper for the class of NMNS processes defined by (2.1). T_0 does not depend on τ because $\langle q(s)q(0)\rangle_{\rm st\,IC}$ only depends on τ through static quantities like $\langle q^2\rangle_{\rm st}$. This dependence cancels in the calculation of T_0.

The inequality of the two correlation functions defined above is an important difference with respect to the NMS example studied in Ref. 3. This inequality is a direct consequence of the intrinsic nonstationarity of the process. In fact, it is easy to see that, in general, this inequality holds for any nonstationary process regardless of its Markovian or non-Markovian character (see below[27]). For NMS processes,[3] the two correlation functions are equal. The existence of this inequality for the NMNS process precludes the use of the operator $L_q(t)$ to obtain an equation for $\langle q(t)\,q(t')\rangle_{\rm st}$. In the following we will derive an approximate equation for this quantity for a general NMNS process specified by (2.1). Let us first summarize the situation for the different cases. We assume that the probability density obeys a Fokker–Planck equation of the form (2.4):

(i) If the process is MNS like the one given by (2.1)–(2.3) and (2.7), then (2.4) is a bonafide Fokker–Planck equation and the conditional probability obeys (2.4) for any t, t'. Thus

$$\frac{d}{dt}\langle q(t)q(t')\rangle = \langle (L_q^+ q)(t)q(t')\rangle, \tag{2.12}$$

where $L_q^+(t)$ is the adjoint operator of $L_q(t)$. This is, of course, true for any Markovian process.

(ii) For an NM process, $\alpha(q, s; q', 0)$ is in general a solution of (2.4):

$$\frac{\partial}{\partial s}\alpha(q, s; q', 0) = L_q(s)\alpha(q, s; q', 0) \tag{2.13}$$

with initial condition $\alpha(q, 0; q', 0) = \delta(q-q')$. Therefore,

$$\frac{d}{ds}\langle q(s)q(0)\rangle = \langle (L_q^+(s)q(s))q(0)\rangle. \tag{2.14}$$

This equation is valid for arbitrary initial conditions.

(iii) Consider now that we choose stationary initial conditions and that the process is then stationary (NMS).[3] For a stationary process we have that $\alpha(q, t; q', t')' = \alpha(q, s; q', 0)$ for arbitrary t' and $t = t' + s$. Therefore, from (2.13)

$$\frac{\partial}{\partial t}\alpha(q, t; q', t') = L_q(s)\alpha(q, t; q', t') \tag{2.15}$$

with

$$\alpha(q, t'; q', t') = \delta(q-q'). \tag{2.16}$$

Thus,

$$\frac{d}{dt}\langle q(t)q(t')\rangle_{\rm st} = \langle (L_q^+(s)q(t))q(t')\rangle_{\rm st} = a(s)\langle (l_q^+ q)(t)q(t')\rangle_{\rm st}, \tag{2.17}$$

where we have used the relation $L_q(t) = a(t)l_q$. For stationary initial conditions in (2.14), (2.17) and (2.14) coincide. This is in agreement with the fact that $\langle q(s)q(0)\rangle_{\rm st\,IC}$ and $\langle q(t)q(t')\rangle_{\rm st}$ coincide for an NMS process. In the example discussed in Ref. 3 the non-Markovicity of the process is explicitly seen because the solution of (2.15)–(2.16) does not satisfy the Chapman–Kolmogoroff equation.

(iv) For the NMNS process (2.1)–(2.2) Eqs. (2.13) and (2.14) remain valid but the argument made above for an NMS process cannot be repeated here. Therefore, no general equation is known for $\langle q(t)q(t')\rangle_{\rm st}$ up to now in this case.

In conclusion, the steady-state correlation function $\langle q(t)q(t')\rangle_{\rm st}$ cannot be identified with $\langle q(s)q(0)\rangle_{\rm st\,IC}$ for a nonstationary process. As a consequence, for an NMNS process we do not have an equation for $\langle q(t)q(t')\rangle_{\rm st}$ expressed only in terms of the Fokker–Planck operator $L_q(t)$. From the point of view of the argument above, the difficulty is the nonstationarity of the process. But we remark that for the class of NMNS processes defined by (2.1), nonstationarity cannot be separated from non-Markovicity. Our goal is now to derive an equation for the correlation function $\langle q(t)q(t')\rangle_{\rm st}$ of this class of NMNS processes.

B. Equations for the correlation functions

In general it is not possible to obtain exact equations for the moments and correlation functions of a process defined by (2.1)–(2.2). Our strategy is then to look for systematic approximation schemes in which the zeroth-order approximation is the Markovian limit $\tau = 0$. We do that by considering τ as a small parameter. Our approximation essentially consists in an expansion in powers of τ. This program was already carried out for the equation satisfied by the probability density $P(q, t)$ in Refs. 5, 14, and 15. From the approximate equation for $P(q, t)$ follows an equation for $\langle q(t)\rangle$. This last equation can be directly obtained as follows: Averaging (2.1) we have

$$\frac{d}{dt}\langle q(t)\rangle = \langle v(q(t))\rangle + \langle g(q(t))\xi(t)\rangle = \langle v(q(t))\rangle + \int_0^t dt_1\, \gamma(t, t_1)\left\langle \frac{\delta g(q(t))}{\delta \xi(t_1)}\right\rangle. \tag{2.18}$$

In the last equality we have used the functional characterization of the Gaussian property of $\xi(t)$.[14,15,28] Performing successive partial integrations over t_1 in (2.18) and using (2.2), we obtain an expansion in powers of τ. This amounts to an approximate calculation of the response function $\delta g(q(t))/\delta\xi(t_1)$ by expanding around $t_1 = t$.

We have

$$\int_0^t dt_1\, \gamma(t, t_1)\left\langle \frac{\delta g(q(t))}{\delta \xi(t_1)}\right\rangle = \frac{D}{\tau}\exp(-t/\tau)\left\{\left[\tau \exp(t_1/\tau) \cdot \left\langle \frac{\delta g(q(t))}{\delta q(t)}\frac{\delta q(t)}{\delta \xi(t_1)}\right\rangle\right]_{t_1=0}^{t_1=t} - \left[\tau^2 \exp\left(\frac{t_1}{\tau}\right) \cdot \left\langle \frac{\delta g(q(t))}{\delta q(t)}\frac{d}{dt_1}\left(\frac{\delta q(t)}{\delta \xi(t_1)}\right)\right\rangle\right]_{t_1=0}^{t_1=t} + \cdots\right\}. \tag{2.19}$$

Recalling that[14,15]

$$\left.\frac{\delta q(t)}{\delta \xi(t_1)}\right|_{t_1\to t} = g(q(t)) \tag{2.20}$$

$$\left.\frac{d}{dt_1}\frac{\delta q(t)}{\delta \xi(t_1)}\right|_{t_1\to t} = v(q(t))\frac{\partial g(q(t))}{\partial q(t)} - \frac{\partial v(q(t))}{\partial q(t)}g(q(t)) \equiv M(q(t)), \tag{2.21}$$

we obtain

$$\int_0^t dt_1\, \gamma(t, t_1)\left\langle \frac{\delta g(q(t))}{\delta \xi(t_1)}\right\rangle \simeq D\left\langle \frac{\partial g(q(t))}{\partial q(t)} g(q(t))\right\rangle - D\tau\left\langle \frac{\partial g(q(t))}{\partial q(t)} M(q(t))\right\rangle - D\exp\left(-\frac{t}{\tau}\right)\left\langle \frac{\partial g(q(t))}{\partial q(t)} \left.\frac{\delta q(t)}{\delta \xi(t_1)}\right|_{t_1=0}\right\rangle, \tag{2.22}$$

where the first terms neglected are proportional to $D\tau^2$ and $D\tau e^{-t/\tau}$. The last term in (2.22) contains the quantity $\delta q(t)/\delta\xi(t_1)|_{t_1=0}$ which is not immediately calculable in general. Nevertheless, this is a transient term that can be safely neglected for $t \gg \tau$. In this approximation and substituting (2.22) in (2.18), we obtain to first order in τ

$$\frac{d}{dt}\langle q(t)\rangle = \langle v(q(t))\rangle + D\left\langle \frac{\partial g(q(t))}{\partial q(t)} g(q(t))\right\rangle - \tau D\left\langle \frac{\partial g(q(t))}{\partial q(t)} M(q(t))\right\rangle. \tag{2.23}$$

The validity of keeping only terms to first order in τ in the above τ expansion depends on the value of τ, the range of values of q of interest and of the particular model. A discussion of this point in a related context is given in Ref. 15. A criterion of validity is obtained comparing the terms of order τ with the following term in the expansion of order τ^2. Past experience with this type of expansion for the calculation of probability densities indicates that it gives good results for reasonably small values of τ.[14]

The equations for the correlation functions can be obtained along the same lines. Since $q(0)$ is statistically independent of $\xi(t)$ the equation for $\langle q(t)q(0)\rangle$ is derived with no extra complication:

$$\frac{d}{ds}\langle q(s)\rangle q(0) = \langle v(q(s))q(0)\rangle + D\left\langle \frac{\partial g(q(s))}{\partial q(s)} g(q(s))q(0)\right\rangle - \tau D\left\langle \frac{\partial g(q(s))}{\partial q(s)} M(q(s))q(0)\right\rangle. \tag{2.24}$$

In (2.24) transient terms have also been neglected. Therefore, this equation cannot be used to calculate T_0 because, by its definition (2.11), T_0 also depends on the time domain $s \lesssim \tau$.

The identical formal structure of (2.23) and (2.24) is a consequence of the fact that both can be obtained from the equation for $P(q, t)$. The equation for $\langle q(t)q(t')\rangle$ requires more care:

$$\frac{d}{dt}\langle q(t)q(t')\rangle = \langle v(q(t))q(t')\rangle + \langle g(q(t))\xi(t)q(t')\rangle. \tag{2.25}$$

Proceeding as we did in (2.18)

$$\langle g(q(t))\xi(t)q(t')\rangle = \int_0^t dt_1\, \gamma(t, t_1)\left\langle \frac{\delta g(q(t))}{\delta \xi(t_1)} q(t')\right\rangle + \int_0^{t'} dt_1\, \gamma(t, t_1)\left\langle g(q(t))\frac{\delta q(t')}{\delta \xi(t_1)}\right\rangle. \tag{2.26}$$

The second term in (2.26) has no counterpart in (2.24) since $\delta q(0)/\delta\xi(t_1) = 0$. The two terms in (2.26) can be calculated again by successive partial integrations. For the first term we have

$$\int_0^t dt_1\, \gamma(t, t_1)\left\langle \frac{\delta g(q(t))}{\delta \xi(t_1)} q(t')\right\rangle = \int_0^t dt_1\, \gamma(t, t_1)\left\langle \frac{\partial g(q(t))}{\partial q(t)}\frac{\delta q(t)}{\delta \xi(t_1)} q(t')\right\rangle \simeq D\left\langle \frac{\partial g(q(t))}{\partial q(t)}\left.\frac{\delta q(t)}{\delta \xi(t_1)}\right|_{t_1=t} q(t')\right\rangle - D\tau\left\langle \frac{\partial g(q(t))}{\partial q(t)}\left.\left(\frac{d}{dt_1}\frac{\delta q(t)}{\delta \xi(t_1)}\right)\right|_{t_1=t} q(t')\right\rangle - D\exp\left(-\frac{t}{\tau}\right)\left\langle \frac{\partial g(q(t))}{\partial q(t)}\left.\frac{\delta q(t)}{\delta \xi(t_1)}\right|_{t_1=0} q(t')\right\rangle, \tag{2.27}$$

where the first terms neglected are proportional to $D\tau^2$ and $D\tau e^{-t/\tau}$. The last term in (2.27) is again a transient term that we neglect. Substituting (2.20) and (2.21), we have to first order in τ

$$\int_0^t dt_1\, \gamma(t,t_1)\left\langle \frac{\delta g(q(t))}{\delta \xi(t_1)} q(t')\right\rangle = D\left\langle \frac{\partial g(q(t))}{\delta q(t)} g(t)q(t')\right\rangle - D\tau\left\langle \frac{\partial g(q(t))}{\partial q(t)} M(q(t))q(t')\right\rangle. \tag{2.28}$$

For the second term in (2.26) we have

$$\int_0^{t'} dt_1\, \gamma(t,t_1)\left\langle g(q(t))\frac{\delta q(t')}{\delta \xi(t_1)}\right\rangle \simeq D\exp\left[-\frac{(t-t')}{\tau}\right]\left\{\left\langle g(q(t))\frac{\delta q(t')}{\delta \xi(t_1)}\Big|_{t_1=t'}\right\rangle - \tau\left\langle g(q(t))\left(\frac{d}{dt_1}\frac{\delta q(t')}{\delta \xi(t_1)}\right)\Big|_{t_1=t'}\right\rangle\right\}$$

$$- D\exp\left(-\frac{t}{\tau}\right)\left\langle g(q(t))\frac{\delta q(t')}{\delta \xi(t_1)}\Big|_{t_1=0}\right\rangle, \tag{2.29}$$

where the first terms neglected are proportional to $D\tau^2 e^{-(t-t')/\tau}$ and $D\tau e^{-t/\tau}$. Once again the last term in (2.29) is a transient term that we neglect. We remark that, given the definition (2.8) in which $t, t' \to \infty$, to neglect transient terms in the calculation of $\langle q(t)q(t')\rangle_{st}$ is not an approximation. In the definition (2.10) of T_{st} we integrate $s = t - t'$ from $s = 0$ to $s = \infty$, and therefore the first two terms in (2.29) cannot be neglected since we cannot assume that $t - t' \gg \tau$. In an expansion in powers of τ aimed to calculate T_{st}, the quantity $e^{-(t-t')/\tau}$ has to be regarded as being of zeroth order in τ. Substituting (2.20) and (2.21) in (2.29), we have to first order

$$\int_0^{t'} dt_1\, \gamma(t,t_1)\left\langle g(q(t))\frac{\delta q(t')}{\delta \xi(t_1)}\right\rangle = D\exp\left[-\frac{(t-t')}{\tau}\right]\{\langle g(q(t))g(q(t'))\rangle - \tau\langle g(q(t))M(q(t'))\rangle\}. \tag{2.30}$$

From (2.25), (2.26), (2.28), and (2.30) we finally have

$$\frac{d}{dt}\langle q(t)q(t')\rangle_{st} = \langle v(q(t))q(t')\rangle_{st} + D\left\langle \frac{\partial g(q(t))}{\partial q(t)} g(q(t))q(t')\right\rangle_{st} - D\tau\left\langle \frac{\partial g(q(t))}{\partial q(t)} M(q(t))q(t')\right\rangle_{st}$$

$$+ D\exp\left[-\frac{(t-t')}{\tau}\right]\{\langle g(q(t))g(q(t'))\rangle_{st} - \tau\langle g(q(t))M(q(t'))\rangle_{st}\}. \tag{2.31}$$

This can be rewritten as

$$\frac{d}{dt}\langle q(t)q(t')\rangle_{st} = \langle (L_q^+(\tau)q)(t)q(t')\rangle_{st} + D\exp\left[-\frac{(t-t')}{\tau}\right]\langle g(q(t))h(q(t'))\rangle_{st}, \tag{2.32}$$

where

$$h(q) = g(q) - \tau M(q) \tag{2.33}$$

and

$$L_q^+(\tau) = \left[v(q) + D\frac{\partial g(q)}{\partial q}h(q)\right]\partial_q + Dg(q)h(q)\,\partial_q^2 \tag{2.34}$$

is the adjoint of the Fokker–Planck operator which appears in the approximate Fokker–Planck equation for the probability density $P(q,t)$ of the process.[14,15] The comments made about the domain of validity of (2.23) apply also for (2.32). This question is further discussed for a particular example in Sec. 3 and in the Appendix.

The two terms in (2.32) contain in a different way the effects of τ being $\neq 0$. The existence of the second term indicates that, as discussed in Sec. 2A, the equation for $\langle q(t)q(t')\rangle_{st}$ cannot be obtained from the equation for the probability density. This term dissappears in the limit $\tau \to 0$. The first term remains in this limit but with $L_q^+(\tau)$ replaced by the Markovian operator $L_q^+(\tau = 0)$. The second term in (2.32) is the main formal difference with (2.24). As seen in (2.29), it appears because $\xi(t)$ and $q(t')$ are correlated for $t > t'$. This correlation dissappears for $\tau = 0$. Such a term vanishes in (2.22) since $\xi(t)$ and $q(0)$ are uncorrelated because the system is prepared at $t = 0$.

Equation (2.32) is the starting point for the dynamical characterization of the steady state of the NMNS processes defined by (2.1) and in particular for the calculation of T_{st}. No general statement about the dependence of T_{st} in τ can be made due to the nonlinear problem involved in the first term in (2.32). A point to remark is the importance of the contributions coming from the second term in (2.32). This is already seen in the case of additive noise [$g(q)$ = const]: The equation for $\langle q(t)q(t')\rangle_{st}$ only depends on the noise parameters D and τ through those terms. Therefore, in the Markovian limit, the dynamics of $\langle q(t)q(t')\rangle_{st}$ is independent of the noise parameters but it is not in the NMNS situation. In the Markovian limit $\langle q(t)q(t')\rangle_{st}$ only depends on the noise parameter D through the initial condition $\langle q^2\rangle_{st}$. Therefore, in this limit, T_{st} is independent of the noise parameters. The importance of the second term in (2.32) is shown in the simple example (2.1)–(2.3). In that example, $\langle g(q(t))\xi(t)q(t')\rangle$ can be calculated exactly, and one obtains an equation whose solution is (2.8). Since T_{st} is linear in τ [see (2.10)], the approximate equation (2.32) leads to a correlation function which reproduces exactly the result (2.10). Explicitly, we have

$$\begin{aligned}\langle g(q(t))\xi(t)q(t')\rangle &= \langle \xi(t)q(t')\rangle \\ &= \frac{D}{1+\tau a}\Big\{\exp\Big[-\frac{(t-t')}{\tau}\Big] \\ &\quad -\exp\Big[-\Big(\frac{t}{\tau}+at'\Big)\Big]\Big\} \\ &\simeq D(1-\tau a)\exp\Big[-\frac{(t-t')}{\tau}\Big].\end{aligned} \tag{2.35}$$

The approximate correlation function solution of (2.32) is the expansion of (2.8) to first order in τ. In this simple example T_{st} is independent of D. The important point to remark is that the τ dependence of T_{st} has its origin in (2.35) which is the term proportional to $\exp[-(t-t')/\tau]$ in (2.32). Therefore, this term originates the τ dependence of T_{st} in this example, and it is the dominant non-Markovian effect.

The equations above are easily generalized to the case of several variables q_i, $i=1,\ldots,N$. We just quote here the results. The generalizations of (2.1) and (2.2) are

$$\dot q_i(t) = v_i(q(t)) + g_{ij}(q(t))\xi_j(t), \tag{2.36}$$

$$\langle \xi_i(t)\rangle = 0,\ \langle \xi_i(t)\xi_j(t')\rangle = \frac{D}{\tau_i}\delta_{ij}\exp\Big[-\frac{|t-t'|}{\tau_i}\Big]. \tag{2.37}$$

We obtain

$$\begin{aligned}\frac{d}{dt}\langle q_i(t)\rangle &= \langle v_i(q(t))\rangle + D\Big\langle \frac{\partial g_{ij}(q(t))}{\partial q_n(t)} g_{nj}(q(t))\Big\rangle \\ &\quad - D\tau_j\Big\langle \frac{\partial g_{ij}(q(t))}{\partial q_n(t)} M_{nj}(q(t))\Big\rangle \\ &\quad - D\tau_j\Big\langle \frac{\partial g_{ij}(q(t))}{\partial q_n(t)} K_{nlj}(q(t))\xi_l(t)\Big\rangle,\end{aligned} \tag{2.38}$$

where[29]

$$M_{nj} = v_l\frac{\partial g_{nj}}{\partial q_l} - \frac{\partial v_n}{\partial q_l}g_{lj}, \tag{2.39}$$

$$K_{nlj} = g_{kl}\frac{\partial g_{nj}}{\partial q_k} - \frac{\partial g_{nl}}{\partial q_k}g_{kj}. \tag{2.40}$$

The last term in (2.38) has to be further elaborated in each particular case. The equation for $\langle q_i(t)q_j(0)\rangle$ is formally identical to (2.38). For $\langle q_i(t)q_j(t')\rangle$ we have[30]

$$\begin{aligned}\frac{d}{dt}\langle q_i(t)q_j(t')\rangle &= \langle v_i(q(t))q_j(t')\rangle \\ &\quad + D\Big\langle \frac{\partial g_{ik}(q(t))}{\partial q_n(t)} g_{nk}(q(t))q_j(t')\Big\rangle \\ &\quad - D\tau_k\Big\langle \frac{\partial g_{ik}(q(t))}{\partial q_n(t)} [M_{nk}(q(t)) \\ &\quad + K_{nlk}(q(t))\xi_l(t)]q_j(t')\Big\rangle \\ &\quad + D\exp\Big[-\frac{(t-t')}{\tau_k}\Big]\{\langle g_{ik}(q(t))g_{jk}(q(t'))\rangle \\ &\quad - \tau_k\langle g_{ik}(q(t))[M_{jk}(q(t')) \\ &\quad + K_{jlk}(q(t'))\xi_l(t')]\rangle\}.\end{aligned} \tag{2.41}$$

3. EXAMPLE I: ONE VARIABLE MODEL WITH MULTIPLICATIVE NOISE

As a first application of the general equations derived above, we consider in this section the model defined by (2.1) and (2.2) with

$$v(q) = -aq + c, \tag{3.1}$$

$$g(q) = -bq. \tag{3.2}$$

a, b, and c are constants, and a is chosen to be positive. This is a simple but nontrivial model because of the multiplicative character of the noise [$g(q)$ being nonconstant]. For $c=0$ the model becomes mathematically much simpler, but it then has a degenerate steady state in the sense that $P(q,t)\to_{t\to\infty} 0$. Besides its intrinsic interest as an illustrative example, this model is the point reactor kinetic equation used to describe the evolution of the number of neutrons in the presence of parametric noise in a nuclear reactor.[16,17] For $c=0$ is a one-variable version of models describing the interaction of an atomic system with a laser whose phase fluctuates.[18,19] The model admits a formal exact solution for $\langle q(t)\rangle$, $\langle q(t)q(t')\rangle$.[33] Here we concentrate in the calculation in first order in τ which gives a more direct way of obtaining concrete information. In this approximation we obtain explicit results which illuminate the general discussion of Sec. 2B.

To calculate T_{st} we need to know $\langle q\rangle_{st}$, $\langle q^2\rangle_{st}$, and $\langle q(t)q(t')\rangle_{st}$. Specifying (2.23) to the model (3.1), (3.2) and setting $(d/dt)\langle q(t)\rangle = 0$, we have

$$\langle q\rangle_{st} = \frac{c(1-\tau Db^2)}{(a-Db^2)} + O(\tau^2). \tag{3.3}$$

An equation for $\langle q^2(t)\rangle$ is obtained following step by step the method used to derive (2.23). We obtain

$$\begin{aligned}\frac{d}{dt}\langle q^2(t)\rangle &= -2a\langle q^2(t)\rangle + 2c\langle q(t)\rangle \\ &\quad + 4Db^2\langle q^2(t)\rangle - 4\tau cDb^2\langle q(t)\rangle\end{aligned} \tag{3.4}$$

and therefore

$$\langle q^2\rangle_{st} = \frac{c^2(1-3\tau Db^2)}{(a-2Db^2)(a-Db^2)} + O(\tau^2). \tag{3.5}$$

We then have to require that $a > 2Db^2$ to guarantee the existence of steady-state values of $\langle q^2\rangle$ and $\langle q(t)q(t')\rangle$.[34] The equation for $\langle q(t)q(t')\rangle_{st}$ follows directly from (2.32) and (3.1)–(3.2) (see the Appendix). Its solution can be written to first order in τ as

$$\begin{aligned}\langle q(t)q(t')\rangle_{st} &= (\langle q^2\rangle_{st} - \langle q\rangle_{st}^2)\exp[(-a+Db^2)(t-t')] \\ &\quad + \langle q\rangle_{st}^2 + \tau Db^2\Big\{1-\exp\Big[-\frac{(t-t')}{\tau}\Big]\Big\} \\ &\quad \times\exp[(-a+Db^2)(t-t')]\langle q^2\rangle_{st} \\ &\quad - \tau Db^2\exp\Big[-\frac{(t-t')}{\tau}\Big] \\ &\quad \times[1-\exp[(-a+Db^2)(t-t')]]\langle q\rangle_{st}^2.\end{aligned} \tag{3.6}$$

A criterion for the validity of Eq. (2.32) for $\langle q(t)q(t')\rangle_{st}$ can be given in this example by comparing the terms of order τ kept in (2.32) with the terms of order τ^2, which are neglected. This comparison is made in the Appendix. We obtain for this model that to justify the approximation it must be $\tau a \ll 1$.

From the definition of T_{st}, (2.10), (3.3), (3.5), and (3.6), we obtain to first order in τ

$$T_{st} = \frac{1}{a - Db^2} + \tau + O(\tau^2). \tag{3.7}$$

The quantity $1/(a - Db^2)$ is the Markovian relaxation time. Therefore, we obtain the same result as that in (2.10): For a small $\tau \neq 0$ a contribution τ is added to the Markovian value of T_{st}. It is easy to see that the terms proportional to $\tau \exp[-(t-t')/\tau]$ in (3.6) contribute in order τ^2 to T_{st}. The τ contribution comes from the term in (3.6) proportional to τ. This term has its origin in the terms proportional to $\exp[-(t-t')/\tau]$ in (2.32). Thus we conclude that for linear processes with additive or multiplicative noise T_{st} depends additively on τ (for small τ) and that this dependence is obtained from the second term in (2.32). For nonlinear processes the first term in (2.32) will also contribute to the τ dependence of T_{st}.

In the simpler case with $c = 0$ it is possible to write explicit exact equations for $\langle q(t)\rangle$ and $\langle q(t)q(t')\rangle$. From (2.18) we have

$$\frac{d}{dt}\langle q(t)\rangle = -a\langle q(t)\rangle - b\int_0^t dt_1 \left\langle \frac{\delta q(t)}{\delta \xi(t_1)}\right\rangle \gamma(t, t_1). \tag{3.8}$$

The response function is calculated from the integral form of (2.1), (3.1), and (3.2) with $c = 0$:

$$\frac{\delta q(t)}{\delta \xi(t_1)} = -bq(t). \tag{3.9}$$

Substituting in (3.8) (see also Ref. 35)

$$\frac{d}{dt}\langle q(t)\rangle = -a\langle q(t)\rangle + Db^2\left[1 - \exp\left(-\frac{t}{\tau}\right)\right]\langle q(t)\rangle. \tag{3.10}$$

Similarly, substituting (3.9) in (2.25) and (2.26), we obtain an exact equation for $\langle q(t)q(t')\rangle$:

$$\begin{aligned}\frac{d}{dt}&\langle q(t)q(t')\rangle \\ &= -a\langle q(t)q(t')\rangle + Db^2\{1 - 2\exp(-t/\tau) \\ &\quad + \exp[-(t-t')/\tau]\}\langle q(t)q(t')\rangle.\end{aligned} \tag{3.11}$$

This equation is in agreement with the one obtained in Ref. 36.

4. EXAMPLE II. *N*-VARIABLE MODEL WITH MULTIPLICATIVE NOISE

We consider here a particular example of (2.36) and (2.37) given by

$$\dot q_i(t) = -\sum_j A_{ij}q_j(t) - \sum_j B_{ij}q_j(t)\xi(t), \quad i,j = 1,\ldots,N. \tag{4.1}$$

A and B are in general noncommuting matrices and $\xi(t)$ is the Ornstein–Uhlenbeck process defined in (2.2). The model is a generalization to N variables of (3.1)–(3.2) with $c = 0$. The complications that appear in this generalization are due to the noncommutativity of A and B. As already mentioned, (4.1) is a general model used in quantum optics.[18,19] For a specific choice of A and B, (4.1) also describes the evolution in phase space of a mechanic oscillator whose frequency fluctuates around a mean value.[20,21] As we did in Sec. 3, we concentrate here on the calculation in first order in τ.

An approximate equation for $\langle q_i(t)\rangle$ follows from (2.38):

$$\begin{aligned}\frac{d}{dt}\langle q_i(t)\rangle = \sum_{j,k}(&-A_{ij} + DB_{ik}B_{kj} \\ &- \tau DB_{ik}[A,B]_{kj})\langle q_j(t)\rangle + O(\tau^2),\end{aligned} \tag{4.2}$$

where $[A, B] = AB - BA$. It is interesting to note that the non-Markovian contribution proportional to τ only appears when A and B do not commute. If A and B commute, no contribution to any order in τ appears in the equation for $\langle q_i(t)\rangle$. For $[A, B] = 0$ the exact equation is a trivial generalization of (3.10):

$$\begin{aligned}\frac{d}{dt}\langle q_i(t)\rangle = \sum_{j,k}\Big(&-A_{ij} + DB_{ik}B_{kj} \\ &\times\left[1 - \exp\left(-\frac{t}{\tau}\right)\right]\Big)\langle q_j(t)\rangle.\end{aligned} \tag{4.3}$$

A different approximation scheme than ours has been proposed in Ref. 18 to obtain an equation for $\langle q_i(t)\rangle$. The first-order equation in such a scheme has been interpreted in terms of an approximation of Bourret's equation.[19] In comparison with our scheme we wish to point out that the first-order approximation in Ref. 18 does not correspond to a first-order approximation in τ. Indeed, when $[A, B] = 0$, the first-order equation of Ref. 18 still contains a correction to the Markovian equation ($\tau = 0$). Such a term does not appear either in the exact equation (4.3) or in our approximation (4.2). On the other hand, Van Kampen's approximation[8] discussed for this model in Ref. 19 reproduces the exact equation (4.3) when $[A, B] = 0$. For $[A, B] \neq 0$ an expansion to first order in τ of van Kampen's equation leads to our result (4.2).

We now consider the equations for the correlation function. The equation for $\langle q_i(t)q_j(0)\rangle$ is formally identical to (4.2). An approximate equation for the correlation function $\langle q_i(t)q_j(t')\rangle$ follows from (2.41):

$$\begin{aligned}\frac{d}{dt}\langle q_i(t)q_j(t')\rangle = \sum_{k,l}(&-A_{ik} + DB_{il}B_{lk} \\ &- \tau DB_{il}[A,B]_{lk})\langle q_k(t)q_j(t')\rangle \\ &+ \sum_{kl} D\exp\left[-\frac{(t-t')}{\tau}\right](B_{ik}B_{jl} \\ &+ \tau B_{ik}[B,A]_{jl})\langle q_k(t)q_l(t')\rangle.\end{aligned} \tag{4.4}$$

This equation is only valid for long times since transient terms have been neglected. In particular, if a steady state exists, (4.4) becomes an equation for $\langle q_i(t)q_j(t')\rangle_{st}$.

A particular case of (4.1) is given by a mechanic oscillator with position q, momentum p, unit mass and stochastic frequency $[\Omega_0^2 + \xi(t)]^{1/2}$ with mean value Ω_0:

$$\dot{q}(t) = p(t), \tag{4.5}$$

$$\dot{p}(t) = -[\Omega_0^2 + \xi(t)]q(t). \tag{4.6}$$

The matrices A, B, and $[A, B]$ are in this case

$$A = \begin{pmatrix} 0 & -1 \\ \Omega_0^2 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{4.7}$$

$$[A, B] = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.8}$$

Equation (4.2) becomes

$$\frac{d}{dt}\langle q(t)\rangle = \langle p(t)\rangle, \tag{4.9}$$

$$\frac{d}{dt}\langle p(t)\rangle = (-\Omega_0^2 + D\tau)\langle q(t)\rangle. \tag{4.10}$$

This means that for small τ and $t \gg \tau$ the mean values oscillate with a frequency $\omega = (\Omega_0^2 - D\tau)^{1/2}$. This effective reduction of the frequency of oscillation is a non-Markovian effect which disappears in the limit $\tau \to 0$. For white noise frequency fluctuations, $\langle q(t)\rangle$ and $\langle p(t)\rangle$ oscillate with the mean frequency Ω_0 independently of the noise intensity D.

The equations for the correlation functions are easily written from (4.4) and (4.7)–(4.8).

5. BROWNIAN OSCILLATOR WITH FLUCTUATION FREQUENCY

The stochastic differential equations that describe the evolution of the position q and momentum p of a Brownian harmonic oscillator of unit mass and fluctuating frequency are[22–24,37]

$$\dot{q}(t) = p(t), \tag{5.1}$$

$$\dot{p}(t) = -2\lambda p(t) - [\Omega_0^2 + \xi(t)]q(t) + \eta(t). \tag{5.2}$$

Here Ω_0 is the mean value of the frequency and $\xi(t)$ the fluctuating part of Ω^2. This fluctuating part is modeled by the Ornstein–Uhlenbeck process (2.2). The damping coefficient is 2λ and $\eta(t)$ is the thermal noise assumed to be Gaussian white noise with zero mean value and satisfying the usual fluctuation–dissipation relation:

$$\langle \eta(t)\eta(t')\rangle = 4\lambda K_B T\,\delta(t - t'). \tag{5.3}$$

The process (5.1) and (5.2) is NMNS due to the fact that $\xi(t)$ is not a white noise. In the following we study non-Markovian dynamical effects (to leading order in the correlation time τ) in the equations for average values, in the stability properties of the system, correlation functions and relaxation time. Equations (5.1)–(5.2) are a particular case of (2.36) in which $\xi_1 = \xi$ and $\xi_2 = \eta$ ($\tau_2 = 0$).

The approximate equation for the first moments follow from (2.38):

$$\frac{d}{dt}\langle q(t)\rangle = \langle p(t)\rangle, \tag{5.4}$$

$$\frac{d}{dt}\langle p(t)\rangle = (-\Omega_0^2 + D\tau)\langle q(t)\rangle - 2\lambda\langle p(t)\rangle. \tag{5.5}$$

The effective reduction of the frequency is the effect already found in (4.10) but here has an important consequence: The solution $\langle q(t)\rangle$ of (5.4)–(5.5) decays to zero in an infradamped oscillatory mode when $\lambda^2 < \Omega_0^2 - D\tau$. It decays to zero in an overdamped mode when $\lambda^2 > \Omega^2 - D\tau$. Therefore, changing the degree of non-Markovicity measured by τ it is possible to change the regime of the oscillator from the overdamped to the infradamped mode and vice versa when $\lambda^2 \simeq \Omega_0^2$.

The equations for the second-order moments are easily obtained by means of the τ expansion used in Sec. 2B. We find

$$\langle q(t)\eta(t)\rangle = 0, \tag{5.6}$$

$$\langle p(t)\eta(t)\rangle = 2\lambda K_B T, \tag{5.7}$$

$$\langle \xi(t) q^2(t)\rangle = -2D\tau\langle q^2(t)\rangle + O(\tau^2), \tag{5.8}$$

$$\langle \xi(t) q(t)p(t)\rangle = -D(1 - 2\lambda\tau)\langle q^2(t)\rangle + O(\tau^2). \tag{5.9}$$

Therefore, the approximate equations to first order in τ are

$$\frac{d}{dt}\langle q^2(t)\rangle = 2\langle p(t)q(t)\rangle, \tag{5.10}$$

$$\frac{d}{dt}\langle p(t)q(t)\rangle = -(\Omega_0^2 - 2D\tau)\langle q^2(t)\rangle + \langle p^2(t)\rangle - 2\lambda\langle p(t)q(t)\rangle, \tag{5.11}$$

$$\frac{d}{dt}\langle p^2(t)\rangle = 2D(1 - 2\lambda\tau)\langle q^2(t)\rangle - 4\lambda\langle p^2(t)\rangle - 2\Omega_0^2\langle p(t)q(t)\rangle + 4\lambda K_B T. \tag{5.12}$$

If a stationary-state value exists for the second-order moments, this is given by the stationary solution of (5.10)–(5.12)

$$\langle pq\rangle_{st} = 0, \tag{5.13}$$

$$\langle p^2\rangle_{st} = \langle q^2\rangle_{st}(\Omega_0^2 - 2D\tau), \tag{5.14}$$

$$\langle q^2\rangle_{st} = \frac{2\lambda K_B T}{2\lambda\Omega_0^2 - D(2\lambda\tau + 1)}. \tag{5.15}$$

Equation (5.15) indicates that a finite stationary value of $\langle q^2\rangle$ is not reached for $t \to \infty$ when $2\lambda\Omega_0^2 - D(2\lambda\tau + 1) < 0$. We then take the stability limit as $D = D_c$, where

$$D_c = \frac{2\lambda\Omega_0^2}{2\lambda\tau + 1} \simeq 2\lambda\Omega_0^2(1 - 2\lambda\tau). \tag{5.16}$$

For $D > D_c$ the oscillator is energetically unstable. It is easy to check that the condition $D < D_c$ guarantees that the linear system (5.10)–(5.12) has no real positive eigenvalue. The non-Markovicity of the process decreases the stability of the system in the sense that D_c is smaller than the critical value of D for $\tau \to 0$.

We now consider the correlation functions in the steady state. From (2.41) we have

$$\frac{d}{dt}\langle q(t)q(t')\rangle_{st} = \langle p(t)q(t')\rangle_{st}, \tag{5.17}$$

$$\frac{d}{dt}\langle p(t)q(t')\rangle_{st} = \left\{-\Omega_0^2 + D\tau + D\tau\exp\left[-\frac{(t-t')}{\tau}\right]\right\} \times \langle q(t)q(t')\rangle_{st} - 2\lambda\langle p(t)q(t')\rangle_{st}, \tag{5.18}$$

$$\frac{d}{dt}\langle q(t)p(t')\rangle_{st} = \langle p(t)p(t')\rangle_{st}, \tag{5.19}$$

$$\frac{d}{dt}\langle p(t)p(t')\rangle_{st}$$
$$= D\exp\left[-\frac{(t-t')}{\tau}\right](1-2\lambda\tau)\langle q(t)q(t')\rangle_{st}$$
$$+\left\{-\Omega_0^2 + D\tau - D\tau\exp\left[-\frac{(t-t')}{\tau}\right]\right\}$$
$$\times\langle q(t)p(t')\rangle_{st} - 2\lambda\langle p(t)p(t')\rangle_{st}. \qquad (5.20)$$

The position–position correlation function solution of (5.17)–(5.20) is given in the overdamped mode by

$$\langle q(t)q(t')\rangle_{st} = \langle q^2\rangle_{st}\exp[-\lambda(t-t')][\cosh\omega(t-t') + (\lambda/2\omega)\sinh\omega(t-t')], \qquad (5.21)$$

where $\omega^2 = \lambda^2 - (\Omega_0^2 - D\tau)$. In the infradamped mode we have an oscillatory decay of the correlation function

$$\langle q(t)q(t')\rangle_{st} = \langle q^2\rangle_{st}\exp[-\lambda(t-t')][\cos\overline{\omega}(t-t') + (\lambda/2\overline{\omega})\sin\overline{\omega}(t-t')], \qquad (5.22)$$

where $\overline{\omega} = (\Omega_0^2 - D\tau) - \lambda^2$. To obtain (5.21) and (5.22), we have neglected the term proportional to $\tau\exp[-(t-t')/\tau]$ in (5.18), which gives a contribution to the correlation function of higher order in τ.

Equations (5.21) and (5.22) exhibit an important non-Markovian dynamical effect: The correlation function depends on the frequency fluctuations not only through the static part $\langle q^2\rangle_{st}$ but also through the effective frequency (ω or $\overline{\omega}$) which characterizes the dynamics. In the Markovian limit $\tau\to 0$, the dynamical evolution of the correlation function is independent of the frequency fluctuations. In this limit the evolution equations for the correlation functions do not depend on the frequency fluctuation. These only come into the expression of $\langle q(t)q(t')\rangle_{st}$ through the initial condition $\langle q^2\rangle_{st}$. The same effect was already discussed in general after (2.32) for a single variable process with additive noise. Here we have a two-variable process with multiplicative noise such that the Markovian dynamics is noise-independent. Analogously to what happens in (2.10), this non-Markovian dynamical effect manifests itself in a dependence of the relaxation time on the frequency fluctuations. This dependence does not exist in the Markovian limit: The relaxation time of $\langle q(t)q(t')\rangle_{st}$ as defined in (2.10) is, in the overdamped mode,

$$T_{st} = \frac{2\lambda}{\Omega_0^2} + \left(\frac{2\lambda}{\Omega_0^2}\right)^2 K_B T\tau + O(\tau^2)$$
$$= T_{st}^{w} + (T_{st}^{w})^2 K_B T\tau + O(\tau^2). \qquad (5.23)$$

Here T_{st}^{w} is the Markovian relaxation time obtained in the limit $\tau\to 0$. This result is qualitatively the same that we found in (2.10) and (3.7), that is, an increase of T_{st} to leading order in τ with respect to the Markovian limit. Nevertheless, this τ dependence does not originate here in the terms proportional to $e^{-(t-t')/\tau}$, which have been neglected in the equations for the correlation functions.

Finally, we wish to compare our results with other studies of the problem of a Brownian harmonic oscillator with fluctuating frequency. In Refs. 23 and 24 the case $\tau = 0$ has been studied. Our results reduce to those of Refs. 23 and 24 in this limit. For $\tau\neq 0$ we have found that the dynamics depends on the frequency fluctuations and that the system is energetically less stable. In Ref. 22, $\xi(t)$ is taken to be a two-state Markov process instead of an Ornstein–Uhlenbeck noise and in Ref. 21 thermal fluctuations modeled by $\eta(t)$ are neglected. Due to the simpler assumption for $\xi(t)$, an exact stability condition was obtained in Ref. 22. To first order in the correlation time of the noise, this condition coincides with our result (5.16). The stability condition of Ref. 21 is just the condition (5.16) with $\tau = 0$. The effect of a finite τ does not appear there due to the crudeness of the approximation. There is no explicit expression for the relaxation time in Ref. 22 that could be compared with (5.23).

ACKNOWLEDGMENTS

We are indebted to J. M. Sancho and L. Pesquera for helpful comments.

APPENDIX

Equation (2.31) becomes, for the model (3.1)–(3.2),

$$\frac{d}{dt}\langle q(t)q(t')\rangle_{st} = (-a + Db^2)\langle q(t)q(t')\rangle_{st}$$
$$+ c(1-\tau Db^2)\langle q\rangle_{st}$$
$$+ D\exp\left[-\frac{(t-t')}{\tau}\right]$$
$$\times(b^2\langle q^2\rangle_{st} - \tau b^2 c\langle q\rangle_{st}). \qquad (A1)$$

Integration of (A1) with (3.3) leads to (3.6) in first order in τ.

We now consider the first terms neglected in (2.31) for this case. The first nontransient terms neglected in (2.27) and (2.29) are, respectively,

$$D\tau^2\left\langle \frac{\partial g(q(t))}{\partial q(t)}\left(\frac{d^2}{dt_1^2}\frac{\delta q(t)}{\delta\xi(t_1)}\right)\Bigg|_{t_1=t} q(t')\right\rangle_{st} \qquad (A2)$$

and

$$D\tau^2\exp\left[-\frac{(t-t')}{\tau}\right]\left\langle g(q(t))\left(\frac{d^2}{dt_1^2}\frac{\delta q(t')}{\delta\xi(t_1)}\right)\Bigg|_{t_1=t'}\right\rangle_{st}. \qquad (A3)$$

For the model (3.1)–(3.2) we have

$$\frac{\delta q(t)}{\delta\xi(t_1)} = -b\exp\left\{-\int_0^t dt'\,[a+b\xi(t')]q(0)\right\}$$
$$-cb\int_0^{t_1}dt'\exp\left\{-\int_{t'}^t dt''\,[a+b\xi(t'')]\right\}$$
$$= -bq(t) + cb\int_{t_1}^t dt'$$
$$\times\exp\left\{-\int_{t'}^t dt''\,[a+b\xi(t'')]\right\}, \qquad (A4)$$

and therefore

$$\frac{d^2}{dt_1^2}\frac{\delta q(t)}{\delta\xi(t_1)}\Bigg|_{t_1=t} = -cb\,[a+b\xi(t)]. \qquad (A5)$$

Substituting in (A2)

$$(A2) = D\tau^2 cb^2[a\langle q\rangle_{st} + b\,\langle \xi(t)q(t')\rangle_{st}]$$
$$= D\tau^2 cb^2\Big[a\langle q\rangle_{st} + b\int_0^{t'} dt_1\,\gamma(t,t_1) \times \Big\langle \frac{\delta q(t')}{\delta\xi(t_1)}\Big\rangle_{st}\Big]$$
$$= D\tau^2 cb^2\Big\{a\langle q\rangle_{st} - D\exp\Big[-\frac{(t-t')}{\tau}\Big]b^2\langle q\rangle_{st} + O(\tau)\Big\}. \quad (A6)$$

Substituting (A5) in (A3),

$$(A3) = D\tau^2 \exp\Big[-\frac{(t-t')}{\tau}\Big]cb^2 \times [a\langle q\rangle_{st} + b\,\langle q(t)\xi(t')\rangle_{st}]$$
$$= D\tau^2 \exp\Big[-\frac{(t-t')}{\tau}\Big]cb^2 \times\Big[a\langle q\rangle_{st} + b\int_0^{t'} dt_1\,\gamma(t',t_1)\Big\langle \frac{\delta q(t)}{\delta\xi(t_1)}\Big\rangle_{st}\Big]$$
$$= D\tau^2 \exp\Big[-\frac{(t-t')}{\tau}\Big]cb^2[a\langle q\rangle_{st} - Db^2 \times\exp[(-a+Db^2)(t-t')]\langle q\rangle_{st} + O(\tau)], \quad (A7)$$

where $\langle \delta q(t)/\delta\xi(t_1)\rangle_{st}$ is calculated from (A4), recalling that $\xi(t)$ is Gaussian and using a trivial cumulant expansion. The first terms neglected in (A1) are given by the addition of (A6) and (A7):

$$D\tau^2 cb^2 a\langle q\rangle_{st} + D\tau^2\exp\Big[-\frac{(t-t')}{\tau}\Big]cb^2 \times\{a - Db^2\langle q\rangle[1+\exp(-a+Db^2)(t-t')]\}. \quad (A8)$$

Comparing these terms of order τ^2 with the corresponding ones of order τ in (A1) our criterion of validity for (A1) gives

$$1 \gg \tau a \quad \text{and} \quad 1 \gg \tau(a-2Db^2); \quad (A9)$$

since $a > 2Db^2$, we finally obtain $\tau \ll a^{-1}$.

[1]S. A. Adelman, J. Chem. Phys. **64**, 124 (1976).
[2]R. F. Fox, J. Stat. Phys. **16**, 259 (1977).
[3]R. F. Fox, J. Math. Phys. **18**, 2331 (1977).
[4]P. Hänggi, H. Thomas, H. Grabert, and P. Talkner, J. Stat. Phys. **18**, 155 (1978).
[5]M. San Miguel and J. M. Sancho, J. Stat. Phys. **22**, 605 (1980).
[6]R. F. Fox, Phys. Rep. **C48**, 171 (1979).
[7]P. Hänggi and H. Thomas, Phys. Rep. **88**, 209 (1982).
[8]N. C. Van Kampen, Phys. Rep. **24**, 171 (1976).
[9]For a microscopic approach to non-Markovian processes see: P. Hänggi and H. Thomas, Z. Phys. B **26**, 85 (1977); H. Grabert, P. Talkner, P. Hänggi, and Z. Phys. B **26**, 398 (1977); H. Grabert, P. Talkner, P. Hänggi, and H. Thomas, Z. Phys. B **29**, 273 (1978); S. Grossman, Z. Phys. B **47**, 251 (1982).
[10]W. Horsthemke and R. Lefever, *Noise-Induced Transitions*, Series in Synergetics (Springer, New York, to be published).
[11]*Stochastic Nonlinear Systems*, edited by R. Lefever and L. Arnold (Springer, New York, 1981), papers by W. Horsthemke, R. Lefever, and M. San Miguel and J. M. Sancho.
[12]S. Kabashima, S. Kogura, T. Kawakubo, and T. Okada, J. Appl. Phys. **50**, 6296 (1979).
[13]J. M. Sancho, M. San Miguel, H. Yamazaki, and T. Kawakubo, Physica A **116**, 5601 (1982).
[14]J. M. Sancho, M. San Miguel, S. L. Katz, and J. D. Gunton, Phys. Rev. A **26**, 1589 (1982).
[15]J. M. Sancho and M. San Miguel, Z. Phys. B **36**, 357 (1980).
[16]E. R. Quabili and M. Karasulu, Ann. Nucl. Energy **6**, 133 (1979).
[17]K. Saito, J. Nucl. Sci. Tech. **17**, 162 (1980).
[18]S. N. Dixit, P. Zoller, and P. Lambropoulos, Phys. Rev. A **21**, 1289 (1980).
[19]P. Zoller, G. Alber, and R. Salvador, Phys. Rev. A **24**, 398 (1981).
[20]U. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Barucha-Reid (Academic, New York, 1968).
[21]R. C. Bourret, Physica **54**, 623 (1971).
[22]R. C. Bourret, U. Frisch, and P. Pouquet, Physica **65**, 303 (1973).
[23]B. J. West, K. Lindenberg, and V. Seshadri, Physica A **102**, 470 (1980); K. Lindenberg, V. Seshadri, and B. J. West, Phys. Rev. A **22**, 2171 (1980).
[24]N. C. Van Kampen, Physica A **102**, 489 (1980).
[25]L. Arnold, *Stochastic Differential Equations* (Wiley, New York, 1974).
[26]It is also possible to think of $\langle q(t)q(t')\rangle_{st}$ as the correlation function of the process defined by (2.1)–(2.3) and arbitrary initial conditions in $t = -\infty$. In this paper we consistently use $t = 0$ as the preparation time.
[27]For the MNS process defined by (2.1), (2.3), and (2.7) the two correlation functions which we have defined coincide accidentally. This does not happen for a general nonlinear MNS process.
[28]E. A. Novikov, Zh. Eksp. Teor. Fiz. **47**, 1919 (1964) [Sov. Phys. JETP **20**, 1290 (1965)].
[29]M. San Miguel and J. M. Sancho, Phys. Lett. A **76**, 97 (1980).
[30]Additional contributions to (2.38) and (2.40) in first order in τ[31,32] vanish in the examples that we will consider. In these examples $K = 0$.
[31]H. Dekker, Phys. Lett. A **90**, 26 (1982).
[32]L. Garrido and J. M. Sancho, Physica A **115**, 479 (1982).
[33]M. A. Rodriguez and L. Pesquera, J. Nucl. Sci. Tech. **20**, 174 (1983).
[34]The divergence for $a = 2Db^2$ is a typical feature of linear multiplicative stochastic processes. It reflects the importance of nonlinear terms in the stability properties. A term proportional to q^3 in (3.1) would eliminate this divergence.
[35]H. Shibata, Y. Takahasi, and N. Hashitsume, J. Stat. Phys. **17**, 171 (1977).
[36]S. Chaturvedi, Phys. Lett. A **90**, 444 (1982).
[37]For recent related work see B. J. West and K. Lindenberg, Phys. Lett. A **95**, 44 (1983).

The generalized function solution of the Fokker–Planck equation for monoenergetic charged particle transport

B. D. Ganapol and E. B. Parlette
Department of Nuclear and Energy Engineering, University of Arizona, Tucson, Arizona 85721

(Received 22 March 1983; accepted for publication 24 June 1983)

The solution of the Fokker–Planck equation describing the motion of an isotropic burst of ions originating at the center of an infinite medium and undergoing small angle deflections is obtained in terms of generalized functions using the multiple collision approach. The resulting solution can be shown to define a functional on the test function space S_α^β for $\alpha \geqslant \frac{1}{2}$ and $\beta \leqslant \frac{1}{2}$ and the space of polynomials specifying the dynamic moments.

PACS numbers: 05.60. + w, 02.30. + g

I. INTRODUCTION

In many fields involving ion transport, the Fokker–Planck characterization continues to enjoy significant popularity.[1–3] To date, however, surprisingly few analytical solutions to model Fokker–Planck transport problems exist. Solutions of this nature, while serving to further our understanding of the mathematical formulation of the physical processes involved, can also provide benchmarks against which numerical solutions can be compared in order to determine the accuracy of the numerical algorithms.[4,6] The focus of this work, therefore, will be on the development of an analytical solution to a simplified problem concerning time-dependent monoenergetic ion transport in an infinite homogeneous medium.

The multiple collision solution technique, which has proven to be very successful in other fields of transport theory,[7,8] will be applied to the time-dependent Fokker–Planck equation for fast ions undergoing small angle scattering without energy loss. The basis for this technique lies in the decomposition of the ion density distribution into its collisional components resulting in a theoretical simplification due to a change in the mathematical character of the governing equation. An expression for the angular density in the class of generalized functions is obtained which, when defined over appropriate test function spaces, defines a valid functional.

II. MULTIPLE COLLISION FORMULATION

For ions emitted isotropically at time $t = 0$ from a plane source located at the position $x = 0$ in a homogeneous infinite medium, the following Fokker–Planck formulation is appropriate if small angle scattering is assumed and energy loss is neglected:

$$\left[\frac{1}{v}\frac{\partial}{\partial t} + \mu\frac{\partial}{\partial x}\right] f(x,\mu,t) = \gamma(v)\frac{\partial}{\partial \mu}(1-\mu^2)\frac{\partial}{\partial \mu} f(x,\mu,t) + \tfrac{1}{2}\delta(x)\delta(t), \tag{1a}$$

$$f(x,\mu,t) = 0, \quad t<0, \tag{1b}$$

$$\lim_{|x|\to\infty} f(x,\mu,t) = 0, \tag{1c}$$

where

$f \equiv$ the ion angular density distribution,

$x \equiv$ the position of ions in the medium $(-\infty < x < \infty)$,

$\mu \equiv$ the cosine of the angle between the ion direction of motion and the positive x direction $(-1 \leqslant \mu \leqslant 1)$,

$t \equiv$ the time after source emission $(0 \leqslant t < \infty)$,

$v \equiv$ the ion velocity,

$\gamma(v) \equiv N\int_{-1}^{1} \sigma_s(\mu_0,v)(1-\mu_0)d\mu_0$,

$\sigma_s(\mu_0,v) \equiv$ the differential scattering cross section for angular deflection,

$N \equiv$ number of scattering centers.

When a series expansion of the form

$$f(x,\mu,t) = \sum_{n=0}^{\infty} \gamma^n f_n(x,\mu,t) \tag{2}$$

is introduced into Eqs. (1), and the coefficients of γ^n are equated, there results the following infinite set of recursive equations:

$$\left[\frac{1}{v}\frac{\partial}{\partial t} + \mu\frac{\partial}{\partial x}\right] f_0(x,\mu,t) = \frac{1}{2}\delta(x)\delta(t), \tag{3a}$$

$$\left[\frac{1}{v}\frac{\partial}{\partial t} + \mu\frac{\partial}{\partial x}\right] f_n(x,\mu,t) = \frac{\partial}{\partial \mu}(1-\mu^2)\frac{\partial}{\partial \mu} f_{n-1}(x,\mu,t) \quad \text{for } n = 1,2,..., \tag{3b}$$

$$f_n(x,\mu,t) = 0, \quad t<0, \quad n = 0,1,..., \tag{3c}$$

$$\lim_{|x|\to\infty} f_n(x,\mu,t) = 0, \quad n = 0,1,... . \tag{3d}$$

Mathematically, the multiple collision formulation corresponds to the singlar perturbation approach where γ is considered a small parameter. Therefore, in general, the series solution representation given by Eq. (2) will not converge but will be asympototic to the solution. In the development to follow, however, it will be shown that Eq. (2) is actually an infinite series of generalized functions and will have, at worst, a finite radius of convergence in an appropriate space of test functions.

III. REDUCED EQUATIONS

By integrating along the particle trajectory, Eqs. (3) are transformed into the following integral form:

$$f_0(x,\mu,t) = (1/2t)\delta(\mu-\eta), \quad \eta \equiv x/vt, \quad t \geqslant 0, \tag{4a}$$

$$f_n(x,\mu,t) = v\int_{0+}^{t} dt'\, q_n(x-\mu v(t-t'),\mu,t'), \quad n=1,2,\ldots, \tag{4b}$$

with

$$q_n(x,\mu,t) \equiv \frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial}{\partial\mu}f_{n-1}(x,\mu,t). \tag{4c}$$

For the first and second collided contributions ($n = 1,2$), we have, for $t > 0$,

$$f_1(x,\mu,t) = vF_1(\mu,\eta), \tag{5a}$$

$$f_2(x,\mu,t) = v^2 tF_2(\mu,\eta), \tag{5b}$$

where

$$F_1(\mu,\eta) \equiv [-(\mu/2)\delta^{(1)}(\mu-\eta) + ((1-\mu^2)/6)\delta^{(2)}(\mu-\eta)], \tag{5c}$$

$$\begin{aligned} F_2(\mu,\eta) = [&(2/3!)\delta^{(1)}(\mu-\eta) + (2/4!)(9\mu^2-4)\delta^{(2)}(\mu-\eta) \\ &- (6^2/5!)\mu(1-\mu^2)\delta^{(3)}(\mu-\eta) \\ &+ (5\cdot 4/6!)(1-\mu^2)^2\delta^{(4)}(\mu-\eta)]. \end{aligned} \tag{5d}$$

It is not surprising that the first and subsequent collided densities are generalized functions (derivatives of delta functions), since the uncollided contribution is a delta function and each scattering operation generates derivatives of the preceeding generation. In other words, the highly directed nature of the density distribution is due to the strongly forward peaked scattering assumed.

Based on the forms of f_0, f_1, and f_2, a natural conjecture for the form of f_n for $t > 0$ is

$$f_n(x,\mu,t) = v(vt)^{n-1}F_n(\mu,\eta), \tag{6}$$

the proof of which follows by induction. Assume the conjectured form for f_{n-1}:

$$f_{n-1}(x,\mu,t) = v(vt)^{n-2}F_{n-1}(\mu,\eta)$$

yielding

$$q_n(x,\mu,t) = v(vt)^{n-2}\frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial F_{n-1}}{\partial\mu}(\mu,\eta).$$

Then, from Eq. (4b),

$$f_n(x,\mu,t) = v\int_{0+}^{t} dt'\, v(vt')^{n-2}\frac{\partial}{\partial\mu}(1-\mu^2)\frac{F_{n-1}}{\partial\mu}(\mu,\eta'),$$

where

$$\eta' \equiv [x - v(t-t')]/t'.$$

Upon a change of variable and the incorporation of the limits into the integrand using step functions θ, we find that the conjecture holds for n with

$$\begin{aligned} F_n(\mu,\eta) \equiv (\eta-\mu)^{n-1}\int_{-\infty}^{\infty} &\frac{[\theta(\eta-\mu)-\theta(\eta-\eta')]}{(\eta'-\mu)^n} \\ &\times \frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial}{\partial\mu}F_{n-1}(\mu,\eta'). \end{aligned} \tag{7}$$

When Eq. (6) is introduced into Eqs. (3b) and (3d), we obtain the following reduced equations for F_n, $n = 1,2,\ldots$:

$$\left[(\mu-\eta)\frac{\partial}{\partial\eta} + n - 1\right]F_n(\mu,\eta) = \frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial}{\partial\mu}F_{n-1}(\mu,\eta), \tag{8a}$$

$$\lim_{|\eta|\to\infty} F_n(\mu,\eta) = 0 \tag{8b}$$

with

$$F_0(\mu,\eta) \equiv \tfrac{1}{2}\delta(\mu-\eta). \tag{8c}$$

These equations are singular and must admit generalized function solutions. In the next section, a transformation will be applied to Eqs. (8) in order to remove the singularity and generate the solution.

IV. SOLUTION OF THE REDUCED EQUATIONS

The reduced equations are most easily solved by introducing the Cauchy representation of F_n in the complex z plane[9]

$$\tilde{F}_n(\mu,z) \equiv \frac{1}{2\pi i}\int_{-\infty}^{\infty} d\eta \frac{F_n(\mu,\eta)}{\eta - z}, \tag{9}$$

which defines a distribution on the test function space B of infinitely differentiable functions which, along with their derivatives, vanish as fast as $1/|\eta|$ as $|\eta|$ goes to infinity.[9] Moreover, $\tilde{F}_n$ is a sectionally analytic function with the real axis as its boundary. Also, $\partial^k \tilde{F}_n/\partial z^k$ vanishes at least as rapidly as $1/|z|^{k+1}$ as z tends to infinity. When this transformation is applied to the reduced equations (8a), they remain operationally invariant with respect to $\tilde{F}_n$,

$$\left[(\mu-z)\frac{\partial}{\partial z} + n - 1\right]\tilde{F}_n(\mu,z) = \frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial}{\partial\mu}\tilde{F}_{n-1}(\mu,z); \tag{10}$$

and, futhermore, the singular nature has been eliminated since z is restricted to the cut plane $\notin(-\infty,\infty)$. F_n is recontructed using one of the Plemelj formulas where one need only form the difference of the boundary values of $\tilde{F}_n$ as z approaches the cut[9]

$$F_n(\mu,\eta) = \tilde{F}_n^{+}(\mu,\eta) - \tilde{F}_n^{-}(\mu,\eta). \tag{11}$$

A recursion relation for the nth derivative of $\tilde{F}_n$ with respect to z is obtained when $n-1$ derivatives of Eq. (10) are taken:

$$(\mu-z)\frac{\partial^n\tilde{F}_n}{\partial z^n}(\mu,z) = \frac{\partial}{\partial\mu}(1-\mu^2)\frac{\partial}{\partial\mu}\frac{\partial^{n-1}\tilde{F}_{n-1}}{\partial z^{n-1}}(\mu,z)$$

which is initiated by

$$\tilde{F}_0(\mu,z) = \frac{1}{2}\frac{1}{2\pi i}\frac{1}{\mu - z},$$

where Eqs. (9) and (8c) have been used. By induction, the nth derivative of F_n can be shown to be

$$\frac{\partial^n\tilde{F}_n}{\partial z^n}(\mu,z) = \frac{1}{2\pi i}\sum_{l=n}^{3n}\frac{A_l^n(\mu)}{(\mu-z)^{l+1}}, \quad \text{for } n = 0,1,\ldots, \tag{12a}$$

where (after some algebra) we have

$$A_l^n(\mu) = \frac{d}{d\mu}(1-\mu^2)\frac{dA_{l-1}^{n-1}(\mu)}{d\mu}$$
$$- (l-1)\left\{\frac{\partial}{\partial\mu}[(1-\mu^2)A_{l-2}^{n-1}(\mu)] + (1-\mu^2)\frac{dA_{l-2}^{n-1}(\mu)}{d\mu}\right\}$$
$$+ (l-1)(l-2)(1-\mu^2)A_{l-3}^{n-1}(\mu) \tag{12b}$$
with

$$A_0^0 = \tfrac{1}{2},$$
$$A_l^n \equiv 0 \text{ for } l<n, \quad l>3n,$$

and, for $n=1$,

$$A_1^1(\mu) = 0, \quad A_2^1(\mu) = \mu, \quad A_3^1(\mu) = 1-\mu^2. \tag{12c}$$

Upon integrating Eq. (12a) n times over the interval $(Re^{i\phi}, z)$, where $0<|\phi|<\pi$ and then taking the limit as R approaches infinity, we find

$$\widetilde{F}_n(\mu,z) = \frac{1}{2\pi i}\sum_{k=1}^{2n}\frac{k!}{(k+n)!}\frac{A_{k+n}^n(\mu)}{(\mu-z)^{k+1}}. \tag{13}$$

Since the boundary values of $(\mu-z)^{-(k+1)}$ are

$$\left[\frac{1}{(\mu-z)^{k+1}}\right]^{\pm} = \mathbf{P}\frac{1}{(\mu-\eta)^{k+1}} \pm i\pi\frac{(-1)^k}{k!}\delta^{(k)}(\mu-\eta),$$

Eq. (11) gives the following expression for F_n in terms of generalized functions:

$$F_n(\mu,\eta) = \sum_{k=0}^{2n}\frac{(-1)^k}{(k+n)!}A_{k+n}^n(\mu)\delta^{(k)}(\mu-\eta). \tag{14}$$

V. FINAL SOLUTION AND CONVERGENCE

From Eqs. (2) and (14), the angular density is given by

$$f(x,\mu,t) = \frac{1}{t}\sum_{n=0}^{\infty}[v\gamma t]^n\sum_{k=0}^{2n}\frac{(-1)^k}{(k+n)!}A_{k+n}^n(\mu)\delta^{(k)}(\mu-\eta). \tag{15}$$

The angular density will therefore be a generalized function of the variable x if, for each element $\phi(x)$ of a test function space, the functional

$$(f,\phi) = \int_{-\infty}^{\infty} dx'\,\phi(x')f(x',\mu,t)$$

exists.[10] The following are two examples of common test function spaces where sufficient conditions can be found for the existence of f as a generalized function.

A. Initial value problem

If f is considered as a Green's function, then the solution $[g(x,\mu,t)]$ to Eqs. (1) for the spatially distributed initial condition (or source)

$$g(x,\mu,0) = \tfrac{1}{2}Q(x)$$

is

$$g(x,\mu,t) = \frac{1}{t}\sum_{n=0}^{\infty}(\gamma vt)^n\sum_{k=0}^{2n}\frac{(-1)^k(vt)^{k+1}}{(k+n)!} \times A_{k+n}^k(\mu)Q^{(k)}(x-\mu vt). \tag{16}$$

The question to be addressed at this point is the following: Under what conditions will the series defining the functional g converge? To determine sufficient conditions for the convergence of Eq. (16), bounds must be specified for both A_{k+n}^k and $Q^{(k)}$. Consider, therefore, that $Q^{(k)}$ belongs to the space of test functions S_α^β which have the following properties:

—infinitely differentiable;
—$|x^jQ^{(k)}(x)| \leqslant CB^k k^{k\beta}j^{j\alpha}$; (17)
—the space is nontrivial for $\alpha+\beta\geqslant 1, \alpha>0, \beta>0$;
—for $\beta<1$, $Q(x)$ is an entire function.

Next a bound on A_{k+n}^k must be found.

Since A_{k+n}^k can be shown to be a k th degree polynomial in μ, A_{k+n}^k is also an entire function when μ is extended into the complex z plane. Therefore from Cauchy's inequality, we have the bound

$$|A_{n+k}^n(z)| \leqslant \chi_{n+k}^n(R), \tag{18}$$

where

$$|z| \leqslant R.$$

In addition, the bound χ_{n+k}^n satisfies [from Eq. (12b)]

$$\chi_{n+k}^n(R) = \frac{2}{r_0}\left[R+\frac{1+R^2}{r_0}\right]\chi_{n+k-1}^{n-1}(R) + 2(n+k-1)\left[R+\frac{1+R^2}{r_0}\right]\chi_{n+k-2}^{n-1}(R) + (n+k-1)(n+k-2)(1+R^2)\chi_{n+k-3}^{n-1}(R) \tag{19a}$$

with

$$\chi_{n+k}^n \equiv 0, \quad k<0, \quad k>2n,$$
$$\chi_k^0 = \tfrac{1}{2}\delta_{k,0},$$
$$0<c\leqslant r_0<R-|z|. \tag{19b}$$

By induction, one can show from Eqs. (19) that χ_{n+k}^n and therefore $|A_{n+k}^n(\mu)|$ is bounded by

$$|A_{n+k}^n(z)| \leqslant (\xi^n/2k^{ks})n^{nr}n^k, \quad n=0,1,\ldots, \; k=0,1,\ldots,2n, \tag{20}$$

where

$$r\geqslant 0, \quad 0\leqslant s\leqslant r/2$$

and

$$\xi(R) = \frac{2}{r_0}\left(R+\frac{1+R^2}{r_0}\right) + 3e^s\left(R+\frac{1+R^2}{r_0}\right) + 3^2(1+R^2)e^{2s}.$$

Thus by choosing

$$R = 1+\epsilon, \quad \epsilon>c,$$

an $r_0>0$ can always be found such that the bound given by Eq. (20) is valid for

$$|z| = |\mu| \leqslant 1.$$

With Eqs. (20) and (17), the series representation given by Eq. (16), can be shown to have the majorant

$$g(x,\mu,t) \leqslant C'\exp[b'(tvB)^{1/(s-\beta)}]\exp[b''(\xi v\gamma t)^{1/(1-r)}] \tag{21a}$$

for

$$r<1,\quad \beta\leqslant s<1/2,$$

$$b'>\frac{s-\beta}{e}\exp\left(\frac{1}{s-\beta}\right),$$

$$b''>\frac{1-r}{e}\exp\left(\frac{1}{1-r}\right),$$

and

$$g<\frac{C''}{(1-e\xi v\gamma t)(1-evtB)} \tag{21b}$$

for

$$r=1,\quad \beta=\tfrac{1}{2},$$

$$vt<\min[1/eB,1/e\xi\gamma].$$

In addition, for the S_α^β spaces to be nontrivial, we must require[10]

$$\alpha\geqslant\tfrac{1}{2}.$$

The above analysis therefore indicates that the multiple collision solution has meaning as a generalized function in the test function space S_α^β for $0<\beta\leqslant\frac{1}{2}$ and $\alpha\geqslant\frac{1}{2}$.

B. Dynamic moments

Another class of test functions over which the distribution f exists is the polynomials that define the dynamic moments

$$M_{2l}(\mu,t)=\int_{-\infty}^{\infty}dx(\mu-\eta)^{2l}f(x,\mu,t),\quad l=0,1,\ldots \tag{22}$$

yielding the representation

$$M_{2l}(\mu,t)=\sum_{n=l}^{\infty}(v\gamma t)^n\frac{(2l)!}{(2l+n)!}A^{\,n}_{2l+n}(\mu). \tag{23}$$

With the bound on $A^{\,n}_{n+k}$ given by Eq. (20), M_{2l} has the majorant

$$M_{2l}\leqslant h_l\exp[b''(v\gamma t\xi)^{1/(1-r)}], \tag{24}$$

and therefore M_{2l} is convergent for $0\leqslant t<\infty$.

VI. CONCLUSION

A generalized function solution of the Fokker–Planck equation has been obtained which is a valid distribution in the test function space S_α^β for $0<\beta\leqslant\frac{1}{2}$, $\alpha\geqslant\frac{1}{2}$ and the space of polynomials which define the dynamic moments. The resulting functionals can be used to generate exact expressions for the moments associated with the ion motion which can then serve as benchmarks for comparison to numerical solutions. The solution methodology has also been applied to the Fokker–Planck equation describing Brownian motion and will be the subject of a future paper.

[1] R. Castaing, Adv. Electron. Electron Phys. **13**, 317 (1960).
[2] T. M. Tran and J. Ligou, Nucl. Sci. Eng. **7.9**, 269 (1981).
[3] K. A. Brickner and S. Jorna, Rev. Mod. Phys. **76**, 325 (1974).
[4] P. A. Haldy and J. Ligou, Nucl. Fusion **17**, 6 (1977).
[5] D. B. Brown and R. E. Ogilvie, J. Appl. Phys. **37**, 4429 (1966).
[6] H. Lanteri, R. Bindi, and P. Rostaing, J. Comput. Phys. **39**, 22 (1981).
[7] B. D. Ganapol and L. M. Grossman, Nucl. Sci. Eng. **52**, 454 (1973).
[8] B. D. Ganapol, JQSRT **22**, 135 (1979).
[9] B. Roos, *Analytic Functions and Distributions in Physics and Engineering* (Wiley, New York, 1969).
[10] I. M. Gel'Fand and G. E. Shilov, *Generalized Functions* (Academic, New York, 1968).

On inequivalent classes of unique-mass–spin relativistic wave equations involving repeated irreducible representations with arbitrary multiplicities

P. M. Mathews[a)]
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
B. Vijayalakshmi
Department of Theoretical Physics, University of Madras, Madras 600 025, India

(Received 4 April 1983; accepted for publication 13 May 1983)

Considering all representations $S(\Lambda)$ of the proper Lorentz group which are equivalent to the direct sum of three unspecified inequivalent irreducible representations, each occurring with arbitrary multiplicity, the question is investigated as to what representations of the above class can support first-order relativistic wave equations for unique-spin, unique-mass particles. An important requirement made from the outset is that the equation shall not be equivalent to any simpler one in the presence of arbitrary interactions. Strong restrictions which result on the irreducible representation content of $S(\Lambda)$ are identified by a consideration of the Jordan canonical form of the matrix β^0 entering such equations. With parity invariance as an additional requirement, it is shown that only 0 and 1 can be physical spins; classes of $S(\Lambda)$ which can lead to new equations for these spins are determined. Finally, the restrictions which are needed to hold down the minimal degree of β^0 to low values ($\leqslant 6$) are also determined.

PACS numbers: 11.10.Qr, 11.30.Cp

1. INTRODUCTION

It has been known for a long time that the introduction of interactions into relativistic wave equations for higher spin particles leads often to problems of consistency.[1,2] The study of consistency problems received a powerful impetus from the work of Velo and Zwanziger,[3] which revealed the existence of a hitherto unsuspected fundamental type of difficulty (noncausality of propagation) in some familiar higher spin theories with external field interactions. The extensive studies which have been made since then have not only revealed further new types of difficulties but also shown that pathologies of one type or another occur in practically all the known higher-spin theories. (References to much of the relevant work may be found in the papers by Wightman,[4] Seetharaman *et al.*,[5] and Mathews *et al.*[6])

These, as well as more recent developments in connection with supergravity theories (see, for instance, the papers by Deser and Witten[7] and Duff and van Nieuwenhuizen[8]) have aroused considerable interest in the search for new relativistic wave equations with a richer structure than the familiar ones. The main feature of such equations[9–13] is the nontrivial multiplicity (i.e., exceeding unity) with which one or more irreducible representations (IR) of the proper Lorentz group appear in the representation $S(\Lambda)$ according to which the wave function transforms. However, most of these equations have turned out to be "barnacled" versions[12,14] of familiar wave equations, and hence no different from them even in the presence of arbitrary interactions; the effort which went into their construction has thus been fruitless.

Since there is an endless number of possibilities for the choice of the representation $S(\Lambda)\sim\Sigma\alpha_\tau D^{(\tau)}(\Lambda)$ according to which ψ is to transform—the $D^{(\tau)}$ being the IR of the proper Lorentz group, α_τ their multiplicities, and the summation being a direct sum—the process of construction and testing, one by one, the equations based on such representations is also endless. To make an appreciable dent on the fundamental problem of determining all possible relativistic wave equations (inequivalent to one another in the interacting situation), one has therefore to try to deal with large classes of equations from which admissible ones would be identified by a systematic application of suitable criteria. The recent work of Cox[15] is a step in this direction; it employs graphical techniques which are particularly useful when repeated IR are not present in $S(\Lambda)$. (Cox's proof of the impossibility of massive spin equations without repeated IR appears to us to be much simpler than the proof given later by Berends *et al.*[16]) However if arbitrarily high multiplicities α_τ are allowed, the problem takes on quite a different complexion. Our aim in this paper is to tackle this problem, which has hardly been attempted so far.

The broad question that we wish to address is that of determining all relativistic wave equations in the first-order form

$$(i\beta^\mu\partial_\mu - m)\psi = 0, \tag{1.1}$$

which:

(i) describe particles of unique physical spin s and mass m (with no degeneracy), and

(ii) are unbarnacled, i.e., will not be equivalent, in the presence of arbitrary interactions, to any simpler equation.

We shall use the term "acceptable theory" to denote any wave equation which satisfies these two conditions; it will be expressly understood that nothing more is implied by this term. It may be mentioned here that the use of condition (ii) is the major new element in our treatment.

We do not have a solution to the above problem in all its generality; indeed that might be too much to expect. But we do find that if the number of *inequivalent* irreducible representations (IIR) in $S(\Lambda)$ is limited to a small number, then one

[a)] Permanent address: Department of Theoretical Physics, University of Madras, Madras 600 025, India.

can make considerable headway in determining the implications of the requirements (i) and (ii). A rudimentary illustration of this was given in an earlier work[17] wherein it was shown that a full determination of all acceptable theories involving just two IIR could be made in an almost trivial way. The present paper deals with the class of all $S(\Lambda)$ which involve three (unspecified) IIR with arbitrary and unspecified multiplicities. The analysis of this case is not trivial, though the mathematical tools required are rather elementary. (They consist of certain properties of nilpotent matrices which are basic, but which the authors have not seen applied in physics hitherto.) It is not clear whether the methods used here would by themselves enable exhaustive classification of equations involving $n = 4,5,\ldots$ IIR; but large subclasses of these also can undoubtedly be dealt with. (A summary statement of a few such results including part of those of the present paper have been published without proofs in a recent note.[18]) It is our hope that most classes of practical interest can eventually be analyzed and any new equations which emerge tested for consistency in the presence of interactions. In any case, we believe that even the present work by itself constitutes a breakthrough in this decades-old field: Never before had a way been found to impose the important physical requirements (i) and (ii) on an infinite class of equations and to determine as end results the values of the physical spin possible in acceptable equations, the sets of Lorentz IR which can occur in them, and the multiplicities with which the IR can enter. (In earlier treatments of classes of equations, such as those of Bhabha[19] and Hurley and Sudarshan,[20] algebraic conditions had been in the foreground, and one had to be content with whatever mass–spin content emerged from the postulated algebra.)

We begin by summarizing in Sec. 2 the basic features of the structure of the matrix β^0 (following from invariance requirements) which are central to our treatment. Most of these are well known, but a few do not seem to have been displayed explicitly or in precise form before. Next, in Sec. 3, we translate our acceptability conditions (i) and (ii) into mathematical requirements on the constituents of β^0. The manifestations of these requirements in the case of theories with three IIR and some of their consequences are dealt with in Sec. 4. Some restrictions on $S(\Lambda)$ which arise from these are deduced in Sec. 5, and the further restrictions which would follow if invariance under space inversion were also required are exhibited in Sec. 6. Section 7 is devoted to an enumeration of all representations $S(\Lambda)$—whether or not they would permit space inversion invariance—which would need to be considered if it were demanded that β^0 should have a minimal equation of degree $\leqslant 6$. Reasons for the special interest of such cases are mentioned at the end of Sec. 3. The closing section presents a discussion of the results obtained and some comments in relation to other work. A summary of certain properties of matrices (especially nilpotent ones) which are essential for our treatment is given in an appendix.

This paper does not deal with the procedures for determination of possible inequivalent skeleton matrices when an admissible $S(\Lambda)$ is given. The execution of that stage of the general program will be illustrated in a separate paper wherein a spin 1 equation belonging to a class newly identified in the present work will be examined in detail. It will be shown that the new equation (which involves repeated IR) will, with minimal electromagnetic interaction, be equivalent to the conventional (Proca)[20] equation with a specific type of *nonminimal* interaction.

2. INVARIANCE CONDITIONS AND THE STRUCTURE OF β^0

A. Invariance under the proper Lorentz group

The mathematical constraints into which the requirements (i) and (ii) of Sec. 1 translate can be expressed most conveniently as conditions on the matrix β^0 in the canonical representation. The basis states of the representation are labeled by τ, j, and λ which identify the Lorentz IR and the quantum numbers of J^2 and J_z to which a state belongs. (When the IR τ occurs with a multiplicity $\alpha_\tau > 1$, an additional label a identifies a particular one of these.) Lorentz invariance of Eq. (1.1) requires that the matrix elements of β^0 should have the following structure:

$$\langle \tau' a' j' \lambda' | \beta^0 | \tau a j \lambda \rangle = c_{a'a}^{(\tau'\tau)} g_j^{(\tau'\tau)} \delta_{j'j} \delta_{\lambda'\lambda}. \tag{2.1}$$

Here, the $c_{a'a}^{(\tau'\tau)}$ are the *reduced* or *skeleton* matrix elements, which may take arbitrary (possibly complex) values. The $g_j^{(\tau'\tau)}$ are Lorentz group Clebsch–Gordan coefficients. With $\tau = (m,n)$ and $\tau' = (m',n')$, $g_j^{(\tau'\tau)}$ can be nonvanishing only if *either*

$$(m' + n') = (m + n), \quad (m' - n') = (m - n) \pm 1$$

and

$$\max(|m - n|, |m' - n'|) \leqslant j \leqslant (m + n), \tag{2.2a}$$

in which case

$$g_j^{(\tau'\tau)} = [j(j + 1) - (m - n)(m' - n')]^{1/2} \tag{2.2b}$$

or

$$(m' - n') = (m - n), \quad (m' + n') = (m + n) \pm 1$$

and

$$|m - n| \leqslant j \leqslant \min((m + n), (m' + n')) \tag{2.3a}$$

in which case

$$g_j^{(\tau'\tau)} = [(m + n + 1)(m' + n' + 1) - j(j + 1)]^{1/2}. \tag{2.3b}$$

It may be noted that

$$g_j^{(\tau\tau')} = g_j^{(\tau'\tau)} = g_j^{(\dot\tau\dot\tau')} = g_j^{(\dot\tau'\dot\tau)}.$$

[If $\tau \equiv (m,n)$, its conjugate IR is $\dot\tau \equiv (n,m)$.] A pair of IR τ, τ' is said to be *linkable* if and only if there is at least one j for which $g_j^{(\tau'\tau)} \neq 0$. Whenever a particular pair τ, τ' is not linkable, the parameters $c_{aa'}^{(\tau\tau')}$ associated with this pair may be taken to vanish without any loss of generality, and it will be understood that this is always done.

According to Eq. (2.1), β^0 is block-diagonal:

$$\beta^0 = \sum_{\oplus} \beta^0_{(j)} \otimes I_j. \tag{2.4}$$

Each term in the direct sum here pertains to a particular spin

j; I_j is the $(2j+1)$-dimensional unit matrix (elements $\delta_{\lambda'\lambda}$); and $\beta^0_{(j)}$, which is called the spin-j block or just the j-block, is made up of an array of subblocks $C^{(\tau'\tau)}g_j^{(\tau'\tau)}$ associated with all those IR pairs τ',τ for which $g_j^{(\tau'\tau)}\neq 0$:

$$\langle \tau' a'|\beta^0_{(j)}|\tau a\rangle \equiv \langle a'|C^{(\tau'\tau)}|a\rangle g_j^{(\tau'\tau)} \equiv c_{a'a}^{(\tau'\tau)} g_j^{(\tau'\tau)}. \tag{2.5}$$

$C^{(\tau'\tau)}$ is an $\alpha_{\tau'}\times\alpha_\tau$ matrix, where α_τ is the number of times the IR τ appears in $S(\Lambda)$; and the skeleton matrix C is constituted by the array of blocks $C^{(\tau'\tau)}$ [with τ',τ ranging over all the IIR contained in $S(\Lambda)$]:

$$C=\|C^{(\tau'\tau)}\|. \tag{2.6}$$

It will be noted that $\beta^0_{(j)}$ for a given j may be obtained from C by deleting all $C^{(\tau'\tau)}$ corresponding to every τ (and every τ') which does not contain j, and then multiplying the remaining $C^{(\tau'\tau)}$ by the respective $g_j^{(\tau'\tau)}$.

As has been observed by Cox,[21] it is always possible to separate the IR in $S(\Lambda)$ into two sets (say K and L) such that no two IR belonging to the same set are linkable. Numbering the IIR of the set K as $\tau=1,2,\ldots,k$ and those of L as $k+1,k+2,\ldots,k+l$, one gets the following structure for the skeleton matrix C:

$$C=\begin{pmatrix}0 & U\\ \overline{U} & 0\end{pmatrix},$$
$$U=\|C^{(\tau'\tau)}\|,\quad \tau'=1,2,\ldots,k,\quad \tau=k+1,\ldots,k+l,$$
$$\overline{U}=\|C^{(\tau'\tau)}\|,\quad \tau'=k+1,\ldots,k+l,\quad \tau=1,2,\ldots,k. \tag{2.7}$$

The j-blocks also then take corresponding forms:

$$\beta^0_{(j)}=\begin{pmatrix}0 & U_j\\ \overline{U}_j & 0\end{pmatrix},$$
$$U_j=\|C^{(\tau'\tau)}g_j^{(\tau'\tau)}\|,\quad \tau'\in K,\ \tau\in L,$$
$$\overline{U}_j=\|C^{(\tau'\tau)}g_j^{(\tau'\tau)}\|,\quad \tau'\in L,\ \tau\in K. \tag{2.8}$$

While U_j and $\overline{U}_j$ need not in general be square matrices, X_j and Y_j defined by the following equations are square:

$$X_j\equiv U_j\overline{U}_j,\quad Y_j\equiv\overline{U}_jU_j,\quad (\beta^0_{(j)})^2=\begin{pmatrix}X_j & 0\\ 0 & Y_j\end{pmatrix}. \tag{2.9}$$

B. Invariance under space inversion

By a suitable definition of phase factors appearing in the definition of the parity operator and taking the IR in a suitable order the requirements for parity invariance can be reduced to the following conditions:

$$(\epsilon^{\tau'}-\epsilon^{\tau})C^{(\tau'\tau)}=0\quad\text{if } \tau=\dot\tau \text{ and } \tau'=\dot\tau', \tag{2.10a}$$
$$C^{(\dot\tau'\tau)}=\epsilon^{\tau}C^{(\tau'\tau)}\quad\text{and}\quad C^{(\tau\dot\tau')}=\epsilon^{\tau}C^{(\tau\tau')}$$
$$\text{if } \tau=\dot\tau,\ \tau'\neq\dot\tau', \tag{2.10b}$$
$$C^{(\dot\tau'\dot\tau)}=C^{(\tau'\tau)}\quad\text{if } \tau\neq\dot\tau,\ \tau'\neq\dot\tau'. \tag{2.10c}$$

The factor ϵ^τ associated with every *self-conjugate* IR in the above is defined to be

$\epsilon^\tau=1$ for scalars, polar vectors,$\cdots$,

$\epsilon^\tau=-1$ for pseudoscalars, pseudovectors,$\cdots$.

It will be noted that when both τ and τ' are self-conjugate (and $\epsilon^\tau=\epsilon^{\tau'}$) there is no restriction on $C^{(\tau'\tau)}$ while, in all other cases, different blocks of this skeleton matrix become interrelated. In theories for half-integer spin, $S(\Lambda)$ cannot include any self-conjugate IR; if τ is in the set K, $\dot\tau$ is in L and vice-versa. As a result, the conditions (2.10) for parity invariance reduce to a requirement that, in Eqs. (2.7) and (2.8),

$$\overline{U}=U,\quad \overline{U}_j=U_j. \tag{2.11}$$

In the integer-spin case, the IR contained in each of the sets K and L fall into three classes: self-conjugate IR and two mutually conjugate sets of non-self-conjugate IR. Consequently, the structure of U and $\overline{U}$ can be reduced to

$$U=\begin{pmatrix}P & Q & \epsilon Q\\ R & S & T\\ \epsilon R & T & S\end{pmatrix},\quad \overline{U}=\begin{pmatrix}\overline{P} & \overline{R} & \epsilon\overline{R}\\ \overline{Q} & \overline{S} & \overline{T}\\ \epsilon\overline{Q} & \overline{T} & \overline{S}\end{pmatrix}, \tag{2.12}$$

where the first row of U pertains to self-conjugate IR present in K, the other two rows to the remaining two classes of IR of K, and the columns of U are similarly associated with the three classes of IR of L. In $\overline{U}$, the roles of K and L are interchanged. There is no necessary relation between the blocks of U and those of $\overline{U}$. The ϵ in (2.12) is the parity factor of the self-conjugate IR, which, according to (2.10), must be the same for all such IR.

3. PHYSICAL REQUIREMENTS

A. Inequivalence to simpler equations

In order to avoid the appearance of barnacles [which would make Eq. (1.1) equivalent to another with β^μ of smaller size even when arbitrary interactions are included], it is necessary and sufficient[14] that the following conditions be met.

Let $C^{(\tau\cdot)}$ be the α_τ-row matrix consisting of the row blocks $C^{(\tau\tau')}$ for given τ and all τ', and let $C^{(\cdot\tau)}$ be the α_τ-column matrix similarly defined. Then the requisite conditions are that

$$\mathrm{Rank}(C^{(\tau\cdot)})=\mathrm{Rank}(C^{(\cdot\tau)})=\alpha_\tau \tag{3.1}$$

for every τ present in $S(\Lambda)$. These are automatically fulfilled if no IR is repeated but impose significant restrictions on possible theories with repeated IR. The conditions (3.1) require that for every τ

$$\alpha_\tau\leqslant\sum_{\tau'}\mathrm{Rank}(C^{(\tau\tau')})\leqslant\sum_{\tau'}{}^{(r)}\alpha_{\tau'}, \tag{3.2a}$$

$$\alpha_\tau\leqslant\sum_{\tau'}\mathrm{Rank}(C^{(\tau'\tau)})\leqslant\sum_{\tau'}{}^{(c)}\alpha_{\tau'}, \tag{3.2b}$$

where the summation in (3.2a) is over those τ' for which there are *nonvanishing* $C^{(\tau'\tau)}$ standing in the τ *row* of C—indicated by the superscript (r)—and the sum in (3.2b) is with reference to the τ *column* of C.

B. Unique mass and spin, and nondegeneracy

The necessary and sufficient conditions for uniqueness and nondegeneracy are that all spin blocks $\beta^0_{(j)}$ with $j\neq s$ be nilpotent while $\beta^0_{(s)}$ must also have the nonzero values $+1$ and -1, occurring just once each. These are equivalent to demanding that the minimal equations of the j-blocks be of the form

$$(\beta^0_{(j)})^{l_j} = (\beta^0_{(j)})^{l_j+2}\delta_{js} \tag{3.3}$$

and further that

$$\mathrm{Tr}(\beta^0_{(j)})^2 = 2\delta_{js}. \tag{3.4}$$

Correspondingly, the square matrices X_j and Y_j defined in (2.9) have to satisfy

$$X_j^{p_j} = X_j^{p_j+1}\delta_{js}, \quad Y_j^{q_j} = Y_j^{q_j+1}\delta_{js}, \tag{3.5}$$

$$\mathrm{Tr}\, X_j = \mathrm{Tr}\, Y_j = \delta_{js} \tag{3.6}$$

with

$$p_j - q_j = 1,0, \text{ or } -1. \tag{3.7}$$

This last property may be readily inferred from the fact that, by virtue of the definitions (2.9), powers of X_j and Y_j are interrelated: $X_j^r = U_j Y_j^{r-1}\overline{U}_j$; $Y_j^r = \overline{U}_j X_j^{r-1} U_j$. As regards the relation between l_j and the corresponding p_j and q_j, there are two possibilities in view of item 1(b) of the Appendix and Eq. (3.7):

$$l_j = p_j + q_j \quad \text{or} \quad l_j = p_j + q_j + 1. \tag{3.8}$$

An obvious consequence of the requirements (3.5) and (3.6) is the following useful lemma.

Lemma 1: $X_j(Y_j)$ cannot be nonsingular in an acceptable theory unless the dimension of $X_j(Y_j)$ is unity and j is the physical spin s.

It may be noted here that the rather cumbersome "root method" of Sokatchev[22] is really nothing more than the use of the condition (3.3), which was pioneered by Capri.[23] Equation (3.3) implies the Harish-Chandra condition[24] that the minimal equation of β^0 be of the form

$$(\beta^0)^{l+2} = (\beta^0)^l.$$

The value of l is $\max(l_j)$. It determines directly the number of levels of constraints (primary, secondary,...) implied by the wave equation, and it appears that if l is large, the equation is prone to a variety of inconsistencies in the presence of interactions.[6,25] For this reason, equations characterized by low values of l are of special interest. However, there is a lower bound[26] on l, depending on the $S(\Lambda)$ involved.

4. THREE-IIR THEORIES: SPECIAL AND COMMON SPINS

Let $\tau = 1,2,3$ stand for three inequivalent IR, of which the last two are linked to the first. Then 2 and 3 cannot be linked to each other, and the most general form of the skeleton matrix is

$$C = \begin{array}{cc} \begin{array}{ccc} \alpha_1 & \alpha_2 & \alpha_3 \end{array} & \\ \begin{pmatrix} 0 & Q & R \\ \overline{Q} & 0 & 0 \\ \overline{R} & 0 & 0 \end{pmatrix} & \begin{array}{c} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \end{array},$$

where $Q \equiv C^{(1,2)}$ is an $\alpha_1 \times \alpha_2$ matrix, and so on.

The conditions (3.1) demand in the present case that

$$\mathrm{Rank}(Q) = \mathrm{Rank}(\overline{Q}) = \alpha_2, \quad \mathrm{Rank}(R) = \mathrm{Rank}(\overline{R}) = \alpha_3, \tag{4.1a}$$

$$\mathrm{Rank}(Q,R) = \mathrm{Rank}\begin{pmatrix} \overline{Q} \\ \overline{R} \end{pmatrix} = \alpha_1, \tag{4.1b}$$

and hence that

$$\alpha_2, \alpha_3 \leqslant \alpha_1 \leqslant \alpha_2 + \alpha_3. \tag{4.2}$$

In particular,

$$\alpha_1 = 1 \quad \text{implies} \quad \alpha_2 = \alpha_3 = 1. \tag{4.3}$$

Lemma 2: In an acceptable 3-IIR theory in which α_1 is equal to $\alpha_2(\alpha_3)$, Q and $\overline{Q}$ (R and $\overline{R}$) must be nonsingular.

This follows trivially from the fact that if $\alpha_1 = \alpha_2$, then Q and $\overline{Q}$ are square matrices with rank [given by Eq. (4.1a)] equal to dimension.

In seeking to implement the conditions on j-blocks which arise from mass–spin requirements, one observes that the nature of a particular j-block depends on whether this j is contained in all three of the IR or in just two, or in only one of them. We shall refer to spins falling under these three types as *common*, *special*, and *trivial* spins, respectively. Trivial spins are ignorable as the corresponding j-blocks are null.

The values of the special and common spins in the six different possible linkage types [i.e., orientations of the 1–2 and 1–3 linkages in the (m,n) plane] are shown in Table I. It may be noted that in one of the linkage types (type V) there are two special spins, one occurring in the IR 1 and 2, the other in 1 and 3. In all other types there is no more than one special spin, occurring in IR 1 and in another IR which will by convention be chosen as 2. It will be taken for granted henceforth that the IR in which any special spin occurs are 1 and 2 (all results being applicable with self-evident changes of wording to the case where it occurs in 1 and 3).

We shall now note the forms of the j-blocks for special and common spins and also the implications of the "no-barnacle" conditions (4.1) for certain properties of the matrices—specifically, the ranks of $\beta^0_{(j)}$, X_j, and Y_j and hence the number of nilpotent irreducible Jordan blocks (niJb) which occur in the Jordan canonical form of each.

(*a*) *When j is a special spin:*

$$\beta^0_{(j)} = \begin{pmatrix} 0 & g_j Q \\ g_j \overline{Q} & 0 \end{pmatrix} \tag{4.4}$$

with $g_j = g_j^{(1,2)}$. Recalling that Q and $\overline{Q}$ have dimensions $\alpha_1 \times \alpha_2$ and $\alpha_2 \times \alpha_1$, respectively, and that both must have rank α_2 by (4.1a), we can readily deduce the properties listed in Table II with the aid of items 2 and 3 of the Appendix. These properties have an important role in the developments of later sections. One can also prove with the aid of Eqs. (4.1) that in the case of special spin blocks, (3.7) can be tightened to

$$p_j - q_j = +1$$

except that

$$p_j - q_j = 0 \quad \text{if } \alpha_1 = \alpha_2 = \alpha_3 = 1.$$

(*b*) *When j is a common spin:*

$$\beta^0_{(j)} = \begin{pmatrix} 0 & g_j Q & g_j' R \\ g_j \overline{Q} & 0 & 0 \\ g_j' \overline{R} & 0 & 0 \end{pmatrix} \tag{4.5}$$

with $g_j' \equiv g_j^{(1,3)}$. Noting that $(g_j Q, g_j' R)$ is an $\alpha_1 \times (\alpha_2 + \alpha_3)$ matrix of rank α_1—same as that of (Q,R)—one can infer the properties listed in Table III. It may further be shown that

TABLE I. Occurrence of special and common spins. [Notation: The IR 1 is (m, n); $(+\ -)$ stands for $(m+1/2, n-1/2)$, and so on. It is presumed that $m \leqslant n$. Other cases follow on taking the conjugates of all three IR.]

Linkage type	IR 2	IR 3	Special spins	Common spins
Ia$(m=n)$	$(-\ +)$	$(+\ -)$	nil	$1 \leqslant j \leqslant 2m$
Ib$(m>n)$	$(-\ +)$	$(+\ -)$	$(m-n)$	$(m-n+1) \leqslant j \leqslant (m+n)$
II	$(+\ +)$	$(-\ -)$	$(m+n)$	$(m-n) \leqslant j \leqslant (m+n-1)$
III	$(+\ +)$	$(+\ -)$	$(m-n)$	$(m-n+1) \leqslant j \leqslant (m+n)$
IV$(m>n)$	$(+\ +)$	$(-\ +)$	nil	$(m-n) \leqslant j \leqslant (m+n)$
V	$(+\ -)$	$(-\ -)$	$(m+n)$, $(m-n)$*	$(m-n+1) \leqslant j \leqslant (m+n-1)$
VI$(m>n)$	$(-\ +)$	$(-\ -)$	$(m+n)$	$(m-n) \leqslant j \leqslant (m+n-1)$

Note: All the special spins occur in IR 1 and 2, except the one indicated by an asterisk, which occurs in 1 and 3.

$$p_j - q_j = -1$$

in this case, with the aid of items 4 and 5 of the Appendix.

5. SOME RESTRICTIONS ON $S(\Lambda)$ ARISING FROM THE ACCEPTABILITY CONDITIONS

We are now in a position to prove that the acceptability conditions (i) and (ii) impose strong restrictions on the representation $S(\Lambda)$.

Theorem 1: No acceptable theory can be based on a representation $S(\Lambda)$ which contains more than two nontrivial spins.

Proof: The requirement (3.6) demands, when $\beta^0_{(j)}$ is given by (4.5) or (4.4), that

$$g_j^2 \operatorname{Tr} Q\overline{Q} + g_j'^2 \operatorname{Tr} R\overline{R} = \delta_{js}. \tag{5.1}$$

(This includes special spins j too, for which $g_j' = 0$.)

Two cases have to be considered.

Case (a): If the IR 1 is self-conjugate, and 2 and 3 are mutually conjugate ($1 = \dot{1}$ and $2 = \dot{3}$, which is the case only in the linkage type Ia), then

$$g_j' = g_j \quad \text{for all } j \tag{5.2}$$

by virtue of (2.2). Then Eq. (5.1) requires that

$$\operatorname{Tr}(Q\overline{Q} + R\overline{R}) = \delta_{js}/g_j^2. \tag{5.3}$$

Once the value of the left-hand member is chosen to validate this equation for $j = s$, the equation cannot hold for any other j. Thus the IR 1,2,3 have to be such that only a single nontrivial spin occurs. Reference to the last two columns of Table I show that the only set of IR allowed in this category is

$$1 \equiv (\tfrac{1}{2}, \tfrac{1}{2}), \quad 2 \equiv (0,1), \quad 3 \equiv (1,0).$$

These stand, of course, for a vector field and an antisymmetric tensor field.

Case (b): In any linkage type other than Ia, if j and J are any two spins contained in $S(\Lambda)$,

$$(g_j'/g_j)^2 \neq (g_J'/g_J)^2 \quad \text{for } j \neq J. \tag{5.4}$$

Consider then a nontrivial spin $J \neq s$ of $S(\Lambda)$; application of Eq. (5.1) to J and s yields the requirement that

$$\operatorname{Tr} Q\overline{Q} = -g_J'^2/(g_J'^2 g_s^2 - g_s'^2 g_J^2),$$
$$\operatorname{Tr} R\overline{R} = g_J'^2/(g_J'^2 g_s^2 - g_s'^2 g_J^2). \tag{5.5}$$

If any other spin $j\ (\neq J,s)$ were to exist in $S(\Lambda)$, Eq. (5.1) would require further that $g_j^2 \operatorname{Tr} Q\overline{Q} + g_j'^2 \operatorname{Tr} R\overline{R} = 0$, but this cannot be, on account of (5.4) and (5.5). Thus $S(\Lambda)$ must be such as to contain no more than two nontrivial spins, which proves the theorem.

What this theorem implies is that (in the notation of Table I) n can have no value other than 1/2, except in linkage type IV where $n = 0$ also is allowed.[27] [In this exceptional case, $S(\Lambda)$ contains only a single spin, $j = m$, and therefore $\operatorname{Tr} Q\overline{Q}$ and $\operatorname{Tr} R\overline{R}$ are not independently determined by (5.1), as in type Ia.]

Theorem 2: An acceptable theory cannot be based on a representation $S(\Lambda)$ unless the multiplicities of the three IR in $S(\Lambda)$ are such that

$$\alpha_2 + \alpha_3 > \alpha_1 \tag{5.6}$$

if $S(\Lambda)$ admits a common spin, and either

$$\alpha_1 > \alpha_2 \tag{5.7a}$$

or

$$\alpha_1 = \alpha_2 = \alpha_3 = 1 \quad \text{and} \quad j = s \tag{5.7b}$$

if $S(\Lambda)$ admits a special spin j.

Proof: Referring to Table III, we note that the matrix Y_j associated with a common-spin block has rank α_1, of which

TABLE II. Properties of the j-block for a special spin.

Matrix	Dimension	Rank	No. of niJb
$\beta^0_{(j)}$	$\alpha_1 + \alpha_2$	$2\alpha_2$	$\alpha_1 - \alpha_2$
X_j	α_1	α_2	$\alpha_1 - \alpha_2$
Y_j	α_2	$\geqslant \max(0, 2\alpha_2 - \alpha_1)$ $\leqslant \alpha_2$	$\alpha_2 -$ rank

TABLE III. Properties of the j-block for a common spin.

Matrix	Dimension	Rank	No. of niJb
$\beta^0_{(j)}$	$\alpha_1 + \alpha_2 + \alpha_3$	$2\alpha_1$	$\alpha_2 + \alpha_3 - \alpha_1$
X_j	α_1	$\leqslant \alpha_1$	$\alpha_1 -$ rank
Y_j	$\alpha_2 + \alpha_3$	α_1	$\alpha_2 + \alpha_3 - \alpha_1$

$(\alpha_1 - \delta_{js})$ is contributed by nilpotent blocks in its Jordan form. This must be not less than $(q_j - 1)$, the rank of the largest of the Jordan blocks. Further, q_j cannot be less than the average size of the niJb (the ratio of total dimension of the nipotent blocks, namely $\alpha_2 + \alpha_3 - \delta_{js}$, to the total number of niJb, $\alpha_2 + \alpha_3 - \alpha_1$). Thus

$$\frac{\alpha_2 + \alpha_3 - \delta_{js}}{\alpha_2 + \alpha_3 - \alpha_1} \leqslant q_j \leqslant \alpha_1 + 1 - \delta_{js} \quad \text{(for a common spin } j\text{)}. \tag{5.8}$$

The inequality of the two extreme members in the above can be reexpressed as

$$(\alpha_1 - \delta_{js})(\alpha_2 + \alpha_3 - \alpha_1 - 1) \geqslant 0, \tag{5.9}$$

whence (5.6) follows.

If a special spin j were present, considering Table II and applying to X_j the same kind of arguments as above, one would get the condition

$$\frac{\alpha_1 - \delta_{js}}{\alpha_1 - \alpha_2} \leqslant p_j \leqslant (\alpha_2 + 1 - \delta_{js}) \quad \text{(for a special spin } j\text{)}, \tag{5.10}$$

whence

$$(\alpha_2 - \delta_{js})(\alpha_1 - \alpha_2 - 1) \geqslant 0. \tag{5.11}$$

If the first factor is nonzero, one immediately gets (5.7a); but if $\alpha_2 = 1$ and $j = s$, no further restrictions on $\alpha_1 - \alpha_2$ arise, thus admitting the additional possibility $\alpha_1 = 1$ (and hence $\alpha_3 = 1$ too) which does not fall under (5.7a). This is (5.7b).

Equations (5.6) and (5.7) tighten the restrictions (4.2); both of them must evidently be required in all cases where $S(\Lambda)$ contains both a common spin and a special spin.

The restrictions on $S(\Lambda)$ pertaining to various classes of linkage types, as embodied in Theorems 1 and 2, are displayed in Table IV.

6. PARITY-INVARIANT EQUATIONS

Only two of the various linkage types can support parity-invariant equations.

(a) *Type Ia:* Since this falls under case (a) of Theorem 1, the only $S(\Lambda)$ to be considered is $\alpha_1(1/2,1/2) \oplus \alpha_2[(1,0) \oplus (0,1)]$ and it involves only a single nontrivial spin, $j = 1$. The equality of α_2 and α_3 is necessary for parity invariance as the IR 2 and 3 are mutually conjugate here. It is also necessary that $R = \epsilon Q$, as we see from Sec. 2. Consequently, we have that $\text{Rank}(Q,R) = \text{Rank}(Q,\epsilon Q) = \text{Rank}(Q) = \alpha_2$. But (4.1b) requires this to be equal to α_1. Therefore, it is necessary that $\alpha_1 = \alpha_2$. Then Lemma 2 requires that Q be nonsingular, making X_1 (which reduces to $2g_1^2 Q\overline{Q}$) also nonsingular. This leads, in turn (by virtue of Lemma 4), to the requirement that $\alpha_1 = \alpha_2 = \alpha_3 = 1$. The final result is that there is just one possibility only for parity-invariant equations under the linkage type Ia, namely one which involves a single vector field and a single antisymmetric tensor field—in other words, the Kemmer[28] equation for spin 1.

(b) *Type II with $m = n$:* The class of representations consistent with Theorem 1 in this case is

$$S(\Lambda) \sim \alpha_1(\tfrac{1}{2}, \tfrac{1}{2}) \oplus \alpha_2(1,1) \oplus \alpha_3(0,0)$$

with α_i subject to both (5.6) and (5.7), i.e., either

$$\alpha_1 = \alpha_2 = \alpha_3 = 1 \quad \text{and} \quad s = 1 \tag{6.1a}$$

or

$$\alpha_2 + \alpha_3 > \alpha_1 > \alpha_2 \quad \text{with } s = 0 \text{ or } 1. \tag{6.1b}$$

The first of these alternatives corresponds to the Hagen–Singh equation[29] for spin 1. The second leads to a new class of equations which has not been studied or known hitherto. The simplest of these is obtained by choosing the smallest values for the α_i allowed by (6.1b), namely $\alpha_1 = 2$, $\alpha_2 = 1, \alpha_3 = 2$. (The wave function then consists of two vector fields, a symmetric traceless tensor field and two scalar fields.) With these values, we see from Tables II and III that the j-blocks for the special spin ($j = 1$) and the common ($j = 0$) have the following properties:

$\beta^0_{(1)}$ has dimension = 3, rank = 2, contains one niJb;

$\beta^0_{(0)}$ has dimension = 5, rank = 4, contains one niJb.

Suppose now that we want a spin 0 equation. Then $\beta^0_{(1)}$ has to be nilpotent, and should be equivalent to a single niJb of dimension 3, while $\beta^0_{(0)}$ would also contain a single niJb of dimension 3 (besides eigenvalues $+1$ and -1). Consequently, the minimal equation of β^0 would be $(\beta^0)^5 = (\beta^0)^3$. If, however, a spin 1 equation is sought, $\beta^0_{(0)}$ would be nilpotent—a single niJb of dimension 5—resulting in the minimal equation $(\beta^0)^7 = (\beta^0)^5$.

7. LIMITATION TO LOW VALUES OF l

We now examine what possibilites might exist for constructing new equations with fairly small values for the minimal degree—specifically, $l \leqslant 4$. Consideration will not be limited to parity-invariant cases, since such a restriction makes no material difference to the arguments, which are very simple. We only need to use the fact that with $l \leqslant 4$, all p_j and q_j must be $\leqslant 2$. The left-hand members of the inequalities (5.8) and (5.10) then require the following restrictions on the multiplicities.

(a) If a common spin j_c exists in the theory,

$$(\alpha_2 + \alpha_3 - \delta_{sj_c})/(\alpha_2 + \alpha_3 - \alpha_1) \leqslant q_{j_c} \leqslant 2, \tag{7.1}$$

which reduces to

$$2\alpha_1 - \delta_{js_c} \leqslant \alpha_2 + \alpha_3.$$

TABLE IV. Classes of linkage types, and restrictions [beyond (4.2)], on the IR in each.

Class	Linkage types	Restrictions on (m, n) and $\{\alpha_i\}$	
A: No special spin			
A_1: Single common spin	Ia	$n = \frac{1}{2}$	$\alpha_2 + \alpha_3 > \alpha_1$
	IV	$n = 0$	
A_2: Two common spins	IV	$n = \frac{1}{2}$	$\alpha_2 + \alpha_3 > \alpha_1$
B: One special spin j_s and one common spin j_c	Ib, II, III, VI	$n = \frac{1}{2}$	$\alpha_1 = \alpha_2 = \alpha_3 = 1$ with $s = j_s$ or $\alpha_2 + \alpha_3 > \alpha_1 > \alpha_2$
C: No common spin, two special spins	V	$n = \frac{1}{2}$	$\alpha_1 > \alpha_2, \alpha_3$

Recalling that we also must have $\alpha_2 < \alpha_1$ and $\alpha_3 < \alpha_1$, we conclude that the multiplicities must be such that *either*

$$\alpha_1 = \alpha_2 = \alpha_3 \quad (= \alpha, \text{say}) \tag{7.2a}$$

or, only when $j_c = s$,

$$\begin{aligned} \alpha_1 = \alpha \quad \text{with} \quad & \alpha_2 = \alpha, \quad \alpha_3 = \alpha - 1 \\ \text{or} \quad & \alpha_2 = \alpha - 1, \ \alpha_3 = \alpha. \end{aligned} \tag{7.2b}$$

In case there exist two common spins, only (7.2a) is admissible.

(b) If a special spin j_s exists in the theory, we have

$$(\alpha_1 - \delta_{sj_s})/(\alpha_1 - \alpha_2) \leqslant p_{j_s} \leqslant 2. \tag{7.3}$$

It is a straighforward matter to see what the above conditions imply for the various classes of linkage types. The results are summarized in column 2 of Table V. Column 3 displays the restrictions arising from parity invariance, found in Sec. 6.

8. DISCUSSION

We have shown in the foregoing sections that the IR $1 \equiv (m,n)$ to which 2 and 3 are linked in any of six ways must (for $m \geqslant n$) have $n = \frac{1}{2}$ (apart from a lone case with $n = 0$) in all acceptable theories. Limitations on the multiplicities and allowed spins have also been established. When space-inversion invariance is required in addition—this being the situation of greatest interest—we see from Table V that s can only be 0 or 1, nothing higher. This happens because of the IR 1 getting pinned down as $(\frac{1}{2}, \frac{1}{2})$ with 2 and 3 either (1,0) and (0,1) (linkage type Ia) or (1,1) and (0,0) (type II). The former yields none other than the Kemmer[28] equation for spin 1. The latter includes the Hagen–Singh[29] spin-1 equation (the case $\alpha_1 = \alpha_2 = \alpha_3 = 1$) and a more general class involving *repeated* IR in which the spin may be arranged to be 0 or 1 by proper choice of the skeleton matrix. The spin-0 equation recently proposed by Cox[30] does not fall under one of these admissible cases, the representation involved $[S(\Lambda) \sim 2(\frac{1}{2}, \frac{1}{2}) \oplus (1,1) \oplus (0,0)]$ being in violation of the requirement $\alpha_2 + \alpha_3 > \alpha_1$ (linkage II, Table V). It may be verified in fact that Cox's equation is barnacled. The simplest of the above-mentioned general class of type II is one with $\alpha_1 = \alpha_3 = 2$ and $\alpha_2 = 1$, already discussed in Sec. 6. It is a member of a subclass characterized by $\alpha_1 = \alpha_3 = \alpha$, $\alpha_2 = \alpha - 1$, i.e., having

$$S(\Lambda) \sim \alpha(\tfrac{1}{2}, \tfrac{1}{2}) \oplus (\alpha - 1)(1,1) \oplus \alpha(0,0),$$

in which the minimal degree $l + 2$ is $(2\alpha + 1)$ if s is chosen to be 0. Proof of this requires a detailed analysis (based on the properties given in Tables II and III) which will not be presented here. The analysis shows also that if s is taken to be 1, there are three possibilities in general for $l + 2$, viz., $(2\alpha + 3)$ or (for $\alpha \geqslant 3$) $(2\alpha + 1)$ or $(2\alpha - 1)$. The existence of several possibilities is a reflection of the fact that we have used only general conditions on the Jordan canonical form of the j-blocks following from (3.3) and (3.4), and have not tried to impose these conditions in full or to construct a skeleton matrix consistent with them. The process of carrying through this final stage will be illustrated in another paper, as already mentioned in the Introduction.

Possibilities for spins higher than 1 are found to come only in representations $S(\Lambda)$ which do not permit parity invariance or construction of a Lagrangian in conventional terms. Some of the allowed $S(\Lambda)$ have earlier made their appearance in the Hurley–Sudarshan[20] (H–S) analysis of theories which have $(\beta^0)^3 = \beta^0$ as the minimal equation for β^0. They are listed in Table VI. (Note that there are no repeated IR in these.) The insistence on the algebra of degree 3 in the H–S work has had the consequence that the equations corresponding to each of these cases have solutions with two different spins. On the other hand, our insistence on a unique spin causes the minimal equation of β^0 in all the cases to be of higher degree, $(\beta^0)^5 = (\beta^0)^3$, as may be seen with the aid of Tables II and III.

TABLE V. Further restrictions on requiring either $l \leqslant 4$ or parity invariance.

Class of linkage types	For $p \leqslant 4$	For parity
A_1	$\alpha_1 = \alpha_2 = \alpha_3 = \alpha$ or $\alpha_1 = \alpha' = \alpha, \alpha'' = \alpha - 1$	$\alpha_1 = \alpha_2 = \alpha_3 = 1, s = 1$
A_2	$\alpha_1 = \alpha_2 = \alpha_3 = \alpha$	not allowed
B	$\alpha_1 = \alpha_2 = \alpha_3 = 1, s = j_s$ or $\alpha_1 = \alpha_3 = 2, \alpha_2 = 1, s = j_c$	only linkage II allowed, with either $\alpha_1 = \alpha_2 = \alpha_3 = 1, s = 1$ or $\alpha_2 + \alpha_3 > \alpha_1 > \alpha_2, s = 0$ or 1
C	$\alpha_2 + \alpha_3 = \alpha_1$	not allowed

ACKNOWLEDGMENTS

P. M. M. takes pleasure in acknowledging the hospitality of Professor Irwin I. Shapiro and the Department of Physics, MIT, where part of this work was done. B. V. is grateful to the CSIR for a Senior Fellowship and subsequently a Research Associateship which enabled her to engage in this work.

APPENDIX

Here we list certain properties of matrices (especially nilpotent ones) which one rarely encounters, but are needed in our work.

Definition 1: A *nilpotent irreducible Jordan block* (niJb) is a square matrix having all elements zero *except* those (all equal to unity) in the positions immediately above the main diagonal. In general, an irreducible Jordan block is (iJb) the sum of a niJb and some multiple of the unit matrix.

Definition 2: If the minimal equation of a matrix M is

TABLE VI. Admissible $S(\Lambda)$ which have appeared in the Hurley–Sudarshan work.[20]

$S(\Lambda)$	Linkage type	Spin described	H–S classification no.[31]
$(s, 1) \oplus (s + 1/2, 1/2) \oplus (s + 1, 0)$	I	s	4
$(s, 1) \oplus (s - 1/2, 1/2) \oplus (s - 1, 0)$	II	$s \geqslant \frac{1}{2}$	2
$(s + 1, 1) \oplus (s + 1/2, 1/2) \oplus (s + 1, 0)$	III	s	2
$(s - 1, 1) \oplus (s - 1/2, 1/2) \oplus (s - 1, 0)$	VI	$s \geqslant 1$	3

$$M^q \underset{i}{\pi} (M - m_i I)^{q_i} = 0 \tag{A1}$$

with $m_i \neq 0$, then q is the *nilpotency index* of M.

1. The following properties of any niJb are self-evident or easily proven. Let J_d be an niJb of dimension d. Then:

(a) The rank of J_d is $(d-1)$;

(b) J_d^2 is equivalent to a direct sum of two niJb, both of which are of dimension $d/2$ if d is even, while one is of dimension $(d+1)/2$ and the other of dimension $(d-1)/2$ if d is odd (the latter is nonexistent for $d=1$).

2. From 1(a) and the fact that any square matrix can be similarly transformed to a direct sum of iJb, it follows that the number N of niJb in the Jordan canonical form of any matrix of dimension d and rank r is

$$N = d - r. \tag{A2}$$

3. If U_1 is an $\alpha \times \beta$ matrix of rank r_1 and U_2 a $\beta \times \sigma$ matrix of rank r_2, then for their product we have

$$\text{rank}(U_1 U_2) \leqslant \min(r_1, r_2). \tag{A3}$$

The only circumstance in which one can assert what the rank is is if

$$\text{rank}(U_1) = \text{rank}(U_2) = \beta, \tag{A4a}$$

in which case

$$\text{rank}(U_1 U_2) = \beta. \tag{A4b}$$

4. If J is an niJb of dimension n and $\tilde{v}$ an n-dimensional row vector, $\tilde{v}J^{n-1}$ is null if $\tilde{v}$ is linearly dependent on the rows of J and nonnull otherwise.

To see this, we note that there exists, associated with J, a column vector u and a row vector $\tilde{v}$ that the vectors

$$u_1 \equiv u, \quad u_2 \equiv Ju, \ldots, u_n \equiv J^{n-1} u \tag{A5}$$

form a linearly independent set of n column vectors and

$$\tilde{v}_n \equiv \tilde{v}, \quad \tilde{v}_{n-1} \equiv \tilde{v}J, \ldots, \tilde{v}_1 \equiv \tilde{v}J^{n-1} \tag{A6}$$

of row vectors, such that

$$\tilde{v}_i u_j = \delta_{ij}. \tag{A7}$$

Further, J can be analyzed in terms of these as

$$J = u_2\tilde{v}_1 + u_3\tilde{v}_2 + \cdots + u_n\tilde{v}_{n-1}. \tag{A8}$$

From these equations it follows readily that $J^{n-1} = u_n\tilde{v}_1$ and hence

$$\tilde{v}_i J^{n-1} = \delta_{ni}\tilde{v}_1. \tag{A9}$$

Since the rows of J are linear combinations of $\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_{n-1}$ only, according to (A8), and $\tilde{v}_n$ is linearly independent of these, Eq. (A9) is the result we seek.

5. Consider the Jordan canonical form of an arbitrary matrix M and the resolution of the space on which M acts into a direct sum of subspaces on which the individual iJb act. Associated with any nilpotent iJb (say J, of dimension n) of M, consider a row vector $\tilde{v}$ which has vanishing components in all subspaces other than the one associated with the particular niJb under consideration and is $\tilde{v}_n$ (as defined in the last paragraph) in the latter. Clearly, this vector is linearly independent of all the rows of M, and

$$\tilde{v}M^{n-1} = \tilde{v}_n J^{n-1} \neq 0 \tag{A10}$$

in view of (A9). There is one such vector for each niJb contained in M, and all these vectors are by themselves a linearly independent set too. Thus, associated with a matrix M, there is a set of linearly independent row vectors (equal in number to the number of niJb in M) which are linearly independent of the rows of M.

If the index of nilpotency of M is q, there is at least one niJb of dimension q; the $\tilde{v}$ associated with it (which is a member of the above-mentioned set) evidently possesses the property that

$$\tilde{v}M^{q-1} \neq 0. \tag{A11}$$

[1] M. Fierz and W. Pauli, Proc. R. Soc. (London) A **173**, 211 (1939).

[2] K. Johnson and E. C. G. Sudarshan, Ann. Phys. (N.Y.) **85**, 126 (1961).

[3] G. Velo and D. Zwanziger, Phys. Rev. **186**, 1337 (1969); **188**, 2218 (1969).

[4] A. S. Wightman, in *Invariant Wave Equations*, edited by G. Velo and A. S. Wightman (Springer-Verlag, Berlin, 1978).

[5] M. Seetharaman, T. R. Govindarajan, and P. M. Mathews, Pramana **13**, 173 (1979).

[6] P. M. Mathews, T. R. Govindarajan, M. Seetharaman, and J. Prabhakaran, J. Math. Phys. **21**, 1495 (1980).

[7] S. Deser and E. Witten, Nucl. Phys. B **178**, 491 (1981).

[8] M. J. Duff and P. van Nieuwenhuizen, Phys. Lett. B **94**, 179 (1980).

[9] A. Z. Capri. Prog. Theor. Phys. **48**, 1364 (1972).

[10] A. S. Glass, Commun. Math. Phys. **23**, 176 (1971).

[11] J. C. Fisk and W. Tait, J. Phys. A **6**, 383 (1973).

[12] W. Hurley and E. C. G. Sudarshan, J. Math. Phys. **16**, 2093 (1975).

[13] M. A. K. Khalil, Phys. Rev. D **15**, 1537 (1977).

[14] M. A. K. Khalil, Prog. Theor. Phys. **60**, 1559 (1978).

[15] W. Cox, J. Phys. A **11**, 1167 (1978).

[16] F. A. Berends, J. W. van Holten, P. Nieuwenhuizen, and B. deWitt, Nucl. Phys. B **154**, 261 (1979).

[17] P. M. Mathews, B. Vijayalakshmi, M. Seetharaman, and Y. Takahashi, J. Phys. A **14**, 1193 (1981).

[18] P. M. Mathews, B. Vijayalakshmi, and M. Sivakumar, J. Phys. A **15**, L579 (1982).

[19] H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945).

[20] W. J. Hurley and E. C. G. Sudarshan, Ann. Phys. (N.Y.) **85**, 546 (1974).

[21] W. Cox, J. Phys. A **7**, 2249 (1974).

[22] E. Sokatchev, Nucl. Phys. B **99**, 96 (1975).

[23] A. Z. Capri, Phys. Rev. **178**, 2427 (1969).

[24] Harish-Chandra, Phys. Rev. **71**, 793 (1947).

[25] T. R. Govindarajan, Ph. D. Thesis, University of Madras, 1979, unpublished.

[26] P. M. Mathews and T. R. Govindarajan, J. Phys. A **15**, 2093 (1982).

[27] It is to be remembered that this is with reference to $S(\Lambda)$ in which the IR $1 \equiv (m,n)$ has $m > n$. Not explicitly mentioned in the paper are the representations conjugate to such $S(\Lambda)$ which will, of course, have $m < n$. For acceptable theories based on the latter, the limitation will evidently be that $m = \frac{1}{2}$ (or 0 in the exceptional case).

[28] N. Kemmer, Proc. R. Soc. (London) A **173**, 91 (1939).

[29] C. R. Hagen and L. P. S. Singh, Phys. Rev. **8**, 3378 (1973).

[30] W. Cox, J. Phys. A **15**, 627 (1982).

[31] Hurley and Sudarshan treat our type IV as conjugate of type III and hence not needing separate consideration. However, this is justified only if $m > n$ in one and $m < n$ in the other. With $m > n$ in both types, their spin contents are different (see Table I). Types III and IV are therefore unrelated. A similar situation obtains for types V and VI too.

A vector wave equation for neutrinos

Frank Reifler
RCA, Government and Commercial Systems, Moorestown, New Jersey 08057

(Received 24 May 1983; accepted for publication 28 October 1983)

The Cartan map gives an isomorphism between spinors and isotropic vectors. Isotropic vectors $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ satisfy the condition $\mathbf{F}\cdot\mathbf{F} = 0$. We show that via the Cartan map, the particle current for neutrinos is given by $j^0 = |\mathbf{E}|$, $\mathbf{j} = \mathbf{E}\times\mathbf{H}/|E|$, and the neutrino wave equation becomes $D^0\mathbf{F} = i\mathbf{D}\times\mathbf{F} - (\mathbf{D}\mathbf{F})\cdot\mathbf{v}$, where $\mathbf{v} = \mathbf{j}/j^0 = \mathbf{E}\times\mathbf{H}/E^2$ = velocity field, $D^0 = i(h/2)(\partial/\partial t) - V^0$, $\mathbf{D} = -i(h/2)\nabla - \mathbf{V}$, where h = Planck's constant and $V = (V^0, \mathbf{V})$ = external potential. This wave equation preserves the isotropic condition, and like the equivalent Dirac equation, causes j to be the conserved current. We show that the isotropic restriction on the vector field $\mathbf{F}$ accounts for the observable properties of a neutrino in an external field, in particular, for the observed spectrum of the energy, momentum, angular momentum, spin, velocity, and position operators.

PACS numbers: 11.10.Qr, 03.50. – z

1. INTRODUCTION

Historically, spinors were introduced into physics for the following four reasons:

(A) To explain the observed spectrum of the energy, momentum, angular momentum, spin, position, and velocity of a spin-half particle in an external field.

(B) To obtain a first-order wave equation.

(C) To obtain a positive probability density which is conserved.

(D) To provide a relativistically covariant theory.

However, there are also two problems with assigning spinors to physical states:

(E) Spinors are tied to a specific Cartesian coordinate frame.[1] The spinor representation is *not coordinate free*, and consequently spinor fields exist only on a very restricted class of space-time manifolds.

(F) For the group of relativistic transformations to act properly, a spinor ψ and its negative $-\psi$ must refer to the same physical state.[2]

These problems can be resolved by a coordinate map from spinors to the isotropic vectors, introduced by Cartan.[3] Isotropic vectors $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ satisfy the condition $\mathbf{F}\cdot\mathbf{F} = 0$. The Cartan map is locally one-to-one onto a manifold of vector states that do not have difficulties (E) and (F).

Isotropic vectors are coordinate free; whereas spinors are tied to a specific Cartesian coordinate frame. To clarify this point consider the spinor

$$\xi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

whose spin is "up the z axis". To identify the z axis, one must specify a particular Cartesian coordinate frame.

Unlike spinors, isotropic vectors transform covariantly under arbitrary coordinate transformations. That spinors do not have any curvilinear components and hence are not generally covariant, is proved in the last theorem of Cartan's book on spinors.[4] This creates enormous difficulties in general relativity, as discussed in Cartan's book.

Generally, isotropic vector fields can be defined on any space-time manifold; whereas spinor fields exist only on a very restricted class of manifolds. Indeed, in order for spinors to be defined on a manifold, the manifold must admit a spin structure.[5] [A spin structure is a map that arbitrarily assigns to each SL(2,C) spinor frame a Cartesian coordinate frame, at every point of the manifold.] However, a spin structure exists only if the manifold is parallelizable, and general space-time manifolds are *not parallelizable*. Since no such restriction is needed for the isotropic vector representation, isotropic vectors have a wider range of applicability than do spinors.

For example, S^4 has been used as a space-time manifold for the Yang–Mills vector fields.[6] The isotropic vectors would allow spin-half particles to be represented as fields on S^4; whereas spinors cannot be used. The Cartan map is an interesting example of a nonlinear homomorphism which preserves physical structure between two representations of the relativity group. Because the Cartan map is not linear, physicists previously considered spinors and isotropic vectors to be physically inequivalent representations. With the Cartan map in mind, perhaps there are other nonlinear equivalences that could classify the irreducible projective representations of the relativity and other groups that model the elementary particles.[7]

The nonlinearity of the Cartan map is a consequence of the fact that $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ satisfy a nonlinear condition, $\mathbf{F}\cdot\mathbf{F} = 0$. It is highly probable that other properties such as isotropic spin, strangeness, etc., also impose nonlinear conditions on the vector fields. These new conditions (which may be similar to the isotropic condition) could be discovered by nonlinear coordinate maps (similar to the Cartan map) from the usual representations into vector fields that satisfy the new conditions.

For the isotropic vectors to be admissable as a model for spin-half particles, they must satisfy properties (A)–(D) above. In this paper we study the Cartan map for a neutrino in an external field.[8] We show that the isotropic condition of the vector field **F** accounts for all the observed properties of the neutrino with regard to (A)–(D).

The Cartan map is locally one-to-one from spinors onto the manifold of isotropic vectors. The theory of spinors and the theory of isotropic vectors are isomorphic theories. The isomorphism extends to all the properties associated with spinors. In particular (see the Appendix), the Cartan map commutes with Lorentz transformations, and preserves the eigenstates and eigenvalues of the energy, momentum, angular momentum, spin, position, and velocity operators [establishing (A) and (D)]. Thus a spinor field in a given state of energy, momentum, angular momentum, and spin is mapped into an isotropic vector field that represents the same physical state. If, for example, the spinor field is spin polarized up the z axis, so is the isotropic vector field.

We show in Secs. 3 and 4 that both theories have the same probability current density, and have equivalent wave equations. These are the main results that are derived in this paper [establishes (B) and (C)]. A straightforward generalization of these results to bispinors representing massive particles will be presented in a forthcoming paper.

An outline of this paper is as follows: In Sec. 2 we introduce the subject of isotropic vector fields and discuss some of their properties. A number of new useful vector identities for the Cartan map are proved in Sec. 3, and we establish (C). Section 4 is devoted to establishing (B) [(A) and (D) are proved in the Appendix]. A short summary and conclusion are given in Sec. 5.

2. ISOTROPIC VECTOR FIELDS

Let (**E**,**H**) be the real (three-dimensional) vector components of a complex vector **F** defined by

$$\mathbf{F} = \mathbf{E} + i\mathbf{H} \in C^3,$$

where $\mathbf{E},\mathbf{H} \in R^3$, C^3 = a complex three-dimensional vector space, and R^3 = a real three-dimensional vector space.

It is well known that if **E** and **H** are regarded as components of an electromagnetic field, then the Lorentz transformations acting on **F** are represented by complex orthogonal matrices.[9] Complex orthogonal matrices leave invariant the Euclidean quadratic form:

$$\begin{aligned}\mathbf{F}\cdot\mathbf{F} &= F_x^2 + F_y^2 + F_z^2 \\ &= E^2 - H^2 + 2i\mathbf{E}\cdot\mathbf{H}.\end{aligned} \tag{2.1}$$

Thus with respect to Lorentz transformations, $E^2 - H^2$ and $\mathbf{E}\cdot\mathbf{H}$ are invariants of **F**. When these invariants vanish, **F** is called an isotropic vector. That is, by definition a complex vector **F** is said to be isotropic if $F^2 = 0$.

The set of isotropic vectors **F** form a submanifold M of C^3,

$$M = \{\mathbf{F} \in C^3 | F^2 = 0\},$$

where C^n = an n-dimensional complex vector space. The Cartan map is a coordinate map of C^2 onto M. Note that C^2 is a flat space; whereas M has the same dimension as C^2, but is not flat.

Isotropic vector fields have special properties not shared by vector fields in general. One such property is that there exists a four-vector j defined by

$$j^0 = |E|,$$
$$\mathbf{j} = \mathbf{E}\times\mathbf{H}/|E|.$$

It is shown in Sec. 3 that via the Cartan map, the four-vector j is the particle current density four-vector for neutrinos.

In this paper we will show how the isotropic vector fields (**E**,**H**), with the four-vector j as the particle current density, give a vector model for neutrinos that is isomorphic to the usual spinor model. The main result of this paper, given in Sec. 4, is the derivation of the equivalent wave equation

$$D^0\mathbf{F} = i\mathbf{D}\times\mathbf{F} - (\mathbf{D}\mathbf{F})\cdot\mathbf{v}, \tag{2.2}$$

where

$$\mathbf{F} = \mathbf{E} + i\mathbf{H} = \text{isotropic vector field},$$
$$\mathbf{v} = \mathbf{j}/j^0 = \text{velocity field},$$
$$D^0 = i\frac{h}{2}\frac{\partial}{\partial t} - V^0,$$
$$\mathbf{D} = -i\frac{h}{2}\nabla - \mathbf{V},$$
$$h = \text{Planck's constant},$$
$$\partial/\partial t = \text{partial differentiation with respect to time},$$
$$\nabla = \text{gradient},$$
$$V = (V^0,\mathbf{V}) = \text{external potential}.$$

The wave equation (2.2) preserves the isotropic condition $F^2 = 0$, and also causes j to be the conserved current.

The reader may notice the similarity of the neutrino wave equation (2.2) with Maxwell's equation, which is given by

$$D^0\mathbf{F} = i\mathbf{D}\times\mathbf{F}. \tag{2.3}$$

However, in general, Maxwell's equation (2.3) does not preserve the isotropic condition $F^2 = 0$. Indeed, Eqs. (2.2) have the form that is necessary in order to preserve the isotropic condition $F^2 = 0$.

Isotropic vector fields **F** of a single frequency are circularly polarized about the axis **v**, where

$$\mathbf{v} = \mathbf{j}/j^0 = \mathbf{E}\times\mathbf{H}/E^2 = \text{velocity field}.$$

(The axis **v** will vary as a function of position.) The fields can be either right or left circularly polarized about **v**, which results in two possible spin states. Actually, by continuity only one spin state is possible for a given neutrino field. A neutrino field **F** (of a single frequency) is either right or left circularly polarized about the axis **v**.

For neutrino fields, angular momentum eigenvalues are integer multiples of $h/2$, where h is Planck's constant. However, the even multiples of $h/2$ do not give rise to fields **F** which satisfy the isotropic condition $F^2 = 0$. Thus the isotropic condition further restricts the possible angular momentum eigenvalues to be odd multiples of $h/2$. This fact correctly accounts for the observed angular momentum spectrum. As a rule, the eigenvalues of any operator may be restricted by considering only those fields that will satisfy the isotropic condition. This rule applies to the energy, momentum, angular momentum, spin, position, and velocity operators.

For example, the z-component of angular momentum J_z for spinors is given by[10]

$$J_z = ih\frac{\partial}{\partial\phi} + \frac{h}{2}\sigma_z\,,$$

where ϕ = spherical azimuth angle, σ_z = Pauli spin matrix for the z axis, and h = Planck's constant.

Via the Cartan map, J_z becomes

$$J_z = \frac{h}{2}\left\{i\frac{\partial}{\partial\phi} + S_z\right\},$$

where S_z = Proca spin matrix for the z axis.

We see that the eigenvalues of J_z are integral multiples of $h/2$, but only odd multiples of $h/2$ will give rise to eigenstates which satisfy the isotropic condition. A similar analysis can be made for each of the operators mentioned above, since they all commute with the Cartan map (see the Appendix).

3. ALGEBRAIC PROPERTIES OF THE CARTAN MAP

In this section we derive some useful algebraic properties of the Cartan map. Prior to defining the Cartan map, let us discuss some notation. Let us denote

$$C^n = \text{a complex } n\text{-dimensional vector space}\,.$$

Primarily, we will be interested in C^2, C^3, and C^4. We will denote elements of C^2 by Greek letters such as

$$\xi = \begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} \in C^2\,.$$

The elements of C^2 will be called spinors. The two components of a spinor $\xi \in C^2$ are denoted by ξ_1 and ξ_2.

We will denote elements of C^3 by Latin letters such as

$$\mathbf{F} = \begin{bmatrix}F_x\\ F_y\\ F_z\end{bmatrix} \in C^3\,.$$

The elements of C^3 will be called vectors. The three components of F along the x, y, and z axes of a Cartesian coordinate system are denoted by F_x, F_y, and F_z.

Finally, we will denote elements of C^4 by Latin letters such as

$$j = \begin{bmatrix}j^0\\ \mathbf{j}\end{bmatrix} = \begin{bmatrix}j^0\\ j_x\\ j_y\\ j_z\end{bmatrix}.$$

Elements of C^4 will be called four-vectors.

Definition: The Cartan map[11] is defined to be the bilinear map b from $C^2\times C^2$ into C^4 given by

$$b^0(\xi,\iota) = -(\xi_1\iota_2 - \xi_2\iota_1)\,,$$
$$\mathbf{b}(\xi,\iota) = \begin{bmatrix}\xi_1\iota_1 - \xi_2\iota_2\\ i(\xi_1\iota_1 + \xi_2\iota_2)\\ -(\xi_1\iota_2 + \xi_2\iota_1)\end{bmatrix}. \tag{3.1}$$

Note from Definition (3.1) that b^0 is antisymmetric, whereas $\mathbf{b}$ is symmetric in the variables ξ and ι. That is,

$$b^0(\xi,\iota) = -b^0(\iota,\xi)\,,$$
$$\mathbf{b}(\xi,\iota) = \mathbf{b}(\iota,\xi)\,. \tag{3.2}$$

In particular, for any spinor ξ,

$$b^0(\xi,\xi) = 0\,. \tag{3.3}$$

We now proceed to state certain algebraic properties of the bilinear map b in the lemmas below:

Lemma 1: For all spinors ρ, ξ, and ι which belong to C^2, the following identities are true:

$$\mathbf{b}(\rho,\xi)\cdot\mathbf{b}(\iota,\iota) = -2b^0(\rho,\iota)b^0(\xi,\iota)\,, \tag{3.4a}$$
$$\mathbf{b}(\rho,\xi)\cdot\mathbf{b}(\xi,\iota) = -b^0(\rho,\xi)b^0(\xi,\iota)\,, \tag{3.4b}$$
$$\mathbf{b}(\rho,\iota)\cdot\mathbf{b}(\xi,\iota) = b^0(\rho,\iota)b^0(\xi,\iota)\,, \tag{3.4c}$$
$$\mathbf{b}(\xi,\xi)\cdot\mathbf{b}(\iota,\iota) = -2b^0(\xi,\iota)^2\,, \tag{3.4d}$$
$$\mathbf{b}(\xi,\iota)\cdot\mathbf{b}(\xi,\iota) = b^0(\xi,\iota)^2\,, \tag{3.4e}$$
$$\mathbf{b}(\xi,\xi)\cdot\mathbf{b}(\xi,\iota) = 0\,. \tag{3.4f}$$

Lemma 2: For all spinors ξ and ι which belong to C^2, the following identity is true:

$$\mathbf{b}(\xi,\xi)\times\mathbf{b}(\iota,\iota) = 2ib^0(\xi,\iota)\mathbf{b}(\xi,\iota)\,. \tag{3.5}$$

Next we define the notion of a conjugate spinor.[12]

Definition: If

$$\xi = \begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} \in C^2$$

is any spinor, then a conjugate spinor $\xi\star$ associated with ξ is defined as

$$\xi\star = \begin{bmatrix}\bar{\xi}_2\\ -\bar{\xi}_1\end{bmatrix} \in C^2\,, \tag{3.6}$$

where $\bar{\xi}_1$ and $\bar{\xi}_2$ are complex conjugates of ξ_1 and ξ_2.

Lemma 3: For all spinors ξ and ι which belong to C^2, the following identities are true:

$$b^0(\xi,\iota\star) = \overline{b^0(\iota,\xi\star)}\,, \tag{3.7a}$$
$$\mathbf{b}(\xi,\iota\star) = \overline{\mathbf{b}(\iota,\xi\star)}\,, \tag{3.7b}$$
$$b^0(\xi\star,\iota\star) = \overline{b^0(\xi,\iota)}\,, \tag{3.7c}$$
$$\mathbf{b}(\xi\star,\iota\star) = -\overline{\mathbf{b}(\xi,\iota)}\,. \tag{3.7d}$$

Now let us define

$$j = b(\xi,\xi\star)\,,$$
$$\mathbf{F} = \mathbf{E} + i\mathbf{H} = i\mathbf{b}(\xi,\xi)\,, \tag{3.8}$$

where **E** and **H** are real vectors. We see from (3.7a) and (3.7b) that j is real. Moreover,

$$\mathbf{F}^\star = \mathbf{E} - i\mathbf{H} = \overline{i\mathbf{b}(\xi,\xi)} = i\mathbf{b}(\xi^\star,\xi^\star)\,. \tag{3.9}$$

Lemma 4: The following identities are true:

$$\mathbf{F}\cdot\mathbf{F} = 0\,, \tag{3.10a}$$

$$\mathbf{E}\cdot\mathbf{E} = \mathbf{H}\cdot\mathbf{H}\,, \tag{3.10b}$$

$$\mathbf{E}\cdot\mathbf{H} = 0\,, \tag{3.10c}$$

$$j^0 = |E|\,, \tag{3.10d}$$

$$\mathbf{j} = \mathbf{E}\times\mathbf{H}/|E|\,. \tag{3.10e}$$

Proof: From (3.4f) and (3.8) we have

$$\mathbf{F}\cdot\mathbf{F} = -\mathbf{b}(\xi,\xi)\cdot\mathbf{b}(\xi,\xi) = 0\,,$$

which proves (3.10a), (3.10b) and (3.10c) are simple consequences of (3.10a). From (3.4d) and (3.8) and (3.9) we get

$$j^0 = \left(\frac{\mathbf{F}\cdot\mathbf{F}^\star}{2}\right)^{1/2} = |E|\,,$$

which proves (3.10d). Lastly, from (3.5) we have, using (3.8) and (3.9),

$$\mathbf{j} = i\,\frac{\mathbf{F}\times\mathbf{F}^\star}{2j^0} = \frac{\mathbf{E}\times\mathbf{H}}{|E|}\,,$$

which proves (3.10e). ∎

Lemma 5: For all spinors ξ which belong to C^2, the following identities are true:

$$j^0 = |E| = |\xi|^2\,, \qquad \mathbf{j} = \frac{\mathbf{E}\times\mathbf{H}}{|E|} = \bar{\xi}^T\boldsymbol{\sigma}\xi\,, \tag{3.11}$$

where $\bar{\xi}^T$ = transpose conjugate of ξ and $\boldsymbol{\sigma} = (\sigma_x,\sigma_y,\sigma_z)$ = Pauli spin-$\frac{1}{2}$ matrices.

It is a consequence of (3.11) that j transforms as a four-vector under relativistic transformations.[13] Also, the vectors **E** and **H** transform as components of an electromagnetic field tensor under relativistic transformations.[14] Furthermore, the Cartan map which associates (**E**,**H**) with ξ maps spin eigenstates to the same spin eigenstates, as can be deduced from Lemma 6(d) below.

Lemma 6: For every pair of spinors ξ and ι belonging to C^2, and any vector **v**, the following identities are true:

(a) $b^0(\mathbf{v}\cdot\boldsymbol{\sigma}\xi,\iota) = \mathbf{v}\cdot\mathbf{b}(\xi,\iota)$,

(b) $\mathbf{b}(\mathbf{v}\cdot\boldsymbol{\sigma}\xi,\iota) = \mathbf{v}b^0(\xi,\iota) + i\mathbf{v}\times\mathbf{b}(\xi,\iota)$,

(c) $\mathbf{b}(\mathbf{v}\cdot\boldsymbol{\sigma}\xi,\xi) = (\mathbf{v}\cdot\mathbf{S})\mathbf{b}(\xi,\xi)$,

(d) If ξ is an eigenspinor of $\mathbf{v}\cdot\boldsymbol{\sigma}$ with eigenvalue λ, then $\mathbf{b}(\xi,\xi)$ is an eigenvector of $\mathbf{v}\cdot\mathbf{S}$ with eigenvalue λ, where $\mathbf{S} = (S_x,S_y,S_z)$ = Proca spin-one matrices.[15]

4. THE WAVE EQUATION

Let $V = (V^0,\mathbf{V})$ be the four-vector potential of an external electromagnetic field. We define momentum operators $p = (p^0,\mathbf{p})$ as follows:

$$p^0 = ih\frac{\partial}{\partial t} - V^0\,,$$

$$\mathbf{p} = -ih\nabla - \mathbf{V}\,,$$

where h = Planck's constant and ∇ = gradient (differential operator).

Dirac's equation for neutrinos is given by[16]

$$p^0\xi = \mathbf{p}\cdot\boldsymbol{\sigma}\xi \tag{4.1}$$

for spinor fields ξ.

Let us define new momentum operators $D = (D^0,\mathbf{D})$ as follows:

$$D^0 = i\frac{h}{2}\frac{\partial}{\partial t} - V^0\,,$$

$$\mathbf{D} = -i\frac{h}{2}\nabla - \mathbf{V}\,.$$

Note that the momentum operators D are similar to the momentum operators p, except for the constants h and $h/2$.

Moreover, when differentiating the bilinear Cartan map b, let p' = the momentum operators p acting only on the first argument of b, and p'' = the momentum operators p acting only on the second argument of b. Since $\mathbf{b}(\xi,\iota) = \mathbf{b}(\iota,\xi)$ is symmetric, we have

$$\begin{aligned} p'\mathbf{b}(\xi,\xi) &= \mathbf{b}(p\xi,\xi) = \mathbf{b}(\xi,p\xi) \\ &= p''\mathbf{b}(\xi,\xi) \\ &= D\mathbf{b}(\xi,\xi)\,. \end{aligned} \tag{4.2}$$

However, $b^0(\xi,\iota) = -b^0(\iota,\xi)$ is antisymmetric, which gives a different result.

Lemma 7: For all spinors ξ which belong to C^2,

$$b^0(p\xi,\xi) = -\{D\mathbf{b}(\xi,\xi)\}\cdot\mathbf{v}\,, \qquad \mathbf{b}(p\xi,\xi) = D\mathbf{b}(\xi,\xi)\,, \tag{4.3}$$

where $\mathbf{v} = \mathbf{j}/j^0$.

Proof: Set $\rho = p\xi$ and $\iota = \xi^\star$. Then by Eq. (3.4b) of Lemma 1, we get, since $j = b(\xi,\xi^\star)$,

$$\begin{aligned} b^0(p\xi,\xi) &= b^0(\rho,\xi) \\ &= -\mathbf{b}(\rho,\xi)\cdot\{\mathbf{b}(\xi,\xi^\star)/b^0(\xi,\xi^\star)\} \\ &= -\{p'\mathbf{b}(\xi,\xi)\}\cdot\mathbf{v} \\ &= -\{D\mathbf{b}(\xi,\xi)\}\cdot\mathbf{v}\,, \end{aligned}$$

where (4.2) is used in the last step. ∎

We now state and prove the main theorem:

Theorem: Via the Cartan map, Dirac's equation for neutrinos becomes the following wave equation:

$$D^0\mathbf{F} = i\mathbf{D}\times\mathbf{F} - (\mathbf{D}\mathbf{F})\cdot\mathbf{v}\,.$$

Proof: Recall from Lemma 6(b) of Sec. 3 that

$$\mathbf{b}(\mathbf{p}\cdot\boldsymbol{\sigma}\xi,\xi) = \mathbf{p}'b^0(\xi,\xi) + i\mathbf{p}'\times\mathbf{b}(\xi,\xi)\,.$$

Using (4.3) this equation becomes

$$\mathbf{b}(\mathbf{p}\cdot\boldsymbol{\sigma}\xi,\xi) = -\{\mathbf{D}\mathbf{b}(\xi,\xi)\}\cdot\mathbf{v} + i\mathbf{D}\times\mathbf{b}(\xi,\xi)\,. \tag{4.4}$$

Together with Dirac's equation (4.1) this last equation (4.4) implies that

$$\begin{aligned} D^0\mathbf{b}(\xi,\xi) &= \mathbf{b}(p^0\xi,\xi) \\ &= \mathbf{b}(\mathbf{p}\cdot\boldsymbol{\sigma}\xi,\xi) \\ &= i\mathbf{D}\times\mathbf{b}(\xi,\xi) - \{\mathbf{D}\mathbf{b}(\xi,\xi)\}\cdot\mathbf{v}\,. \end{aligned} \tag{4.5}$$

Substituting $\mathbf{F} = i\mathbf{b}(\xi,\xi)$ into (4.5), we get the result

$$D^0\mathbf{F} = i\mathbf{D}\times\mathbf{F} - (\mathbf{D}\mathbf{F})\cdot\mathbf{v}\,.$$ ∎

5. SUMMARY AND CONCLUSION

Isotropic vector fields satisfy the properties (A-D) of Sec. 1, which make them an admissible model for spin-half particles. Moreover, isotropic vector fields do not have the difficulties (E) and (F) of the spinor fields, for assigning them physical states.

The two theories are isomorphic. We showed that both have the same probability current density, and satisfy equivalent wave equations (Secs. 3 and 4). We established the isomorphism using the nonlinear Cartan map, which has some novel algebraic properties, which are derived in Sec. 3.

We conclude from these results that spin-half particles have a coordinate-free representation as vector fields $\mathbf{F} = \mathbf{E} + i\mathbf{H}$ satisfying the isotropic condition $\mathbf{F}\cdot\mathbf{F} = 0$. It is conceivable that all other properties of spin-half particles, such as isotropic spin, strangeness, etc., may also be represented by additional conditions on the vector fields.

APPENDIX: COMMUTATION OF THE CARTAN MAPS WITH THE OPERATORS REPRESENTING OBSERVABLES

Let ξ be a spinor field, and let $\underline{\tilde{A}}$ be an operator acting on ξ. Let the Cartan map $\mathbf{b}$ send ξ to its isotropic vector field $\mathbf{F} = i\mathbf{b}(\xi,\xi)$. We will say that $\underline{\tilde{A}}$ commutes with the Cartan map and becomes the operator $\underline{A}$, acting on $\mathbf{F}$, if

$$\begin{aligned}\underline{A}\,\mathbf{F} &= i\underline{A}\mathbf{b}(\xi,\xi)\\ &= i\mathbf{b}(\underline{\tilde{A}}\xi,\xi)\,.\end{aligned} \tag{A1}$$

Thus if ξ is an eigenstate of $\underline{\tilde{A}}$ with eigenvalue λ, we have from (A1) since $\mathbf{b}$ is bilinear,

$$\underline{A}\,\mathbf{F} = ib(\lambda\xi,\xi) = \lambda\,\mathbf{F}\,,$$

so then $\mathbf{F}$ is an eigenstate of $\underline{A}$ with the same eigenvalue λ.

Using Lemmas 6 and 7 from Secs. 3 and 4, the reader can easily establish that the operators listed in Table I commute with the Cartan map. Generally, any linear combination of these operators, with coefficients that are functions of $\mathbf{x}$, will commute with the Cartan map.

The following theorem gives two simple consequences of the commutation with the operators above:

Theorem: (a) The Cartan map preserves the eigenstates and eigenvalues of these operators.

(b) The Cartan map commutes with relativistic transformations.

Proof: The proof of (a) has already been established by Formula (A1). To prove (b), note that the operators representing the infinitesimal generators of the relativity group, for example $i\mathbf{J}$, commute with the Cartan map. Let $\underline{U}(\tau)$ be a one-parameter subgroup of relativistic transformations. Let

$$2\underline{A} = \underline{U}'(\tau)\underline{U}(\tau)^{-1}$$

be its infinitesimal generator. Then one shows that

$$\underline{U}(\tau)\mathbf{F} = i\mathbf{b}(\underline{\tilde{U}}(\tau)\xi,\underline{\tilde{U}}(\tau)\xi) \tag{A2}$$

by considering the unique solution of the differential equation

$$\frac{d\mathbf{F}}{d\tau}(\tau) = 2\underline{A}\,\mathbf{F}(\tau) \tag{A3}$$

TABLE I. Operators commuting with the Cartan map.

Operator symbol	Spinor representation	Isotropic vector representation
ϵ	$ih\dfrac{\partial}{\partial t}$	$i\dfrac{h}{2}\dfrac{\partial}{\partial t}$
$\mathbf{p}$	$-ih\nabla$	$-i\dfrac{h}{2}\nabla$
$\mathbf{x}$	$\mathbf{x}$	$\mathbf{x}$
$V(\mathbf{x})$	$V(\mathbf{x})$	$V(\mathbf{x})$
$\mathbf{s}$	$\frac{1}{2}h\boldsymbol{\sigma}$	$\frac{1}{2}h\mathbf{S}$
$\mathbf{L}$	$\mathbf{x}\times\mathbf{p}$	$\mathbf{x}\times\mathbf{p}$
$\mathbf{J}$	$\mathbf{L}+\mathbf{s}$	$\mathbf{L}+\mathbf{s}$
$\mathbf{v}$	$\boldsymbol{\sigma}$	$\mathbf{S}$

with $\mathbf{F}(0) = \mathbf{F}$. Clearly, the left-hand side of (A2) is the unique solution of (A3). Set

$$\mathbf{F}(\tau) = i\mathbf{b}(\underline{\tilde{U}}(\tau)\xi,\underline{\tilde{U}}(\tau)\xi)\,.$$

Then, $\mathbf{F}(0) = \mathbf{F}$, and also,

$$\begin{aligned}\frac{d\mathbf{F}}{d\tau} &= 2i\mathbf{b}(\underline{\tilde{U}}'\xi,\underline{\tilde{U}}\xi)\\ &= 2i\mathbf{b}(\underline{\tilde{A}}\underline{\tilde{U}}\xi,\underline{\tilde{U}}\xi)\\ &= 2i\underline{A}\tilde{b}(\underline{\tilde{U}}\xi,\underline{\tilde{U}}\xi)\\ &= 2\underline{A}\,\mathbf{F}\,.\end{aligned}$$

Therefore the right-hand side of (A2) is also the unique solution of (A3). Thus Formula (A2) is proved. Formula (A2) establishes that the Cartan map commutes with relativistic transformations.

[1] D. Hestenes, *Space-Time Algebra* (Gordon and Breach, New York, 1966).

[2] By a spinor field Ψ, one really means a projective spinor field Ψ for which $\Psi = -\Psi$. D. J. Simms, *Lie Groups and Quantum Mechanics* (Springer-Verlag, New York, 1968).

[3] E. Cartan, *Theory of Spinors* (MIT, Cambridge, MA, 1966). This map will be referred to as the Cartan map.

[4] E. Cartan, *Theory of Spinors* (MIT, Cambridge, MA, 1966).

[5] R. Geroch, "Spinor structure of space-times in general relativity. I," J. Math. Phys. **9** (11), 1739 (1968).

[6] M. Daniel and C. M. Viallet, "The Geometrical Setting of Gauge Theory of the Yang–Mills Type," Rev. Mod. Phys. **52** (1), 175 (1980).

[7] D. J. Simms, *Lie Groups and Quantum Mechanics* (Springer-Verlag, New York, 1968).

[8] A neutrino is a massless spin-half particle, described by Dirac's equation. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory Part I* (Pergamon, New York, 1979).

[9] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, New York, 1971).

[10] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory Part I* (Pergamon, New York, 1979).

[11] E. Cartan, *Theory of Spinors* (MIT, Cambridge, MA, 1966).

[12] A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Dover, New York, 1980).

[13] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory Part I* (Pergamon, New York, 1979).

[14] E. Cartan, *Theory of Spinors* (MIT, Cambridge, MA, 1966).

[15] The ith Proca spin-one matrix S_i has components $i(\epsilon_i)_{jk}$, where ϵ_{ijk} is the permutation symbol.

[16] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory Part I* (Pergamon, New York, 1979).

Explicit evaluation of group-invariant measure as by-product of path integration over Yang–Mills fields

Claudio Teitelboim
Center for Theoretical Physics, The University of Texas at Austin, Austin, Texas 78712
and Istituto di Fisica Teorica dell' Università di Torino, 10125, Torino, Italy

(Received 8 December 1982; accepted for publication 15 April 1983)

A method is presented for explicitly evaluating the invariant measure over a Lie group, for any parametrization in which a generic group element is written as a product of exponentials of elements of the Lie algebra. The measure is expressed as a transition amplitude associated with the quantum mechanical evolution operator of the "ghost fields" which appear in the path integral over Yang–Mills fields. As an illustration of the method, the measure is evaluated in two cases: (i) in canonical coordinates, for an arbitrary Lie group which admits them; (ii) in terms of Euler angles, for the group of rotations in three-dimensional space.

PACS numbers: 11.15. – q, 03.65. – w, 02.20.Sv

I. INTRODUCTION

The path integral approach is without doubt the most appropriate formulation available of the quantum mechanics of the Yang–Mills field.[1] At the heart of its success is the fact that the path integral is particularly suited to take into account the geometric properties of the field, which are reflected in the gauge invariance of the action.

As shown by Faddeev and Popov,[1] the manner in which the gauge invariance is dealt with in the path integral is based on the principles of the theory of invariant integration over groups. However, the explicit form of the invariant group measure plays a role neither in establishing the path integral scheme nor in its usual applications. It is sufficient for those purposes to know that the measure exists.

It is the point of this paper to indicate that one can turn the procedure upside down and use the path integral to obtain explicit formulas for the invariant measure in different parametrizations of the group. We shall describe the general method for doing this and will apply it to two examples:

(i) the measure in canonical coordinates for an arbitrary Lie group which admits them;

(ii) the measure for the rotation group in three dimensions in terms of Euler angles.

Strictly speaking, nothing new will be obtained since the above-mentioned measures have been found long ago by other techniques. However, we feel that the method described here is of interest since, besides its practical utility, it provides an interplay of concepts and techniques of group theory, gauge fields, functional integration and Hilbert space—all of them coming together under the common roof of quantum mechanics.

The presentation is organized as follows: Section II reviews the Hamiltonian action principle for the Yang–Mills field, including its gauge invariance. Section III presents those points of the path integral procedure which are essential for our aim. The key issue of boundary conditions and admissible ways of fixing the gauge is dealt with in Sec. IV. Section V reviews another needed concept, the representation of determinants as functional integrals (or transition amplitudes). Finally, Secs. VI and VII present the manner in which the group measure is written as a "ghost" transition amplitude as well as its explicit evaluation for the two examples indicated above.

The method presented here arose from work on the Hamiltonian path integral quantization of the gravitational field.[2]

II. HAMILTONIAN ACTION FOR THE YANG–MILLS FIELD

In order to establish both the notation and some properties that will be needed below, we shall briefly review here the properties of the action integral for the Yang–Mills field.

The components of the field will be denoted by $A^a{}_\mu(x,t)$. The space-time index μ takes the values $0,i$, with $i = 1,2,3$, and the group index a runs from 1 to n, with n being the number of generators of the Lie group on which the theory is based.

The action takes the form,

$$S = \int_{t_1}^{t_2} dt \left[\int d^3x \,(\dot{A}^a{}_i\, \pi_a{}^i - A_0{}^a \gamma_a) - H \right]. \tag{2.1}$$

Here the $\pi_{a\,i}$ are the momenta canonically conjugate to the spatial components $A^a{}_i$. No conjugate to the time component $A_0{}^a$ appears; that variable enters as a Lagrange multiplier for the constraint,

$$\gamma_a = \pi_a{}^i{}_{,i} + C_{ab}{}^c A^b{}_i \pi_c{}^i. \tag{2.2}$$

The function γ_a is called a constraint because it vanishes on classical trajectories as a consequence of the demand that the action be an extremum under variations of $A_0{}^a$. The numbers $C_{ab}{}^c$ in (2.2) are the structure constants of the Lie group.

The γ_a's obey the Poisson bracket relations,

$$[\gamma_a(x),\gamma_b(x')] = C_{ab}{}^c \gamma_c(x)\, \delta(x, x'), \tag{2.3}$$

and, for the Yang–Mills field, the functional H is given by

$$H = \int \frac{1}{2}(\pi^2 + B^2)\, d^3x, \tag{2.4}$$

with $\pi^2 = \pi_a{}^i \pi^a{}_i$, $B^2 = B^a{}_i B_a{}^i$. Here B is the "gauge covariant" curl of $A^a{}_i$ and the group indices are raised and lowered with the group metric and its inverse [which are hence needed to write (2.4)].

However, in the present case we shall only need from H the property

$$[H, \gamma_a(x)] = 0, \tag{2.5}$$

which (2.4) satisfies together with many other functions [all powers of (2.4), for example].

The action (2.1) is invariant under the infinitesimal gauge transformation

$$\delta A^a{}_i = \left[A^a{}_i, \int \epsilon^b \gamma_b\, d^3x\right], \tag{2.6a}$$

$$\delta \pi_a{}^i = \left[\pi_a{}^i, \int \epsilon^b \gamma_b\, d^3x\right], \tag{2.6b}$$

$$\delta A_0{}^a = \dot\epsilon^a - C_{bc}{}^a A_0{}^b \epsilon^c, \tag{2.6c}$$

where the ϵ^a are arbitrary functions of x and t.

Equations (2.6a) and (2.6b) show that the constraints γ_a are the generators of gauge transformations on the canonical variables, an observation that combined with (2.3), implies that the gauge group of the theory is the product of infinitely many copies (one for each space-time point) of the Lie group with structure constants $C_{ab}{}^c$.

The action principle for the Yang–Mills field states that the classical equations of motion are obtained by demanding that (2.1) should not change under variations of $A^a{}_i$, $\pi_a{}^i$ and $A_0{}^a$. The only restriction on the variations being that $A^a{}_i$ should be fixed up to a gauge transformation of the form (2.6a) at the endpoints t_1, t_2. In other words, it is the gauge-invariant content of $A^a{}_i$, rather than $A^a{}_i$ itself, which is fixed at the end points.

III. PATH INTEGRAL OVER GAUGE FIELDS

According to Faddeev and Popov,[1] the transition amplitude between two field configurations is given by a path integral of the form

$$\int \prod_{x,t,i,a} dA^a{}_i(x,t)\, d\pi_a{}^i(x,t)\, dA_0{}^a(x,t) \times \exp(iS)\, \delta[\chi] \det M_\chi. \tag{3.1}$$

In (3.1) the "gauge condition" χ is a certain functional of $A^a{}_\mu$ and $\pi_a{}^i$ which depends on x, t. The effect of $\delta[\chi] = \Pi_{x,t}\, \delta(\chi(x,t))$ is to select in the path integral only the representative with $\chi = 0$ from each class of gauge equivalent histories [which are related to each other by iterations of the infinitesimal transformation (2.6)].

The linear operator M_χ describes the response of the gauge condition to a change in gauge. If we denote by $\delta_\epsilon \chi$ the effect on χ of having its arguments undergo the transformation (2.6), then M_χ is defined by,

$$M_\chi\, \epsilon = \delta_\epsilon\, \chi. \tag{3.2}$$

The presence of the factor $\det M_\chi$ in the integration measure in (3.1) ensures that the value of the integral is independent of the choice of χ or, in other words, converts (3.1) into an integral over classes of gauge-equivalent histories.

IV. BOUNDARY CONDITIONS AND ADMISSIBLE GAUGE CONDITIONS

The quantities which are fixed in the classical action principle become the arguments of the quantum mechanical transition amplitude.

In the present case those arguments are the gauge-invariant content A_1, A_2 of $A^a{}_i(x, t_1)$ and $A^a{}_i(x, t_2)$. Here, A_1 and A_2 are classes of equivalence of gauge-related field configurations at the initial and final times. (These classes are the boundaries of the classes of higher dimensionality over which the path integral is performed.)

Now, in practice one wants to fix $A^a{}_i(x, t_1)$ and $A^a{}_i(x, t_2)$ rather than the more abstract notion of a class. In order to do this while still effectively dealing with the gauge invariant contents A_1, A_2, *one must leave the gauge freedom at the end points untouched by the gauge condition.*

Thus, the gauge transformations that can (and must) be eliminated by the gauge condition χ are not all those which leave invariant the action (2.1), but, rather, only the ones which do not touch the endpoints. In other words, we must add to the transformations (2.6) the boundary condition,

$$\epsilon^a(t_1, x) = \epsilon^a(t_2, x) = 0, \quad \text{for all } x. \tag{4.1}$$

The restrictions (4.1) limit the class of permissible gauge conditions. Indeed, χ must be such that, starting from any given history, one can reach, by means of a gauge transformation, one and only one history satisfying $\chi = 0$. Here the gauge transformation employed must be an iteration of the transformations (2.6), with the boundary conditions (4.1).

A necessary condition for χ to be admissible in the above sense is that the operator M_χ defined by (3.2) should be invertible under the boundary conditions (4.1). This ensures that $\det M_\chi \neq 0$, as it should be the case for the measure in (3.1) not to vanish identically. The invertibility of M_χ means that if a history with $\chi = 0$ exists, it is accessible from any history in a neighborhood and also unique within a neighborhood. Nothing is said about the global problem which is much harder to deal with.

If M_χ acting on the space (4.1) is to have an inverse, it must be a second-order differential operator in time. Thus we should look for gauge conditions which are such that $\delta_\epsilon\, \chi$ should involve second time derivatives of ϵ. This means that, in the present context, canonical conditions such as $\chi = \partial^i A^a{}_i$ for which $\delta_\epsilon\, \chi$ involves no time derivatives of ϵ [Eq. (2.6a)] are not permissible. Neither is it allowed the use of $\chi = A_0{}^a$, for which $\delta_\epsilon\, \chi$ involves only first time derivatives of ϵ [Eq. (2.6c)]. [These conditions can, however, be used if one folds the amplitude (3.1) with gauge-invariant wave functionals, a step not considered here.]

A wide class of permissible gauge conditions are those of the form,

$$\chi^a(x, t) = A_0{}^a(x, t) + f^a(x, t)[A^a{}_i, \pi_a{}^i, A_0{}^a], \tag{4.2}$$

where f contains no time derivatives of $A^a{}_i$, $\pi_a{}^i$, $A_0{}^a$, but may depend on them even nonlocally in space. The point here is that the variation of (4.2) yields an $\ddot\epsilon^a$ on account of (2.6c).

If a permissible gauge condition is used, then the path integral (3.1) must be carried over all histories $A_a{}^i(x, t)$, $\pi_a{}^i(x, t)$, $A_0{}^a(x, t)$ for which $A^a{}_i(x, t)$ is fixed at the end points t_1, t_2:

$$A^a{}_i(x, t_1) = A^a{}_i(1), \quad A^a{}_i(x, t_2) = A^2{}_i(2). \tag{4.3}$$

In(4.3) one may choose for $A^a{}_i(1)$ any representative within the class A_1 of the gauge related fields. Similarly for $A^a{}_i(2)$. The amplitude (3.1) should then become automatically dependent only on A_1, A_2, that is it should not change its value if either $A^a{}_i(1)$ or $A^a{}_i(2)$ undergo a gauge transformation (2.6a). This observation will be crucial for the identification of the group invariant measure.

V. REPRESENTATION OF DETERMINANTS AS TRANSITION AMPLITUDES

In a finite number of dimensions one has the formula for a Gaussian integral,

$$\int \exp[iv_i\, a^i{}_k\, u^k]\, dv\, du = (\det a)^{-1}, \tag{5.1}$$

where we have left out an irrelevant numerical factor.

The formula (5.1) may be used in turn to represent the determinant of a linear operator acting on an infinite-dimensional function space by means of a functional integral. The functional integral may then be defined as the matrix element of a quantum mechanical evolution operator in Hilbert space.

This way of evaluating determinants appears to be much more efficient for our purposes than, for example, using directly the Fredholm formula $\det M = \exp(\operatorname{tr} \ln M)$. It also has the advantage, in the examples at hand, of providing an answer already regularized, whereas in the Fredholm case one has to first divide M by a regularizing operator M_0 (that is one must replace M by $M_0^{-1}M$).

Thus for the operator M_χ appearing in the path integral we write the analog of the argument of the exponential in (5.1) in the form

$$S_\chi = \int_{t_1 \atop \text{all } x}^{t_2} \bar{\epsilon}_a(x, t)(M_\chi\, \epsilon)^a(x, t)\, dx\, dt, \tag{5.2}$$

where the gauge parameter $\epsilon^a(x, t)$ takes the place of u^k in (5.1) whereas the independent set of fields $\bar{\epsilon}_a(x, t)$ plays the role of v_i.

The determinant of M_χ has then the representation

$$(\det M_\chi)^{-1} = \int \prod_{x, t, a} d\bar{\epsilon}_a(x, t)\, d\epsilon^a(x, t) \times \exp(iS_\chi), \tag{5.3}$$

with S_χ given by (5.2).

[In field theory applications $\bar{\epsilon}$ and ϵ are taken to be anticommuting "ghost" fields and the integration in (5.3) is understood to be that introduced by Berezin.[1] This has the effect of giving $(\det M_\chi)^{+1}$ instead of $(\det M_\chi)^{-1}$ on the left side of (5.3). For the present purpose we have preferred to use ordinary real-valued fields in order not to inject another concept in the discussion. All the analysis that follows can be carried out equally well if ϵ and $\bar{\epsilon}$ are anticommuting.]

In (5.3) the functional integral is extended over all fields ϵ^a satisfying the boundary conditions (4.1), because that is the space on which M_χ acts. The boundary conditions on $\bar{\epsilon}_a$, on the other hand, must be determined by a different argument, since those functions do not play a direct role in the gauge transformations.

In our case, as indicated before, we ultimately want to define the functional integral as the matrix element of an evolution operator in Hilbert space. In order for this to be feasible, one must be able to write the integrand under the time integral in (5.2) as a Lagrangian which depends on time derivatives of up to the first order only. This is necessary[3] in order for the amplitudes to obey the basic composition law for processes occurring in succession, which we may write symbolically as

$$\langle 3|1\rangle = \sum_2 \langle 3|2\rangle\, \langle 2|1\rangle. \tag{5.4}$$

Now, as explained earlier, the operator M_χ contains second time derivatives, so these must be eliminated from the action (5.2). When the gauge condition is of the form (4.2), this is easily done, provided one imposes the boundary conditions

$$\bar{\epsilon}_a(x, t_1) = \bar{\epsilon}_a(x, t_2) = 0, \tag{5.5}$$

since in that case the second time derivatives disappear upon integration by parts without bringing in an endpoint contribution. Hence, we shall adopt (5.5) for gauge conditions of the form (4.2).

The functional integral (5.3) is then to be evaluated under the boundary conditions (4.1) and (5.5) and its value may therefore be written as a matrix element,

$$(\det M_\chi)^{-1} = \langle \epsilon^a = \bar{\epsilon}_a = 0 | U_\chi(t_2, t_1) | \epsilon^a = \bar{\epsilon}_a = 0 \rangle, \tag{5.6}$$

where U_χ is the evolution operator obtained from the Hamiltonian associated with the action (5.2).

VI. INVARIANT MEASURE IN CANONICAL GROUP COORDINATES

As a first illustration and application of the method we will derive the expression for the group-invariant measure of an arbitrary Lie group which admits canonical coordinates.

The canonical coordinates are such that any group element can be written as the exponential of a single element of the Lie algebra. In other words, if A^a are the canonical coordinates of the group element $g(A)$ and if T_a are the group generators obeying

$$[T_a, T_b] = C_{ab}{}^c\, T_c, \tag{6.1}$$

then

$$g(A) = \exp(A^a\, T_a). \tag{6.2}$$

A. Identification of measure

To achieve our aim, we will use the simplest of the gauge conditions (4.2), namely,

$$\chi^a(x, t) = \dot{A}_0{}^a(x, t). \tag{6.3}$$

Consider first the Faddeev–Popov operator M_χ. Using (6.3), (2.6c), and (3.2), one obtains

$$(M_\chi\,\epsilon)^a = \frac{\partial}{\partial t}(\dot\epsilon^{\,a} - C_{bc}{}^a A_0{}^b \epsilon^c). \tag{6.4}$$

Here we can assume that $A_0{}^b$ is time-independent since, due to (6.3) and the factor $\delta[\chi]$ in (3.1), only histories with $\dot A_0{}^a = 0$ contribute to the path integral.

The operator (6.4) does not involve spatial derivatives of ϵ, and therefore its determinant is a product over all points x of space, of the determinants of the operators defined by

$$(M_\chi(x)\,\epsilon)^a = \frac{d}{dt}(\dot\epsilon^{\,a} - C_{bc}{}^a A_0{}^b \epsilon^c), \tag{6.5}$$

where now the ϵ's are regarded just as functions of time, and not as fields $\epsilon(x,t)$.

Furthermore, as will be explicitly seen below, the determinant of the operator defined by (6.5) takes the form

$$(t_2 - t_1)^n \mu_L(\Lambda(x)), \tag{6.6}$$

(the index a runs from 1 to n), with

$$\Lambda^a(x) = (t_2 - t_1) A_0{}^a(x). \tag{6.7}$$

Therefore, we can write

$$\det M_\chi = \prod_x [\mu_L(\Lambda(x))(t_2 - t_1)^n]. \tag{6.8}$$

Next, before further analyzing (6.8), turn back to the transition amplitude (3.1). On account of the δ functional, the integration over $A_0{}^a(x,t)$ may be carried out for all times except one, t_1 say. Thus, using (6.8), the amplitude takes the form

$$\langle A_2,t_2|A_1,t_1\rangle = \int \prod_{x,a}[d\Lambda^a(x)\,\mu(\Lambda(x))] \times K[A^a{}_i(2), A^a{}_i(1); \Lambda^a, t_2, t_1], \tag{6.9}$$

where $\Lambda^a(x)$ may be thought of as being equal to $(t_2 - t_1) A_0{}^a(x, t_1)$.

The auxiliary amplitude K is given by

$$K = \int \prod_{x,t,i,a} dA^a{}_i(x,t)\, d\pi_a{}^i(x,t) \exp(iS). \tag{6.10}$$

It has been possible to factorize the determinant (6.8) out of the integration over $A^a{}_i$, $\pi_a{}^i$ because, due to the simple nature of the gauge condition, M_χ does not depend on $A^a{}_i$ or $\pi_a{}^i$.

Now, the path integral (6.10) is nothing but the matrix element of the evolution operator derived from the Hamiltonian,

$$H + \int A_0{}^a \gamma_a\, d^3x, \tag{6.11}$$

which depends on $A_0{}^a$ as a parameter and which, due to $\dot A_0{}^a = 0$, does not depend explicitly on time. Therefore, we may write

$$\langle A_2,t_2|A_1,t_1\rangle = \int \prod_{x,a}[d\Lambda^a(x)\,\mu_L(\Lambda(x))]\Big\langle A^a{}_i(2)\Big|\exp\Big[-i(t_2 - t_1)H - i\int \Lambda^a \gamma_a\, d^3x\Big]\Big|A^a{}_i(1)\Big\rangle, \tag{6.12}$$

which on account of (2.5) may be rewritten as

$$\langle A_2,t_2|A_1,t_1\rangle = \int \prod_{x,a}[d\Lambda^a(x)\,\mu_L(\Lambda(x))]\Big\langle A^a{}_i(2)\Big|\exp[-i(t_2 - t_1)H]\exp\Big(-i\int \Lambda^a \gamma_a\, d^3x\Big)\Big|A^a{}_i(1)\Big\rangle. \tag{6.13}$$

The content of (6.13) may be understood by recalling that, on account of (2.6a) and (2.6b) the operator γ_a is the generator of gauge transformations on $A^a{}_i$, $\pi_a{}^i$. This means that the effect of the factor $\exp(-i\int\Lambda^a\gamma_a)$ is simply to change,

$$A^a{}_i(1) \to (A^a(1))_{g(\Lambda)}, \tag{6.14}$$

with $g(\Lambda)$ given by (6.2), and where $(A^a{}_i)_g$ denotes the result of a (finite) gauge transformation by the group element g on the field $A^a{}_i(x)$. [The quantum operator $-i\gamma_a(x)$ is a representation of the group generator T_a at the point x. The factor $-i$ arises when passing from the Poisson brackets in (2.3) to commutators.] Thus we see that the quantities $(t_2 - t_1)A_0{}^a(x)$ play the role of canonical coordinates in the group copy at the point x.

Using (6.14), we may express (6.13) as

$$\langle A_2,t_2|A_1,t_1\rangle = \int \prod_{x,a}[d\Lambda^a(x)\,\mu_L(\Lambda(x))]\langle A^a{}_i(2)|\exp[-i(t_2 - t_1)H]|(A^a{}_i(1))_{g(\Lambda)}\rangle, \tag{6.15}$$

which shows that the dependence of the amplitude on the argument $A^a{}_i(1)$ is given by an average over the group at each point of space with a measure $\mu(\Lambda)$, where the Λ^a are canonical coordinates.

But, according to the discussion in Sec. IV, the amplitude must be invariant under gauge transformations in $A^a{}_i(1)$ [and also in $A^a{}_i(2)$, this comes next]. Hence $\mu(\Lambda)$ must be the group-invariant measure corresponding to the transformation (6.14), i.e., the left measure, in canonical coordinates. This last implication follows from the fact that the previous discussion goes through for any H that is group invariant [Eq. (2.5)] but otherwise arbitrary.

Lastly we analyze the behavior of (6.15) under gauge transformations of $A^a{}_i(2)$. To start, we observe that, on account of (2.5), we may pass the $\exp(-i\int\Lambda^a\gamma_a)$ factor to the left of $\exp[-i(t_2 - t_1)H]$ in (6.13), which yields the adjoint, $\exp(+i\int\Lambda^a\gamma_a^\dagger)$ acting on $A^a{}_i(2)$. However, we have from (2.2)

$$i\int \Lambda^a(x)\,[\gamma_a^\dagger(x) - \gamma_a(x)]\, d^3x = -\operatorname{Tr}\Omega, \tag{6.16}$$

where

$$\Omega^a{}_b(x,y) = \delta(x,y)\,\omega^a{}_b(\Lambda(x)), \tag{6.17}$$

with

$$\omega^a{}_b(\Lambda) = \Lambda^c C_{cb}{}^a, \tag{6.18}$$

being a matrix whose exponential gives the group element (6.18) in the adjoint representation $(T_a)^b{}_c = C_{ab}{}^c$. The trace operation in (6.16) involves both the discrete and continuous labels of Ω. [The factor ordering in (2.2), which yields (6.16), has been chosen because it is the one for which the action of $\exp(-i\int \Lambda^a \gamma_a)$ on $|A^a{}_i(1)\rangle$ is simply given by the transformation (6.14) of the eigenvalue, without bringing in a factor in front of the state vector.]

But, on account of (6.17), Ω has block form in the continuous labels, which yields for the exponential of (6.16),

$$\exp(-\operatorname{Tr}\Omega) = \det[\exp(-\Omega)]$$
$$= \prod_x \det\{\exp[-\omega(\Lambda(x))]\}. \tag{6.19}$$

Thus, we see that the effect of $\exp(+i\int \Lambda^a \gamma_a^\dagger)$ on $|A^a{}_i(2)\rangle$ amounts to inducing at each point a gauge transformation

$$A^a{}_i(2) \to (A^a{}_i(2))_{g^{-1}(\Lambda)}, \tag{6.20}$$

while simultaneously multiplying the amplitude by the factor (6.19). [In canonical coordinates $g(-\Lambda) = g^{-1}(\Lambda)$.] Hence we obtain the analogous relation to (6.15),

$$\langle A_2, t_2 | A_1, t_1\rangle$$
$$= \int \prod_{x,a} [d\Lambda^a(x)\, \mu_R(\Lambda(x))]$$
$$\times \langle (A^a{}_i(2))_{g^{-1}(\Lambda)} |\exp[-i(t_2 - t_1) H]| A^a{}_i(1)\rangle, \tag{6.21}$$

with

$$\mu_R(\Lambda) = \det\{\exp[-\omega(\Lambda)]\}\, \mu_L(\Lambda), \tag{6.22}$$

from which we conclude that (6.22) must be the invariant measure appropriated to the transformation (6.20), that is, the right measure. This conclusion will be explicitly confirmed after evaluating the measure.

[Several of the key points behind the above discussion may be found in the following simple example which is useful for fixing the ideas. Consider the quantum mechanical motion of a particle with Hamiltonian H in two space dimensions. Assume that H is invariant under rotations. Then the matrix element $\langle \mathbf{r}_2 |\exp[-i(t_2 - t_1) H]| \mathbf{r}_1\rangle$ takes the form $K(r_2, r_1, \phi)$, where $r_1 = |\mathbf{r}_1|$, $r_2 = |\mathbf{r}_2|$, and ϕ is the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$. Next expand K as

$$K(r_2, r_1, \phi) = \sum_m K_m(r_2, r_1)\, e^{im\phi}, \tag{6.23}$$

and find the response

$$K_\lambda(r_2, r_1, \phi) = K(r_2, r_1, \phi + \lambda)$$
$$= \sum_m k_m(r, r_1)\, e^{im\phi} e^{im\lambda}, \tag{6.24}$$

of K under a rotation of $\mathbf{r}_1$ by an angle λ. Further, assume that the average over λ of (6.24) with a measure μ:

$$\int_{-\pi}^{+\pi} K_\lambda(r_2, r_1, \phi)\, \mu(\lambda)\, d\lambda \tag{6.25}$$

is invariant under rotations of $\mathbf{r}_1$, i.e., independent of ϕ (which clearly makes also invariant under rotations in $\mathbf{r}_2$). This can only be true for an arbitrary choice of the functions $k_m(r_2, r_1)$—which characterize the rotationally invariant Hamiltonian H—if

$$\int_{-\pi}^{+\pi} e^{im\lambda} \mu(\lambda)\, d\lambda \propto \delta_{m,0}, \tag{6.26}$$

which means that $\mu(\lambda)$ must be a constant, independent of λ. But this is jut the (left and right) invariant measure for the group of rotations in two space dimensions.]

B. Evaluation of measure

Having identified $\mu_L(\Lambda)$ appearing in (6.13) as the left-invariant group measure expressed in canonical coordinates, we proceed to its explicit evaluation. For that purpose we will write the determinant of the differential operator defined by (6.5) in the form (5.6). Since we are working at a fixed space point, the integrals and products over x in (5.2) and (5.3) will be omitted.

According to (6.6) and (5.6), we have then

$$[\mu_L(\Lambda)]^{-1} = (t_2 - t_1)^n \langle \epsilon^a = \bar{\epsilon}_a = 0$$
$$\times |\exp[-i(t_2 - t_1) h]| \epsilon^a = \bar{\epsilon}_a = 0\rangle, \tag{6.27}$$

where

$$h = (-\pi^a + C_{bc}{}^a A_0{}^b \epsilon^c)\, \bar{\pi}_a \tag{6.28}$$

is the Hamiltonian derived from the Lagrangian,

$$l = \dot{\bar{\epsilon}}_a(\dot{\epsilon}^a - C_{bc}{}^a A_0{}^b \epsilon^c), \tag{6.29}$$

and A_0 is related to Λ by (6.7). Note that (6.28) does not depend explicitly on time due to the gauge condition $\dot{A}_0{}^a = 0$.

The matrix element (6.27) may be evaluated with the help of the change of variables,

$$\phi = \exp[-(t_2 - t_1)^{-1}(t - t_1)\,\omega]\, \epsilon, \tag{6.30}$$

$$\bar{p}^T = \bar{\pi}^T \exp[(t_2 - t_1)^{-1}(t - t_1)\,\omega], \tag{6.31}$$

where ω is given by (6.18) and we have used matrix notation, with T denoting transposition.

The transformation (6.30), (6.31) is canonical but time-dependent and hence changes the value of the Hamiltonian. The new Hamiltonian is

$$\tilde{h} = -p^T \exp[-(t_2 - t_1)^{-1}(t - t_1)\,\omega]\, \pi, \tag{6.32}$$

and depends only on the momenta.

The operator (6.32) is explicitly time-dependent, but has the property

$$[\tilde{h}(t), \tilde{h}(t')] = 0, \tag{6.33}$$

for any t, t', hence no time ordering is needed for its evolution operator which reads

$$\tilde{U}(t_2, t_1) = \exp\left[-i\int_{t_1}^{t_2} \tilde{h}(t)\, dt\right]$$
$$= \exp\{i\bar{p}^T[(1 - e^{-\omega})/\omega](t_2 - t_1)\,\pi\}. \tag{6.34}$$

Next, we must evaluate the matrix element of (6.34) between states with $\epsilon = \bar{\epsilon} = 0$. However, (6.34) is constructed in terms of the conjugates of ϕ and $\bar{\epsilon}$. In order to reexpress

(6.27) as a matrix element between orthonormalized eigenstates of ϕ and $\bar{\epsilon}$, we must introduce a normalizing factor. This factor is most easily obtained by changing the integration variables in the time-sliced version of the Hamiltonian path integral expression of (6.27), and recalling that in a "q to q" matrix element there is an "extra integration" over p at the upper end point [see, for example, Eq. (1.12) of Ref. 1]. One obtains

$$[\mu_L(A)]^{-1} = (t_2 - t_1)^n \det[\exp(-\omega)]\langle \phi = \bar{\epsilon} = 0 \times |\tilde{U}(t_2, t_1)| \phi = \bar{\epsilon} = 0\rangle. \quad (6.35)$$

The matrix element in (6.35) is easily evaluated by going to eigenstates of $\bar{p}$ and π. This gives a Gaussian integral over $\bar{p}$ and π, and one can apply (5.1) to obtain

$$[\mu_L(A)]^{-1} = (t_2 - t_1)^n \det[\exp(-\omega)] \times \det\{[(1 - e^{-\omega})/\omega](t_2 - t_1)\}^{-1}, \quad (6.36)$$

which yields finally

$$\mu_L(A) = \det[(e^{\omega(A)} - 1)/\omega(A)]. \quad (6.37)$$

Equation (6.37) is known[4] as the expression for the left-invariant measure over a Lie group in canonical coordinates. The right invariant measure is obtained by changing $\omega \to -\omega$, which amounts to multiplying (6.37) by $\exp(-\omega)$, in agreement with (6.22). Note, incidentally, that the right-invariant measure may be also represented in the form (6.27) provided we change in that equation $h \to h^\dagger$. The verification of this statement follows the same lines as the discussion between (6.15) and (6.22).

VII. INVARIANT MEASURE FOR THE ROTATION GROUP IN TERMS OF EULER ANGLES

Why did we obtain above the group measure in canonical coordinates rather than in any other parametrization of the group? Clearly the only choice ever was made when selecting the gauge condition. So the condition $\dot{A}_0{}^a = 0$ corresponds to choosing canonical coordinates in the group.

If one looks back at the analysis in the previous section, one sees that the canonical coordinates arose because the condition $\dot{A}_0{}^a = 0$ produced in (6.13) a factor $\exp[-i(t_2 - t_1)\int A_0{}^a \gamma_a]$, which is just the defining relation (6.2) of the canonical coordinates. This means that if we want to evaluate group measures in terms of other group coordinates, we must tailor the gauge condition to the chosen group coordinates.

To explain more clearly what is meant here by "tailoring," we will derive in what follows the invariant measure for the rotation group in three dimensions, in terms of Euler angles. It will be apparent, however, that the method is applicable to any case in which the group coordinates are defined by expressing a generic group element as a product of exponentials of elements of the Lie algebra. The canonical coordinates are the simplest such case, with only one factor being presented in the product.

A. Choice of gauge condition and identification of measure

For the rotation group we have

$$C_{bc}{}^a = \epsilon_{abc}, \quad (7.1)$$

and the Euler angles are defined by writing any group element in the form

$$g(\psi, \theta, \phi) = e^{\psi T_1} e^{\theta T_2} e^{\phi T_1}. \quad (7.2)$$

In order to choose a gauge condition tailored to (7.2), we divide the total time interval $t_1 < t < t_2$ in three subintervals, which we shall call I, II, III, and which are defined respectively by $t_1 < t < t'$, $t' < t < t''$, $t'' < t < t_2$. The dividing times t' and t'' are arbitrary (but such that $t_1 < t' < t'' < t_2$). They will drop out from the final answer for the measure. (And t_1, t_2, too.)

We impose then *a different gauge condition for each subinterval*, namely,

$$\dot{A}_0{}^1 = 0, \quad A_0{}^2 = 0, \quad A_0{}^3 = 0, \quad \text{for interval I}, \quad (7.3a)$$
$$A_0{}^1 = 0, \quad \dot{A}_0{}^2 = 0, \quad A_0{}^3 = 0, \quad \text{for interval II}, \quad (7.3b)$$
$$\dot{A}_0{}^1 = 0, \quad A_0{}^2 = 0, \quad A_0{}^3 = 0, \quad \text{for interval III}. \quad (7.3c)$$

The reason for (7.3) is that the product of exponentials in (7.2) will arise from the fact that the gauge transformations generated by the term $A_0{}^a \gamma_a$ in (6.11) consists of three successive steps in time according to (7.3). It will be seen below that, although not of the form (4.2), (7.3) is indeed a permissible gauge condition.

In order to find the transition amplitude from $A^a{}_i(1)$ at time t_1 to $A^a{}_i(2)$ at time t_2, we will first evaluate separately the amplitudes for the subintervals I, II, and III and will subsequently fold them together.

Due to the simple nature of the gauge conditions (7.3), the ghost contributions will not involve the fields $A^a{}_i$, $\pi_a{}^i$. Therefore, in each interval, one can factorize the $\delta[\chi] \det M_\chi$ term in (3.1) out of the functional integration in $A^a{}_i$, $\pi_a{}^i$, just as in (6.9), (6.10). Furthermore, one can also here use the $\delta[\chi]$ to perform all "integrations but one" on $A_0{}^a$ for each interval. In this way one gets auxiliary amplitudes analogous to (6.10), which are given by

$$K_{\rm I} = \langle A^a{}_i(t')|\exp[-i(t' - t_1)H] \times \exp\left[-i(t' - t_1)\int A_0{}^1({\rm I})\gamma_1 d^3x\right]\left| A^a{}_i(t_1)\right\rangle, \quad (7.4a)$$

$$K_{\rm II} = \langle A^a{}_i(t'')|\exp[-i(t'' - t')H] \times \exp\left[-i(t'' - t')\int A_0{}^2({\rm II})\gamma_2 d^3x\right]\left| A^a{}_i(t')\right\rangle, \quad (7.4b)$$

$$K_{\rm III} = \langle A^a{}_i(t_2)|\exp[-i(t_2 - t'')H] \times \exp\left[-i(t_2 - t'')\int A_0{}^1({\rm III})\gamma_1 d^3x\right]\left| A^a{}_i(t'')\right\rangle, \quad (7.4c)$$

In Eqs. (7.4), $A_0^1(\mathrm{I})$ is the time-independent value of A_0^1 in interval I, $A_0^2(\mathrm{II})$ is the time-independent value of A_0^2 in interval II, and $A_0^1(\mathrm{III})$ is the time-independent value of A_0^1 in interval III. These functions of x are to be integrated over as independent variables at the end, since the path integral (3.1) involves an integration over $A_0^a(x, t)$ for all x and for all t in the interval $t_1 < t < t_2$. Thus the A_0^1 in (7.4a) is not related to the A_0^1 in (7.4c).

In order to complete the integration over $A^a_i(x, t)$ in (3.1), we must multiply together the amplitudes (7.4) and integrate over the field variables $A^a_i(t')$ and $A^a_i(t'')$ at the joining points. This gives for the total auxiliary amplitude the formula

$$K[\phi, \theta, \psi] = \Big\langle A^a_i(2) \Big| \exp[-i(t_2 - t_1)H] \cdot\exp\Big(-i\int \psi\gamma_1\, d^3x\Big)\cdot\exp\Big(-i\int \theta\gamma_2\, d^3x\Big) \cdot\exp\Big(-i\int \phi\gamma_1\, d^3x\Big)\Big| A^a_i(1)\Big\rangle, \tag{7.5}$$

where we have defined

$$\phi(x) = (t' - t_1) A_0^1(x)(\mathrm{I}), \tag{7.6a}$$

$$\theta(x) = (t'' - t') A_0^2(x)(\mathrm{II}), \tag{7.6b}$$

$$\psi(x) = (t_2 - t'') A_0^1(x)(\mathrm{III}), \tag{7.6c}$$

the amplitude (3.1) is obtained from (7.5) by the analogous relation to (6.9), namely,

$$\langle A_2, t_2 | A_1, t_1\rangle = \int \prod_x [d\phi(x)\, d\theta(x)\, d\psi(x)(t_2 - t'')^{-1} \times (t'' - t')^{-1}(t' - t_1)^{-1}] \det M_\chi\, K[\phi, \theta, \psi]. \tag{7.7}$$

Equation (7.7) shows, by the same reasoning that followed (6.13) and led to the identification of (6.8) as the group measure in canonical coordinates, that $\mu(\phi, \theta, \psi)$ appearing in

$$\det M_\chi = \prod_x [\mu(\phi(x), \theta(x), \psi(x)) \times (t_2 - t'')(t'' - t')(t' - t_1)] \tag{7.8}$$

is the group-invariant measure in terms of Euler angles. [Here the right and left measures coincide because (7.1) has zero trace.]

B. Evaluation of measure

The operator M_χ appearing in (7.8) is defined in the whole interval $t_1 < t < t_2$ by the collection of three different expressions for the subintervals I, II, III. Those expressions are obtained by applying the gauge transformation (2.6c) to the gauge conditions (7.3) and using the definition (3.2). One obtains:

Interval I:

$$(M_\chi\, \epsilon)^1 = \ddot\epsilon^1, \tag{7.9a}$$

$$(M_\chi\, \epsilon)^2 = \dot\epsilon^2 + (t' - t_1)^{-1}\phi\epsilon^3, \tag{7.9b}$$

$$(M_\chi\, \epsilon)^3 = \dot\epsilon^3 - (t' - t_1)^{-1}\phi\epsilon^2. \tag{7.9c}$$

Interval II:

$$(M_\chi\, \epsilon)^1 = \dot\epsilon^1 - (t'' - t')^{-1}\theta\epsilon^3, \tag{7.10a}$$

$$(M_\chi\, \epsilon\omega)^2 = \ddot\epsilon^2, \tag{7.10b}$$

$$(M_\chi\, \epsilon)^3 = \dot\epsilon^3 + (t'' - t')^{-1}\,\theta\epsilon^1. \tag{7.10c}$$

Interval III:

$$(M_\chi\, \epsilon)^1 = \ddot\epsilon^1, \tag{7.11a}$$

$$(M_\chi\, \epsilon)^2 = \dot\epsilon^2 + (t_2 - t'')^{-1}\psi\epsilon^3, \tag{7.11b}$$

$$(M_\chi\, \epsilon)^3 = \dot\epsilon^3 - (t_2 - t'')^{-1}\psi\epsilon^2. \tag{7.11c}$$

It is straightforward to verify that the equation

$$M_\chi\, \epsilon = 0, \tag{7.12}$$

for $t_1 < t < t_2$, with the boundary conditions

$$\epsilon^a(t_1) = \epsilon^a(t_2) = 0, \tag{7.13}$$

has the only solution $\epsilon = 0$ provided one assumes continuity at t' and t''. This means that M has an inverse and therefore the gauge conditions (7.3) are (locally) permissible. Actually it is somewhat unnecessary to check this since: (i) One knows that the Euler angles are good coordinates for the rotation group and (ii) the gauge conditions (7.3) just establish that one goes from the initial "gauge frame" to the final one by acting on the fields with a group element of the form (7.2) at every space point.

Next we observe that M_χ does not involve space derivatives so that its determinant is indeed an infinite product over x of the form (7.8). Therefore, we will consider ϵ as a function of time only just as in Sec. VIB. With that understanding, (7.8) will be valid with the x dependence omitted throughout, with the infinite product deleted, and with M_χ defined by (7.9)–(7.11).

We would now like to represent the determinant of M_χ as a transition amplitude. Just as (7.5), that amplitude will result from folding together the three different amplitudes corresponding to the subintervals I, II, and III. The folding operation is, however, subtler than the one which led from (7.4) to (7.5). The subtlety resides in the boundary conditions that the additional fields $\bar\epsilon$ which appear in the action S_χ must satisfy.

The analysis proceeds as follows. Consider the first interval. There we have, from (5.2),

$$S_\chi{}^{\mathrm{I}} = \int_{t_1}^{t'} dt\, \{\bar\epsilon_1\, \ddot\epsilon^1 + \bar\epsilon_2 + [\dot\epsilon^2 + (t' - t_1)^{-1}\phi\epsilon^3] + \bar\epsilon_3[\dot\epsilon^3 - (t' - t_1)^{-1}\phi\epsilon^2]\}. \tag{7.14}$$

To start with, the only thing we know is fixed in this action is $\epsilon^a(t_1) = 0$. The value $\epsilon^a(t')$ is free since nothing restricts the gauge freedom at that intermediate time. However the situation with $\bar\epsilon_a$ is different. First of all we see that, in order to eliminate the second time derivative of ϵ^1 in (7.14), we must demand

$$\bar\epsilon_1(t_1) = \bar\epsilon(t') = 0. \tag{7.15}$$

Second, we observe that (7.14) is already in Hamiltonian form in ϵ^2 and ϵ^3, with $\bar\epsilon_2$ and $\bar\epsilon_3$ being nothing but the momenta $\bar\pi_2$ and $\bar\pi_3$ canonically conjugate to ϵ^2 and ϵ^3. This means that $\bar\epsilon_2$ and $\bar\epsilon_3$ are not fixed at t_1 and t'.

Thus, if we pass to Hamiltonian form in $\bar{\epsilon}_1$, ϵ^1 as well, we find that the partial amplitude in the first interval has the form

$$K_\chi{}^{\rm I}[\epsilon^a(t'), \epsilon^a(t_1)=0]$$
$$= \langle \epsilon^a(t'), \bar{\epsilon}_1(t')=0 | \exp[-i(t'-t_1)h^{\rm I}] |$$
$$\times | \epsilon^a(t_1)=0, \bar{\epsilon}_1(t_1)=0 \rangle, \tag{7.16}$$

where

$$h^{\rm I} = -\bar{\pi}_1\pi^1 - (t'-t_1)^{-1}\phi(\epsilon^3\bar{\pi}_2 - \epsilon^2\bar{\pi}_3). \tag{7.17}$$

Similarly, we find, for the other two intervals,

$$K_\chi{}^{\rm II}[\epsilon^a(t''), \epsilon^a(t')]$$
$$= \langle \epsilon^a(t''), \bar{\epsilon}_2(t')=0 | \exp[-i(t''-t')h^{\rm II}] |$$
$$\times | \epsilon^a(t'), \bar{\epsilon}_2(t')=0 \rangle, \tag{7.18}$$

with

$$h^{\rm II} = -\bar{\pi}_2\pi^2 - (t''-t')^{-1}\theta(\epsilon^1\bar{\pi}_3 - \epsilon^3\bar{\pi}_1), \tag{7.19}$$

and

$$K_\chi{}^{\rm II}[\epsilon^a(t_2)=0, \epsilon^a(t'')]$$
$$= \langle \epsilon^a(t_2)=0, \bar{\epsilon}_1(t_2)=0 | \exp[-i(t_2-t'')h^{\rm III}] |$$
$$\times | \epsilon^a(t''), \bar{\epsilon}_1(t'')=0 \rangle, \tag{7.20}$$

with

$$h^{\rm III} = -\bar{\pi}_1\pi^1 - (t_2-t'')^{-1}\psi(\epsilon^3\bar{\pi}_2 - \epsilon^2\bar{\pi}_3). \tag{7.21}$$

The total amplitude is obtained by multplying together $K_\chi{}^{\rm I}$, $K_\chi{}^{\rm II}$, $K_\chi{}^{\rm III}$ and integrating the product over $\epsilon^a(t')$ and $\epsilon^a(t'')$. Hence, the folding of the amplitudes $K_\chi{}^{\rm I}$, $K_\chi{}^{\rm II}$ and $K_\chi{}^{\rm III}$ is done only over the variables ϵ^a which appear uniformly in the three intervals with a term $\bar{\pi}_a\dot{\epsilon}^a$ in the corresponding Hamiltonian action. In fact, the need for integrating over $\epsilon^a(t')$ and $\epsilon^a(t'')$ in the amplitude may be thought of as being the quantum mechanical origin of the continuity assumption for ϵ^a at t' and t'' in the classical history determined by (7.12) and (7.13).

The additional variables $\bar{\epsilon}_a$ are treated in a different way. As the Hamiltonian analysis shows, there is really one $\bar{\epsilon}$ coming in into each interval instead of all three of them, and that one is fixed at both end points of the interval. Here the Hamiltonian form is most useful in preventing one from erroneously treating the $\bar{\epsilon}_a$, which appear in the Lagrangian form (7.14) and its analogs for the other intervals, as being different parts of one overall history for $t_1 < t < t_2$.

The partial amplitudes $K_\chi{}^{\rm I}$ and $K_\chi{}^{\rm II}$ are most easily evaluated. Indeed the two terms in $h^{\rm I}$ commute, so the exponential in (7.16) factors into the product of the exponentials of the two terms. Furthermore, the operator $\epsilon^3\bar{\pi}_2 - \epsilon^2\bar{\pi}_3$ is nothing but the generator of a rotation in the ϵ^2, ϵ^3 plane, an operation which has no effect on ϵ^1 and also leaves invariant $\epsilon^2 = \epsilon^3 = 0$. Hence we find

$$K_\chi{}^{\rm I} = \langle \epsilon^a(t'), \bar{\epsilon}_1(t')=0 |$$
$$\times \exp[-i(t'-t_1)\bar{\pi}_1\pi^1] | \epsilon^a(t_1)=0, \bar{\epsilon}_1(t_1)=0 \rangle$$
$$= \delta(\epsilon^2(t'))\,\delta(\epsilon^3(t'))(t'-t_1)^{-1}, \tag{7.22}$$

and, similarly,

$$K_\chi{}^{\rm III} = \delta(\epsilon^2(t''))\,\delta(\epsilon^3(t''))(t_2-t'')^{-1}. \tag{7.23}$$

From these equations we see in particular that the dependence on the first and last Euler angles, ϕ and ψ, disappears so the measure will depend only on θ.

If we now multiply (7.22), (7.23), and (7.18), and integrate the result over $\epsilon^a(t'')$ and $\epsilon^a(t')$ for $a = 1,2,3$, we find, recalling (7.8),

$$\mu^{-1}(\theta) = (t''-t')\int d\epsilon^1(t')\,d\epsilon^1(t'')\langle \epsilon^1(t''),$$
$$\epsilon^2 = \epsilon^3 = \bar{\epsilon}_2 = 0 |$$
$$\times \exp[-i(t''-t')h^{\rm II}] | | \epsilon^1(t'),$$
$$\epsilon^2 = \epsilon^3 = \bar{\epsilon}_2 = 0 \rangle. \tag{7.24}$$

But again here the two terms in $h^{\rm II}$ commute, and we obtain in (7.24) a product of exponentials, which permits us to reduce immediately that equation to

$$\mu^{-1}(\theta) = \int d\epsilon^1(t')\,d\epsilon^1(t'')\,\langle \epsilon^1(t''), \epsilon^3 = 0 |$$
$$\times \exp[-i\theta(\epsilon^3\bar{\pi}_1 - \epsilon^1\bar{\pi}_3)] | | \epsilon^1(t'), \epsilon^3 = 0 \rangle. \tag{7.25}$$

Now, $\epsilon^2\bar{\pi}_1 - \epsilon^1\bar{\pi}_3$ is the generator of rotations in the ϵ^1, ϵ^3 plane. Hence the action of its exponential on $|\epsilon^1, \epsilon^3\rangle$ is to rotate the eigenvalues as the components of a vector, by an angle θ. It follows that

$$\mu^{-1}(\theta) = \int d\epsilon^1)(t')\,d\epsilon^1(t'')\,\langle \epsilon^1(t''), \epsilon^3 = 0 |$$
$$\times \cos\theta\epsilon^1(t'), \sin\theta\epsilon^1(t')\rangle$$
$$= \int d\epsilon^1(t')\,d\epsilon^1(t'')\,\delta[\epsilon^1(t'')$$
$$-\cos\theta\epsilon^1(t')]\,\delta[\sin\theta\epsilon^1(t')] = (\sin\theta)^{-1}. \tag{7.26}$$

So we conclude

$$\mu(\theta) = \sin\theta, \tag{7.27}$$

which is the well-known expression for the invariant measure over the rotation group in terms of the Euler angles (7.2).

ACKNOWLEDGMENTS

I would like to thank Marc Henneaux for many useful discussions and John Klauder for valuable suggestions. This work was supported in part by a U. S. National Science Foundation Grant to the University of Texas at Austin and by the Educational Program of the Region Piemonte/Istituto di Fisica Teorica dell' Università di Torino, Torino, Italy. The author is a recipient of a John S. Guggenheim Fellowship and of a University of Texas Faculty Research Award. For kind hospitality he would like to thank G. Ferrero of the Regione Piemonte, M. Rasetti of the Educational Program and T. Regge of the Istituto di Fisica Teorica dell' Università di Torino. Finally, thanks are expressed to David Brown and Alvaro Díaz for carefully reading the manuscript.

[1]See L. D. Faddeev and A. A. Slavnov, *Gauge Fields, Introduction to Quantum Theory* (Benjamin/Cummings, Reading, MA, 1980), and references therein.
[2]C. Teitelboim, Phys. Rev. D **25**, 3159 (1982); **28**, 297 (1983); Phys. Rev. Lett. **50**, 705 (1983).
[3]R. P. Feynmann, Rev. Mod. Phys. **20**, 267 (1948).
[4]See, for example, B. S. DeWitt, in *Relativity Groups and Topology, Les Houches 1963*, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1964); S. Helgason, in *Batelle Recontres 12967*, edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

The Hamilton–Jacobi theory of De Donder and Weyl applied to some relativistic field theories[a]

J. von Rieth
Institut für Theoretische Physik, RWTH Aachen, 5100 Aachen, Federal Republic of Germany

(Received 14 February 1983; accepted for publication 25 May 1983)

The Hamilton–Jacobi (HJ) theory for fields initiated by De Donder and Weyl is applied to Klein–Gordon, Dirac, and gauge fields. Conserved HJ currents are derived for each parameter on which a solution of the field equations depends, and the concept of a "complete" integral of the De Donder–Weyl HJ equation is discussed from which solutions of the field equations may be constructed.

PACS numbers: 11.15. – q, 11.10.Qr

1. INTRODUCTION

In the 1930s De Donder[1] and Weyl[2] developed the concept of a Hamilton–Jacobi (HJ) theory for field equations derived from a variational action integral with a Lagrangian density $\mathscr{L} = \mathscr{L}(x, z, v)$, where $x = (x^1,..., x^m)$, $z = (z^1,..., z^n)$, $v = (v_1^1,..., v_m^1, v_1^2,..., v_m^2,..., v_1^n,..., v_m^n)$ such that $v_\mu^a = \partial_\mu f^a(x)$, $a = 1,..., n$, $\mu = 1,..., m$, on the extremals $z^a = f^a(x)$. It is the purpose of this paper to discuss some applications of the De Donder–Weyl HJ theory to relativistic fields.

In Sec. 2 we shall briefly summarize the basic features of the De Donder–Weyl HJ theory which recently was discussed extensively in a review article by Kastrup[3]: In the canonical framework of De Donder and Weyl the generalization of the HJ equation

$$\partial_t S(t, q) + H(t, q, p = \partial S) = 0, \quad p_j = \partial_j S(t, q)$$

for the function $S(t, q)$ in mechanics is the partial differential equation (pde)

$$\begin{aligned} &\partial_\mu S^\mu(x, z) + \mathscr{H}(x, z, \pi^\mu_a = \partial_a S^\mu(x, z)) = 0\,, \\ &\pi^\mu_a := \frac{\partial \mathscr{L}}{\partial v_\mu^a}, \quad \mathscr{H}(x, z, \pi) = \pi^\mu_a v_\mu^a - \mathscr{L}\,, \qquad (1.1) \\ &\partial_\mu := \frac{\partial}{\partial x^\mu}, \quad \partial_a := \frac{\partial}{\partial z^a} \end{aligned}$$

for the m functions $S^\mu(x, z)$ [in Eq. (1.1) and in the following the summation convention is assumed to hold]. It is obvious from Eq. (1.1) that it reduces to the one of mechanics if $m = 1$.

In Sec. 3 we discuss the following applications of solutions $S^\mu(x, z)$:

(i) If $S^\mu(x, z)$ depends on a parameter α, then the current

$$G^\mu(x; \alpha) := \frac{\partial S^\mu}{\partial \alpha}\bigg|_{z^a = f^a(x)}, \quad \mu = 1,..., m\,, \qquad (1.2)$$

is conserved for any extremal $z^a = f^a(x)$ for which the "transversality" property

$$\pi^\mu_a(x) = \partial_a S^\mu(x, z)\big|_{z^b = f^b(x)}, \quad \pi^\mu_a(x) := \frac{\partial \mathscr{L}}{\partial v_\mu^a}(x, f(x), \partial f(x)) \qquad (1.3)$$

holds. A special application of this result is the following: Given an extremal, $z^a = f^a(x)$, which depends on a parameter α. Then it is possible to construct a solution $S^\mu(x, z; \alpha)$ such that the property (1.3) holds for that particular extremal $z^a = f^a(x; \alpha)$ [the extremal $z = f(x)$ is said to be "embedded" (weakly) in a solution $S^\mu(x, z)$ of the HJ equation (1.1)].

(ii) In mechanics a complete set of solutions of the equations of motion can be constructed from a solution $S(t, q; \alpha)$ of the HJ equation, which depends on n parameters α_j, $j = 1,..., n$ such that

$$\det\left[\frac{\partial^2 S}{\partial q^j\, \partial \alpha_k}\right] \neq 0$$

by solving a set of n algebraic equations for n functions $q^j = f^j(t)$. Similarly, suppose a solution $S^\mu(x, z; \alpha)$ depends on $m \cdot n$ parameters α^μ_a, $\mu = 1,..., m$, $a = 1,..., n$ such that

$$\det\left[\frac{\partial^2 S^\mu}{\partial z^a\, \partial \alpha^\nu_b}\right] \neq 0\,,$$

then it is possible to construct solutions $z^a = f^a(x)$ of the Euler–Lagrange equations, provided a set of $m^2 \cdot n$ algebraic equations in the variables x and z possesses a degeneracy such that it can be solved for the n functions $z^a = f^a(x)$.

In Sec. 4 we discuss two methods for solving the HJ equation (1.1). The first one uses Cauchy's method of characteristics, and the second one uses a polynomial ansatz in the variables z already introduced by Weyl.[2] The resulting differential equations of the second approach which are of the matrix Riccati type are closely related to Jacobi's equations of the second variation for field theories. In Sec. 5A we apply the results of the preceding chapters to the following relativistic field theories: free Klein–Gordon field, E-dynamics with an external current, Klein–Gordon field in an external electromagnetic field, the system of coupled electromagnetic and scalar fields, the Dirac field without and with external electromagnetic field, and the SU(2)–Yang–Mills field. In Sec. 5B we derive a general form of a conserved current associated with a parameter α on which a solution $f^a(x; \alpha)$ of the Euler–Lagrange equations depends and apply the results to plane waves and instantons. Section 5C contains a simple example for a "complete" integral of the HJ equation (1.1), from which solutions of the Euler–Lagrange equations can be constructed by solving algebraic equations.

[a] Work supported by the Deutsche Forschungsgemeinschaft.

 0022-2488/84/041102-14$02.50

2. THE DE DONDER–WEYL HJ EQUATION

In this chapter the De Donder–Weyl HJ equation for fields with m independent variables will be deduced, following ideas of Caratheodory,[4] introduced by him in the framework of mechanics and generalized by Boerner[5] to the case of field theories.

For reasons of convenience we restrict ourselves here to the physically important case $m = 4$.

Here we have four independent variables x^μ, $\mu = 0,...,3$, $x = (x^0, x^1, x^2, x^3) \in V \subset M^4$ [M^4 is the four-dimensional Minkowski space with metric $(+---)$] and variables $z = (z^1,..., z^n) \in G^n (n \geqslant 1$, G^n is a subset of $\mathbb{C}^n$ or $\mathbb{R}^n$), which become dependent variables on four-dimensional submanifolds $\Sigma_4 \subset V \otimes G^n$,

$$z^a = z^a(x), \quad x \in V, \quad a = 1,..., n\,. \tag{2.1}$$

The values of the functions (2,1) on the boundary ∂V of V are assumed to be the same for all functions.

For a given Lagrangian $\mathscr{L} = \mathscr{L}(x^\mu, z^a, v^a_\mu)$ [the variables v^a_μ become dependent variables $\partial_\mu z^a(x)$ on Σ_4] we shall now minimize the action integral

$$A(\Sigma_4) := \int_V d^4x\, \mathscr{L}(x^\mu, z^a(x), \partial_\mu z^a(x))\,. \tag{2.2}$$

Instead of the variational problem for $\mathscr{L}$ we can look at the corresponding problem for

$$\tilde{\mathscr{L}}(x, z, \partial z) := \mathscr{L}(x, z, \partial z) - \Phi(x, z, \partial z)\,, \tag{2.3}$$

where the function Φ is defined by

$$\Phi(x, z, \partial z) := \frac{dS^\mu}{dx^\mu}(x, z) \equiv \partial_\mu S^\mu(x, z) + \partial_a S^\mu(x, z)\partial_\mu z^a(x)\,. \tag{2.4}$$

The quantities $S^\mu(x, z)$ are functions of x and z which are still to be determined as follows: Because the integral

$$\int_V d^4x \frac{dS^\mu}{dx^\mu} = \int_{\partial V} S^\mu\, d^3\Sigma_\mu\,,$$

$$d^3\Sigma_\mu := \frac{1}{3!}\epsilon_{\mu\nu\rho\sigma}\, dx^\nu \wedge dx^\rho \wedge dx^\sigma\,,$$

depends only on the value of the functions $S^\mu(x, z)$ on the boundary ∂V, the integral

$$\tilde{A}(\Sigma_4) := \int_V d^4x\, \tilde{\mathscr{L}}(x, z(x), \partial z(x)) \tag{2.5}$$

will be extremal iff this is the case for the integral (2.2), too. In this sense the two variational problems are equivalent. We now seek functions $S^\mu(x, z)$ and functions $\varphi^a_\mu(x, z)$ (φ^a_μ is called a *geodesic field*) such that the properties

$$\tilde{\mathscr{L}}(x, z, \varphi) = 0 \tag{2.6}$$

and

$$\tilde{\mathscr{L}}(x, z, v) > 0, \quad v^a_\mu \neq \varphi^a_\mu\,, \tag{2.7}$$

hold. Obviously the solutions of

$$\partial_\mu z^a(x) = \varphi^a_\mu(x, z(x)) \tag{2.8}$$

minimize the integral (2.5) because of the properties (2.6), (2.7), and therefore the action integral (2.2) becomes minimal, too. [We remark that the existence of functions S^μ and φ^a_μ which satisfy the conditions (2.6) and (2.7) is *one* sufficient but not necessary criterion for having minima of (2.2).] A necessary condition for $\tilde{\mathscr{L}}$ to have a local minimum at $v^a_\mu = \varphi^a_\mu$ is

$$\left.\frac{\partial \mathscr{L}}{\partial v^a_\mu}\right|_{v^b_\nu = \varphi^b_\nu} = \partial_a S^\mu\,. \tag{2.9}$$

It follows from Eqs. (2.6), (2.8), and (2.9) that

$$\partial_\mu S^\mu(x, z) = \mathscr{L}(x, z, \varphi) - \left.\frac{\partial \mathscr{L}}{\partial v^a_\mu}\right|_{v^b_\nu = \varphi^b_\nu} \cdot \varphi^a_\mu(x, z)\,. \tag{2.10}$$

Equations (2.4), (2.9), and (2.10) imply that the integral

$$\hat{A}(\hat{\Sigma}_4) := \int_V d^4x \Big\{ \mathscr{L}[x, \hat{z}(x), \varphi(x, \hat{z}(x))] + \frac{\partial \mathscr{L}}{\partial v^a_\mu}[x, \hat{z}(x), \varphi(x, \hat{z}(x))]\,[\partial_\mu \hat{z}^a(x) - \varphi^a_\mu(x, \hat{z}(x))] \Big\} \tag{2.11}$$

is independent of the choice of the functions $\hat{z}(x)$ inside V. (This is a generalization of Hilbert's famous "independent integral" for mechanics.[6])

We now introduce canonical momenta

$$\pi^\mu_a := \frac{\partial \mathscr{L}}{\partial v^a_\mu} \tag{2.12}$$

and assume that

$$\det\left[\frac{\partial^2 \mathscr{L}}{\partial v^a_\mu \partial v^b_\nu}\right] \neq 0\,. \tag{2.13}$$

Then the Legendre transformation $v^a_\mu \to \pi^\mu_a$ is regular. The Hamilton function associated with the Legendre transformation is

$$\mathscr{H} := \pi^\mu_a v^a_\mu - \mathscr{L} = \mathscr{H}(x, z, \pi)\,. \tag{2.14}$$

For the geodesic field we get, according to Eq. (2.9),

$$\psi^\mu_a(x, z) := \left.\frac{\partial \mathscr{L}}{\partial v^a_\mu}\right|_{v^b_\nu = \varphi^b_\nu} = \partial_a S^\mu(x, z)\,, \tag{2.15}$$

whereas the relation (2.10) can be written as

$$\partial_\mu S^\mu(x, z) = \mathscr{L}(x, z, \varphi) - \psi^\mu_a(x, z)\varphi^a_\mu(x, z)\,. \tag{2.16}$$

Combining Eqs. (2.14) and (2.16), we obtain the so-called *Hamilton–Jacobi equation*

$$\pi^\mu_a = \partial_a S^\mu(x, z) = \psi^\mu_a(x, z)\,, \tag{2.17a}$$

$$\partial_\mu S^\mu(x, z) + \mathscr{H}(x, z, \pi^\mu_a = \partial_a S^\mu(x, z)) = 0\,. \tag{2.17b}$$

We see that the functions S^μ introduced above must satisfy Eqs. (2.17). Taking the differential of the function $\mathscr{H}$, we obtain $v^a_\mu = \partial\mathscr{H}/\partial\pi^\mu_a$ and therefore the geodesic field

$$\varphi^a_\mu(x, z) = \left.\frac{\partial \mathscr{H}}{\partial \pi^\mu_a}\right|_{\pi^\nu_b = \psi^\nu_b}\,. \tag{2.18}$$

As already mentioned in connection with Eqs. (2.8), the solutions $z^a(x)$ of

$$\partial_\mu z^a(x) = \varphi^a_\mu(x, z(x)) = \left.\frac{\partial \mathscr{H}}{\partial \pi^\mu_a}\right|_{\pi^\nu_b = \psi^\nu_b} \tag{2.19}$$

are functions which fulfill the necessary conditions for the existence of an extremum, that is, they also satisfy[7,8]

$$\partial_\mu \pi^\mu_a = -\partial_a \mathscr{H}. \tag{2.20}$$

The pde's (2.19) and (2.20) are the two sets of canonical or Hamiltonian equations. The Euler–Lagrange equations

$$\frac{d}{dx^\mu}\left(\frac{\partial \mathscr{L}}{\partial v^a_\mu}\right) = \partial_a \mathscr{L} \tag{2.21}$$

are equivalent to the canonical equations, the solutions of which we will call *extremals* of the variational problem.

For a more detailed discussion of the De Donder–Weyl canonical theory we refer to Refs. 1–3, 9, and 10.

3. SOME IMPORTANT PROPERTIES OF SOLUTIONS OF THE HJ EQUATION

In the following we will show how the knowledge of solutions of the HJ equation can simplify the problem of solving the field equations (2.21): First, we observe that Eq. (2.17) is *one* equation for *four* unknown functions S^μ; consequently, we have considerable freedom in the choice of the S^μ. There is, however, a restriction which does not occur in classical mechanics, because the functions φ^a_μ of Eqs. (2.18) must be subjected to certain integrability conditions. These conditions are the following:

$$\partial_\nu \varphi^a_\mu + \partial_b \varphi^a_\mu \cdot \varphi^b_\nu \overset{!}{=} \partial_\mu \varphi^a_\nu + \partial_b \varphi^a_\nu \cdot \varphi^b_\mu,$$

$$a = 1,\ldots, n, \quad \mu, \nu = 0,\ldots,3. \tag{3.1}$$

They impose stringent restrictions on the class of functions S^μ, which are of physical interest.

One possibility of removing the arbitrariness in the choice of the S^μ consists of the condition that a given extremal $z^a = f^a(x)$ be embedded weakly in a solution of the HJ equation. (The weak embedding is locally always possible; see Sec. 4.) This means requiring the so-called *transversality condition*

$$\pi^\mu_a(x) \overset{!}{=} \partial_a S^\mu(x, z)|_{z=f(x)}. \tag{3.2}$$

[$\pi^\mu_a(x)$ denotes the canonical momenta belonging to the solutions $z^a = f^a(x)$, according to Eq. (2.12).]

It can happen that a solution of the HJ equation (2.17) in which an extremal is weakly embedded does not fulfill conditions (3.1) for $z^a \neq f^a(x)$, i.e., beyond the extremal. If the relations (3.1) are valid in an open neighborhood of $(x^\mu, f^a(x))$, too, we will speak of a *strong embedding* of the extremal.

A physically important application of solutions of the HJ equation is the following: If the S^μ depend differentiably on a parameter α, i.e., $S^\mu = S^\mu(x, z; \alpha)$, then the functions

$$\tilde{G}^\mu(x, z; \alpha) := \frac{\partial S^\mu}{\partial \alpha}(x, z; \alpha) \tag{3.3}$$

will yield a conserved current in the case that we substitute z^b by such extremals $f^b(x)$ which fulfill $\pi^\mu_b(x) = \partial_b S^\mu|_{z=f(x)}$, i.e., we have to solve Eqs. (2.19) in principal. Therefore, the vector field

$$G^\mu(x, f(x); \alpha) := \tilde{G}^\mu(x, z; \alpha)|_{z=f(x)} \tag{3.4a}$$

has a vanishing divergence

$$\frac{dG^\mu}{dx^\mu}(x, f(x); \alpha) = 0. \tag{3.4b}$$

This can be proved as follows:

$$\frac{dG^\mu}{dx^\mu} = \frac{\partial \tilde{G}^\mu}{\partial x^\mu} + \frac{\partial \tilde{G}^\mu}{\partial z^b}\frac{\partial f^b}{\partial x^\mu}$$

$$\underset{(3.3)}{=} \frac{\partial^2 S^\mu}{\partial x^\mu \partial \alpha}\bigg|_{z=f(x)} + \frac{\partial^2 S^\mu}{\partial z^b \partial \alpha}\bigg|_{z=f(x)} \frac{\partial f^b}{\partial x^\mu}.$$

As S^μ is a solution of Eq. (2.17) and $f^b(x)$ an extremal, we obtain, in accordance with Eq. (2.19),

$$\frac{dG^\mu}{dx^\mu} = -\frac{\partial \mathscr{H}}{\partial \alpha}\left(x, z, \frac{\partial S}{\partial z}\right)\bigg|_{z=f(x)} + \frac{\partial \psi^\mu_b}{\partial \alpha}(x, z)\bigg|_{z=f(x)} \frac{\partial f^b}{\partial x^\mu}(x)$$

$$= \left[-\frac{\partial \mathscr{H}(x, z, \partial S/\partial z)}{\partial \pi^\mu_b}\bigg|_{z=f(x)} + \frac{\partial f^b}{\partial x^\mu}(x)\right] \frac{\partial \psi^\mu_b(x, z)}{\partial \alpha}\bigg|_{z=f(x)} = 0.$$

Noether's currents for classical field theories[11] are just a special case of the above HJ currents, for which the parameter α becomes the parameter of a transformation group.[7]

We next mention a method which allows us to calculate extremals from certain solutions of the HJ equation[3]: Suppose $S^\mu(x, z; \alpha^\nu_b)$ is a solution of Eq. (2.17) which depends differentiably on $4n$ parameters α^ν_b, fulfills the integrability conditions (3.1), and satisfies the inequality

$$\det\left(\frac{\partial^2 S^\mu}{\partial z^a \partial \alpha^\nu_b}\right) \neq 0. \tag{3.5}$$

Such a solution will be called "complete integral" of the HJ equation. We then can compute the functions

$$\tilde{G}^{\mu;b}_\nu(x, z; \alpha) := \frac{\partial S^\mu}{\partial \alpha^\nu_b}(x, z; \alpha),$$

$$\mu, \nu = 0,\ldots,3, \quad b = 1,\ldots, n. \tag{3.6}$$

Furthermore, we seek $16n$ functions $g^{\mu;b}_\nu(x)$ which fulfill the following two conditions:

1. The $g^{\mu;b}_\nu(x)$ have vanishing divergence, that is,

$$\partial_\mu g^{\mu;b}_\nu(x) = 0, \quad \nu = 0,\ldots,3, \quad b = 1,\ldots, n. \tag{3.7}$$

2. The $16n$ equations

$$\tilde{G}^{\mu;b}_\nu(x, z; \alpha) \overset{!}{=} g^{\mu;b}_\nu(x) \tag{3.8}$$

should be uniquely solvable for the n (dependent) variables z^a.

Then $15n$ of the $16n$ equations (3.8) necessarily must depend on the rest. The resulting functions $z^a = f^a(x)$ are extremals. We can prove this in the following way: If $z^a = f^a(x)$ is such a solution, then we have the relation [see Eq. (3.7)]

$$\frac{d\tilde{G}^{\mu;b}_\nu}{dx^\mu}(x, f(x); \alpha) \equiv \frac{dG^{\mu;b}_\nu}{dx^\mu}(x, f(x); \alpha) = 0.$$

As in the proof of Eq. (3.4), we obtain, on the other hand,

$$\frac{dG^{\mu;b}_{\nu}(x,f(x);\alpha)}{dx^{\mu}} = \left[\frac{\partial f^c}{\partial x^{\mu}}(x) - \frac{\partial \mathscr{H}(x,z,\partial S/\partial z)}{\partial \pi^{\mu}_c}\bigg|_{z=f(x)}\right] \times \frac{\partial^2 S^{\mu}}{\partial z^c \partial \alpha^{\nu}_b}(x,z;\alpha),$$

and, therefore, because of the inequality (3.5) we get

$$\frac{\partial f^c}{\partial x^{\mu}}(x) - \frac{\partial \mathscr{H}(x,z,\partial S/\partial z)}{\partial \pi^{\mu}_c}\bigg|_{z=f(x)} = 0.$$

Because of our assumption that the solution S^{μ} fulfills the integrability conditions (3.1), the last equations represent the first set of canonical equations [see (2.19)]. Because of Eqs. (2.19) the solutions $f^a(x)$ are extremals. Further applications of HJ theories can be found in Ref. 12.

4. TWO METHODS FOR SOLVING THE HJ EQUATION

A. Cauchy's method of characteristics

First we shall deal with the problem of embedding a given extremal $z^a = f^a(x)$ weakly in a solution of the HJ equation. For that purpose we use Cauchy's method of characteristics.[8,13,14] This approach in principle yields a method for the explicit construction of solutions of the HJ equation.

In order to remove part of the arbitrariness in the choice of the functions S^{μ}, we require that the functions S^j, $j = 1,2,3$, obey the transversality conditions (3.2). That can be achieved, for example, by the simple ansatz

$$S^j(x,z) = \pi^j_a(x)[z^a - f^a(x)], \quad j = 1,2,3. \tag{4.1}$$

The remaining function S^0 is determined by the HJ equation, which now takes the reduced form

$$\partial_0 S^0(x,z) + \hat{\mathscr{H}}(x,z,\pi^0_a = \partial_a S^0, \quad \pi^j_a = \partial_a S^j(x,z)) = 0 \tag{4.2a}$$

with

$$\hat{\mathscr{H}}\left(x,z,\pi = \frac{\partial S}{\partial z}\right) := \partial_j S^j(x,z) + \mathscr{H}(x,z,\pi^0_a = \partial_a S^0, \ \pi^j_a = \partial_a S^j). \tag{4.2b}$$

In addition, S^0 should satisfy the transversality condition (3.2), too, i.e.,

$$\pi^0_a(x) \overset{!}{=} \partial_a S^0(x,z)|_{z=f(x)}. \tag{4.2c}$$

Equation (4.2a) obviously is equivalent to a problem in mechanics with an in general explicitly time-dependent Hamilton function $\hat{\mathscr{H}}$. In Eq. (4.2a) the space variables $\mathbf{x}$ only appear as parameters which are kept fixed.

The differential equations for the characteristics associated with the pde (4.2a) are

$$\dot{x}^j(t) = 0, \quad j = 1,2,3, \tag{4.3a}$$

$$\dot{z}^a(t) = \frac{\partial \hat{\mathscr{H}}}{\partial \pi^0_a}, \quad \dot{\pi}^0_a(t) = -\frac{\partial \hat{\mathscr{H}}}{\partial z^a}. \tag{4.3b}$$

(Here we denote the differentiation with respect to t by an overdot.) If we employ the ansatz (4.1), the second of Eqs. (4.3b) takes the form $\dot{\pi}^0_a = -\partial_a \mathscr{H} - \partial_j \pi^j_a(x)$, in which the term $\partial_j \pi^j_a(x)$ plays the role of an additional, explicitly time-dependent force.

Equations (4.3a) imply: $\mathbf{x} = \mathbf{c} = \text{const}$, whereas Eqs. (4.3b) are the canonical equations of a mechanical system with the reduced Hamilton function $\hat{\mathscr{H}}$. Solutions of Eqs. (4.3b) can immediately be given[14]: The extremals $f^a(t,\mathbf{c})$ and $\pi^0_a(t,\mathbf{c})$ are solutions of Eqs. (4.3b) as functions of t and constant $\mathbf{c}$.

We determine solutions $z^a = \zeta^a(t)$ and $\pi^0_a = \chi_a(t)$ such that the following initial conditions are fulfilled:

$$\zeta^a(0) = u^a, \quad \chi_a(0) = \pi^0_a(0,\mathbf{c}), \quad a = 1,\dots,n. \tag{4.4}$$

Thus, the solutions depend on the n free parameters u^a: $\zeta^a = \zeta^a(t;u^a)$, $\chi_a = \chi_a(t;u^a)$. Due to the uniqueness theorem for the solutions of ordinary differential equations (ode's), we have

$$\zeta^a(t;u^a_0) = f^a(t,\mathbf{c}), \quad \chi_a(t;u^a_0) = \pi^0_a(t,\mathbf{c}), \qquad u^a_0 := f^a(0,\mathbf{c}). \tag{4.5}$$

Furthermore, the initial conditions (4.4) imply

$$\frac{\partial(\zeta^1,\dots,\zeta^n)}{\partial(u^1,\dots,u^n)} = 1 \neq 0 \quad \text{for } t = 0, \tag{4.6}$$

i.e., the equations $z^a = \zeta^a(t;u^a)$ are solvable for the u^a in an open neighborhood of $t = 0$.

Because of the initial conditions (4.4) a lengthy but simple calculation[14] shows that the differential form

$$\left[-\hat{\mathscr{H}}(x,\zeta,\chi) + \chi_a \frac{\partial \zeta^a}{\partial t}\right] dt + \chi_a \frac{\partial \zeta^a}{\partial u^b} du^b \tag{4.7}$$

is a total differential. Therefore, the line integral

$$\sigma(t;\mathbf{u}) := \int_{(0,\mathbf{u}_0)}^{(t,\mathbf{u})} \left\{\left[-\hat{\mathscr{H}}(x,\zeta,\chi) + \chi_a \frac{\partial \zeta^a}{\partial t}\right] dt + \chi_a \frac{\partial \zeta^a}{\partial u^b} du^b\right\} \tag{4.8}$$

[$\mathbf{u} := (u^1,\dots,u^n)$, $\mathbf{u}_0 := (u^1_0,\dots,u^n_0)$] is path-independent. After solving the equations $z^a = \zeta^a(t;u^b)$ for the u^a: $u^a = \gamma^a(t,z^b)$ and replacing the constant space coordinates $\mathbf{c}$ by the vector $\mathbf{x}$, we obtain the solution of the differential equation (4.2a),

$$S^0(x,z) := \sigma(x, u^a = \gamma^a(t,z)). \tag{4.9}$$

[(4.9) is indeed a solution of Eq. (4.2a) which can be verified with the help of the definitions (4.8) and (4.9) and the identity[14] $z^a = \zeta^a(t;\gamma)$.] We have

$$\partial_a S^0 = \frac{\partial \sigma}{\partial u^b}\frac{\partial \gamma^b}{\partial z^a} = \chi_c \frac{\partial \zeta^c}{\partial u^b}\frac{\partial \gamma^b}{\partial z^a} = \chi_c \delta^c_a = \chi_a,$$

which because of the relations (4.5) implies

$$\partial_a S^0(x,z)|_{z=f(x)} = \pi^0_a(x);$$

therefore, the transversality condition is satisfied, and the problem is solved. As a consequence, the weak embedding of an extremal $z^a = f^a(x)$ is (locally) always possible.

The construction of explicit solutions of Eq. (4.2a) essentially depends on the solvability of Eqs. (4.3b); solving this system of differential equations analytically will not be possible in general.

B. The "polynomial ansatz" technique

A more fruitful procedure for an explicit construction of solutions of the HJ equation represents the following "polynomial method."[2] Especially the derivation of conserved currents will be much easier.

For that purpose we assume that the variables z^a, v^a_μ, π^μ_a are real numbers (the complex case runs analogously). We start with the ansatz

$$S^j(x,z) := \pi^j_b(x)[z^b - f^b(x)], \quad j=1,2,3,$$

$$\begin{aligned} S^0(x,z) := & \int_0^{x^0} dt\, g(t,\mathbf{x}) + \pi^0_b(x)[z^b - f^b(x)] \\ & + \tfrac{1}{2} h_{bc}(x)[z^b - f^b(x)][z^c - f^c(x)] \\ & + (1/3!) l_{bcd}(x)[z^b - f^b(x)][z^c - f^c(x)] \\ & \times [z^d - f^d(x)] + \cdots, \end{aligned} \tag{4.10}$$

where $h_{bc}(x) = h_{cb}(x)$, $l_{bcd}(x) = l_{cbd}(x) = l_{cdb}(x), \cdots$. Again it will be our aim to embed a given extremal $f^a(x)$ with momenta $\pi^\mu_a(x)$ weakly in a solution of the HJ equation. We obviously achieve this aim through the ansatz (4.10), if we succeed in determining functions $g(x)$ and $h_{bc}(x)$, $l_{bcd}(x), \cdots$ such that (4.10) becomes a solution of the HJ equation. For this purpose the function

$$M(z;x) := -\mathscr{H}(x, z, \pi^\mu_a = \partial_a S^\mu(x,z)) \tag{4.11}$$

will now be expanded in a Taylor series with respect to the variables z^a around the extremal. The independent variables x are kept fixed as parameters. The result is [the asterisked quantities are to be taken on the extremal $z^a = f^a(x)$ and on their momenta $\pi^\mu_a(x)$]

$$\begin{aligned} M(z;x) = & -\mathscr{H}^* + \Big(-\frac{\partial \mathscr{H}^*}{\partial z^a} \\ & - \frac{\partial \mathscr{H}^*}{\partial \pi^0_b} \frac{\partial^2 S^0}{\partial z^b \partial z^a}\Big|_{z=f(x)} \Big)(z^a - f^a(x)) \\ & + \frac{1}{2}\Big(-\frac{\partial^2 \mathscr{H}^*}{\partial z^a \partial z^b} - \frac{\partial^2 \mathscr{H}^*}{\partial z^a \partial \pi^0_c} \frac{\partial^2 S^0}{\partial z^c \partial z^b}\Big|_{z=f(x)} \\ & - \frac{\partial^2 \mathscr{H}^*}{\partial \pi^0_c \partial z^b} \frac{\partial^2 S^0}{\partial z^c \partial z^a}\Big|_{z=f(x)} \\ & - \frac{\partial^2 \mathscr{H}^*}{\partial \pi^0_c \partial \pi^0_d} \frac{\partial^2 S^0}{\partial z^d \partial z^b}\Big|_{z=f(x)} \frac{\partial^2 S^0}{\partial z^c \partial z^a}\Big|_{z=f(x)} \\ & - \frac{\partial \mathscr{H}^*}{\partial \pi^0_c} \frac{\partial^3 S^0}{\partial z^c \partial z^a \partial z^b}\Big|_{z=f(x)} \Big) \\ & \times [z^a - f^a(x)][z^b - f^b(x)] + \cdots. \end{aligned} \tag{4.12}$$

Comparing the coefficients of the corresponding powers of $(z^a - f^a(x))$, we obtain from Eqs. (4.10) and (4.12)

$$g(x) = \pi^\mu_b(x) v^b_\mu(x) - \mathscr{H}^* = \mathscr{L}^*, \tag{4.13a}$$

$$\partial_\mu \pi^\mu_a(x) = -\partial_a \mathscr{H}^*, \quad a = 1,\dots,n, \tag{4.13b}$$

$$\begin{aligned} & \partial_0\Big(\frac{\partial^2 S^0}{\partial z^a \partial z^b}\Big|_{z=f(x)}\Big) + \Big(\frac{\partial^2 S^0}{\partial z^a \partial z^c}\Big|_{z=f(x)}\Big) \\ & \times \frac{\partial^2 \mathscr{H}^*}{\partial \pi^0_c \partial \pi^0_d}\Big(\frac{\partial^2 S^0}{\partial z^d \partial z^b}\Big|_{z=f(x)}\Big) \\ & + \frac{\partial^2 \mathscr{H}^*}{\partial z^a \partial \pi^0_c}\Big(\frac{\partial^2 S^0}{\partial z^c \partial z^b}\Big|_{z=f(x)}\Big) \\ & + \frac{\partial^2 \mathscr{H}^*}{\partial z^b \partial \pi^0_c}\Big(\frac{\partial^2 S^0}{\partial z^c \partial z^a}\Big|_{z=f(x)}\Big) \\ & + \frac{\partial^2 \mathscr{H}^*}{\partial z^a \partial z^b} = 0, \quad a, b = 1,\dots, n. \end{aligned} \tag{4.13c}$$

Obviously, Eqs. (4.13b) are just the second set of the canonical equations (2.20). Equation (4.13c) is a matrix differential equation of the first order for the functions $h_{ab}(x)$. We want to write this relation somewhat more compactly. For this purpose we introduce the following abbreviations:

$$\begin{aligned} & S := \Big(\frac{\partial^2 S^0}{\partial z^a \partial z^b}\Big|_{z=f(x)}\Big), \quad \mathscr{H}_{\pi^0\pi^0} := \Big(\frac{\partial^2 \mathscr{H}^*}{\partial \pi^0_c \partial \pi^0_d}\Big), \\ & \mathscr{H}_{z\pi^0} := \Big(\frac{\partial^2 \mathscr{H}^*}{\partial z^a \partial \pi^0_c}\Big), \quad \mathscr{H}_{zz} := \Big(\frac{\partial^2 \mathscr{H}^*}{\partial z^a \partial z^b}\Big), \\ & \mathscr{H}_{\pi^0 z} := \mathscr{H}^T_{z\pi^0}. \end{aligned} \tag{4.14}$$

The quantities (4.14) are real $n \times n$ matrices of which S, $\mathscr{H}_{\pi^0\pi^0}$, and $\mathscr{H}_{zz}$ are symmetric. With the help of this notation, Eqs. (4.13c) assume the form

$$\dot{S} + S\mathscr{H}_{\pi^0\pi^0}S + \mathscr{H}_{z\pi^0}S + S\mathscr{H}_{\pi^0 z} + \mathscr{H}_{zz} = 0. \tag{4.15}$$

Equation (4.15) is a matrix Riccati differential equation[15] with the independent variable t ($\mathbf{x}$ appears only as a parameter) whose coefficients $\mathscr{H}_{\pi^0\pi^0}$, $\mathscr{H}_{z\pi^0}$, and $\mathscr{H}_{zz}$ will in general be explicitly t-dependent. Equation (4.15) can be associated with a linear system of matrix differential equations

$$\begin{bmatrix} \dot{U} \\ \dot{V} \end{bmatrix} = \begin{bmatrix} \mathscr{H}_{\pi^0 z} & \mathscr{H}_{\pi^0\pi^0} \\ -\mathscr{H}_{zz} & -\mathscr{H}_{z\pi^0} \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix}. \tag{4.16}$$

(U and V are t-dependent $n \times n$ matrices.) The connection between the solutions of Eqs. (4.16) and (4.15) is the following: If $U(t)$ and $V(t)$ are solutions of Eqs. (4.16) and if $U(t)$ is nonsingular, then $S(t) := V(t)U^{-1}(t)$ will be a solution of Eq. (4.15); on the other hand, if a solution of (4.15) can be written in the form $S(t) = V(t)U^{-1}(t)$, then U and V will solve (4.16). [It can be proven that the general solution of Eq. (4.15) can be expressed through *four* independent special solutions of the same equation.[16]]

Equations of the type (4.16) we also find in the theory of the second variation in which necessary and sufficient criteria are developed for a solution $z^a = f^a(x)$ of the Euler–Lagrange equations to minimize the action integral (2.2). A necessary condition for this to be the case is that the following quadratic form $Q(\xi)$ is positive semidefinite, that is,

$$\frac{1}{2} Q(\xi) := \int_V d^4x\, L\Big(x, \xi(x), \frac{\partial \xi}{\partial x}(x)\Big) \geqslant 0, \tag{4.17}$$

in which

$$L(x,\zeta^a,\omega^a_\mu):=\frac{1}{2}\left(\frac{\partial^2\mathscr{L}^*}{\partial z^a\partial z^b}\zeta^a\zeta^b+2\frac{\partial^2\mathscr{L}^*}{\partial z^a\partial v^b_\mu}\zeta^a\omega^b_\mu+\frac{\partial^2\mathscr{L}^*}{\partial v^a_\mu\partial v^b_\nu}\omega^a_\mu\omega^b_\nu\right). \tag{4.18}$$

The functions $\zeta^a(x)$ are assumed to obey the normalization condition: $\|\zeta\|_{\mathscr{L}^2(V)}=1$, where $\mathscr{L}^2(V)$ denotes the Hilbert space of the absolutely square integrable functions on V. [If $Q(\zeta)$ in Eq. (4.17) is even positive definite, then this property will also be a sufficient condition for the existence of a minimum.] We try to solve this sign problem through an analogous procedure as in the first variation. This leads to the so-called "accessory variational problem"[17,18]: We deduce the Euler–Lagrange equations for the Lagrange function

$$L^0:=L-\frac{1}{2}\sum_{a=1}^{n}\lambda\zeta^a\zeta^a \tag{4.19}$$

(λ is a Lagrangian multiplier):

$$\frac{d}{dx^\mu}\left(\frac{\partial L^0}{\partial\omega^a_\mu}\right)-\frac{\partial L^0}{\partial\zeta^a}=0. \tag{4.20}$$

For $\lambda=0$ Eqs. (4.20) are called *Jacobi equations* of the second variation. Equations (4.20) constitute a linear eigenvalue problem from which we will get the solution of the sign problem (4.17), if the sign of the smallest eigenvalue λ can be determined. (For a more detailed discussion see Ref. 18.)

We calculate the canonical momenta and the Hamilton function for the Lagrange function (4.18) by a Legendre transformation; this gives[19]

$$\eta^\mu_a:=\frac{\partial L}{\partial\omega^a_\mu}=\frac{\partial^2\mathscr{L}^*}{\partial v^a_\mu\partial z^b}\zeta^b+\frac{\partial^2\mathscr{L}^*}{\partial v^a_\mu\partial v^b_\nu}\omega^b_\nu, \tag{4.21a}$$

$$H(x,\zeta,\eta):=\frac{1}{2}\left(\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial z^b}\zeta^a\zeta^b+2\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial\pi^\mu_b}\zeta^a\eta^\mu_b+\frac{\partial^2\mathscr{H}^*}{\partial\pi^\mu_a\partial\pi^\nu_b}\eta^\mu_a\eta^\nu_b\right). \tag{4.21b}$$

Jacobi's equations of the second variation are equivalent to the canonical equations

$$\omega^a_\mu=\frac{\partial H}{\partial\eta^\mu_a}=\frac{\partial^2\mathscr{H}^*}{\partial\pi^\mu_a\partial z^b}\zeta^b+\frac{\partial^2\mathscr{H}^*}{\partial\pi^\mu_a\partial\pi^\nu_b}\eta^\nu_b, \tag{4.22a}$$

$$\partial_\mu\eta^\mu_a=-\frac{\partial H}{\partial\zeta^a}=-\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial z^b}\zeta^b-\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial\pi^\nu_b}\eta^\nu_b. \tag{4.22b}$$

We easily find extremals to the Lagrange function (4.18), i.e., solutions of Jacobi's equations: If $z^a=f^a(x;\alpha)$ is an extremal of the first variation which depends on a parameter α, then

$$\zeta^a=T^a(x):=\frac{\partial f^a}{\partial\alpha}(x;\alpha)$$

will be a solution of Jacobi's equations (see, for example, Ref. 18). It is a trivial fact that $T^a(x)\equiv 0$, also, is an extremal of the second variation with $\|T^a\|=0$.

For the Hamilton function (4.21b) we now can set up the HJ equation

$$\partial_\mu W^\mu(x,\zeta)+H\left(x,\zeta,\eta=\frac{\partial W}{\partial\zeta}\right)=0. \tag{4.23}$$

We want to embed a given solution $\zeta^a=T^a(x)$ of Jacobi's equations weakly in a solution of the HJ equation (4.23): If $\eta^\mu_a(x)$ are the canonical momenta, we will make the ansatz

$$W^j(x,\zeta):=\eta^j_a(x)[\zeta^a-T^a(x)],\quad j=1,2,3. \tag{4.24}$$

From this we deduce the reduced HJ equation [see (4.2)]:

$$\partial_0W^0(x,\zeta)+\hat{H}\left(x,\ \zeta,\ \eta^0_a=\frac{\partial W^0}{\partial\zeta^a},\ \eta^j_a=\eta^j_a(x)\right)=0, \tag{4.25a}$$

in which

$$\hat{H}(x,\zeta,\eta):=\partial_j\eta^j_a(x)[\zeta^a-T^a(x)]-\eta^j_a(x)\partial_jT^a(x)+H\left(x,\ \zeta,\ \eta^0_a=\frac{\partial W^0}{\partial\zeta^a},\ \eta^j_a=\eta^j_a(x)\right). \tag{4.25b}$$

In accordance with (4.3b) we obtain the differential equations for the characteristics:

$$\begin{aligned}\dot{\zeta}^a(t)&=\frac{\partial\hat{H}}{\partial\eta^0_a}=\frac{\partial^2\mathscr{H}^*}{\partial z^b\partial\pi^0_a}\zeta^b(t)+\frac{\partial^2\mathscr{H}^*}{\partial\pi^0_a\partial\pi^0_b}\eta^0_b(t)+\frac{\partial^2\mathscr{H}^*}{\partial\pi^0_a\partial\pi^j_b}\eta^j_b(x),\\ \dot{\eta}^0_a(t)&=-\frac{\partial\hat{H}}{\partial\zeta^a}=-\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial z^b}\zeta^b(t)-\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial\pi^0_b}\eta^0_b(t)-\frac{\partial^2\mathscr{H}^*}{\partial z^a\partial\pi^j_b}\eta^j_b(x)-\partial_j\eta^j_a(x).\end{aligned} \tag{4.26}$$

Equations (4.26) form an inhomogeneous, linear system of differential equations with varying coefficients. The homogeneous part of Eqs. (4.26) is the same as Eqs. (4.16). We automatically get this homogeneous part, if we embed the trivial extremal $T^a(x)\equiv 0$ with $\eta^\mu_a(x)=0$. Therefore, we find the following connection between Eq. (4.15) and the HJ equation (4.23):

If $S(x)$ is a symmetric $n\times n$ matrix, then

$$W^0(x,\zeta):=\tfrac{1}{2}\zeta^TS(x)\zeta,\quad \zeta^T:=(\zeta^1,\dots,\zeta^n)$$

will be a solution of Eq. (4.25a) for the extremal $T^a(x)\equiv 0$ iff $S(x)$ is a solution of (4.15). [In general we easily show that if $S^\mu(x,z)$ is a solution of the HJ equation (2.17) in which an extremal $z^a=f^a(x)$ is weakly embedded, then

$$W^\mu(x,\zeta):=\frac{1}{2}\zeta^a\zeta^b\left.\frac{\partial^2S^\mu(x,z)}{\partial z^a\partial z^b}\right|_{z=f(x)},\quad \mu=0,\dots,3$$

solves Eq. (4.23).]

Summarizing, we can draw the following conclusion: The "polynomial method" will be appropriate, if the Hamilton function $\mathscr{H}$ is a polynomial up to the second order in the variables π^μ_a and z^a; for in this case the solution of Eq. (2.17) is given by a polynomial ansatz for S^0 up to the second order [see Eq. (4.10)]. The only problem left consists of solving the corresponding matrix Riccati equation (4.15). But even the physically more interesting cases in which $\mathscr{H}$ is of higher order in the variables z^a can be treated with the above method to some extent (see Sec. 5A).

5. EXAMPLES

With some examples from relativistic field theory we now want to illustrate the methods developed above.

A. Solutions of the HJ equation

First we construct some explicit solutions of the HJ equation:

(i) The simplest example from field theory is the real, scalar Klein–Gordon field. It has the Lagrange function (LF)

$$\mathscr{L} = \tfrac{1}{2} v^{\mu} v_{\mu} - \tfrac{1}{2} m^2 z^2 \quad (n = 1!) \tag{5.1}$$

in which z and v_μ become

$$z \to z(x), \quad v_\mu \to v_\mu(x) = \partial_\mu z(x)$$

on the submanifolds Σ_4 [see Eq. (2.1)]. The canonical momenta are

$$\pi^{\mu} := \frac{\partial \mathscr{L}}{\partial v_\mu} = v^{\mu}.$$

We see immediately that the Legendre transformation (LT) is regular [see Eq. (2.13)]. Therefore, we can calculate the Hamilton function (HF)

$$\mathscr{H} = \pi^{\mu} v_\mu - \mathscr{L} = \tfrac{1}{2} \pi^{\mu} \pi_\mu + \tfrac{1}{2} m^2 z^2 .$$

For the slope functions $\varphi_\mu(x, z)$ we obtain here

$$\varphi_\mu(x, z) := \left.\frac{\partial \mathscr{H}}{\partial \pi^{\mu}}\right|_{\pi = \partial S/\partial z} = g_{\mu\nu} \frac{\partial S^{\nu}}{\partial z}(x, z)$$

($g_{\mu\nu}$ = Minkowski metric tensor), and the integrability conditions (3.1) take the form

$$g_{\mu\rho} \frac{\partial^2 S^{\rho}}{\partial x^{\nu} \partial z} + g_{\mu\rho}\, g_{\nu\sigma} \frac{\partial^2 S^{\rho}}{\partial z^2} \frac{\partial S^{\sigma}}{\partial z} \overset{!}{=} g_{\nu\rho} \frac{\partial^2 S^{\rho}}{\partial x^{\mu} \partial z} + g_{\nu\rho}\, g_{\mu\sigma} \frac{\partial^2 S^{\rho}}{\partial z^2} \frac{\partial S^{\sigma}}{\partial z}. \tag{5.2}$$

These are nonlinear pde's of the second order.

If we want to embed a given extremal $z = \varphi(x)$ with momenta $\pi^{\mu}(x) = \partial^{\mu}\varphi(x)$ weakly in a solution of the HJ equation, the ansatz (4.10) will lead to

$$S^{\mu} = g^{\mu 0} \int_0^{x^0} dt\, \mathscr{L}^*(t, \mathbf{x}) + \pi^{\mu}(x)[z - \varphi(x)] + g^{\mu 0} h(x)[z - \varphi(x)]^2 .$$

$h(x)$ here has to satisfy the Riccati equation

$$\dot{h} = -2h^2 - \tfrac{1}{2} m^2 ,$$

the general solution of which is

$$h(x) = \tfrac{1}{2} m \tan[-m x^0 + F(\mathbf{x})] .$$

The integrability conditions (5.2) here give

$$\partial_j h(x)[z - \varphi(x)] = 0, \quad j = 1,2,3 .$$

These conditions are fulfilled either on the extremals (that is always true) or if $\partial_j h(x) = 0$, $j = 1,2,3$. The result shows that solutions of the HJ equation exist in which a given extremal can be embedded weakly but not strongly.

A symmetric form of the ansatz (4.10) is possible, too; for instance,

$$S^{\mu}(x, z) = \partial^{\mu} f(x) + g^{\mu}(x) z + \tfrac{1}{2} h^{\mu}(x) z^2 . \tag{5.3}$$

The coefficient functions are determined by comparing the coefficients of z; this gives

$$\partial_\mu h^{\mu} = -h^{\mu} h_\mu - m^2 , \tag{5.4a}$$

$$g^{\mu}(x) = \pi^{\mu}(x) - 2h^{\mu}(x)\varphi(x) , \tag{5.4b}$$

$$\Box f = -\tfrac{1}{2} g^{\mu}(x) g_\mu(x) \quad (\Box := \partial_\mu \partial^{\mu}) . \tag{5.4c}$$

A separation ansatz yields special solutions of Eq. (5.4a) (see Sec. 5C). If the functions $h^{\mu}(x)$ are known, Eq. (5.4b) defines the functions $g^{\mu}(x)$. This definition of the functions $g^{\mu}(x)$ ensures the validity of the transversality condition (3.2). For the ansatz (5.3) we obtain the integrability conditions (3.1),

$$[\partial_\nu h_\mu(x) - \partial_\mu h_\nu(x)][z - \varphi(x)] = 0 ,$$

$$\nu, \mu = 0,\dots,3, \ \mu \neq \nu .$$

(ii) The calculations for the complex (charged), scalar field are completely analogous: Here we have the LF

$$\mathscr{L} = \bar{v}_\mu v^{\mu} - m^2 \bar{z} z \quad (n = 2!)$$

(the overbar denotes the complex conjugation), from which we find the canonical momenta and the HF

$$\pi^{\mu} := \frac{\partial \mathscr{L}}{\partial v_\mu} = \bar{v}^{\mu}, \quad \bar{\pi}^{\mu} := \frac{\partial \mathscr{L}}{\partial \bar{v}_\mu} = v^{\mu} ,$$

$$\mathscr{H} = \pi^{\mu} v_\mu + \bar{\pi}^{\mu} \bar{v}_\mu - \mathscr{L} = \pi_\mu \bar{\pi}^{\mu} + m^2 \bar{z} z .$$

With the help of Eq. (4.10) we get the following solution of the HJ equation:

$$S^{\mu}(x, z, \bar{z}) = g^{\mu 0} \int_0^{x^0} dt\, \mathscr{L}^*(t,\mathbf{x}) + \pi^{\mu}(x)[z - \varphi(x)] + \bar{\pi}^{\mu}(x)[\bar{z} - \overline{\varphi}(x)] + g^{\mu 0} h(x) \times [\bar{z} - \overline{\varphi}(x)][z - \varphi(x)] ,$$

where the function $h(x)$ is a solution of the Riccati equation

$$\dot{h} = -h^2 - m^2 , \tag{5.5}$$

with the general solution

$$h(x) = m \tan[-m x^0 + F(\mathbf{x})] . \tag{5.6}$$

(iii) (a) In the case of E-dynamics without external current we start with the LF

$$\mathscr{L} = -\tfrac{1}{4} f_{\mu\nu} f^{\mu\nu}, \quad f_{\mu\nu} := v_{\mu\nu} - v_{\nu\mu} ,$$

where $v_{\mu\nu} \to v_{\mu\nu}(x) := g_{\rho\nu} \partial_\mu A^{\rho}(x)$ on the extremals $A^{\rho}(x)$. The canonical momenta are

$$\pi^{\mu\nu} := \frac{\partial \mathscr{L}}{\partial v_{\mu\nu}} = -f^{\mu\nu} .$$

It is clear that the LT is not regular [see Eq. (2.13)]. Nevertheless, it is possible to calculate the HF by means of the relation $\pi^{\mu\nu} v_{\mu\nu} = \tfrac{1}{2} \pi^{\mu\nu}(v_{\mu\nu} - v_{\nu\mu})$. The result is

$$\mathscr{H} = \pi^{\mu\nu} v_{\mu\nu} - \mathscr{L} = -\tfrac{1}{4} \pi^{\mu\nu} \pi_{\mu\nu} .$$

Although the LT is singular, a linear ansatz of the type (4.10) leads to the solution of the weak embedding problem: If $A_\mu(x)$ and $\pi^{\mu\nu}(x)$ are given extremals and associated momenta, then

$$S^{\mu}(x, z_\nu) = g^{\mu 0} \int_0^{x^0} dt\, \mathscr{L}^*(t,\mathbf{x}) + \pi^{\mu\nu}(x)[z_\nu - A_\nu(x)] \tag{5.7}$$

will solve the HJ equation

$$\partial_\mu S^{\mu} - \frac{1}{4}\frac{\partial S^{\mu}}{\partial z_\nu}\frac{\partial S_\mu}{\partial z^\nu} = 0 .$$

There is no integrability problem here, because the canonical equations (2.19) and (2.20) are not equivalent to the Euler–Lagrange equations, for the LT is singular.

A symmetric solution of the HJ equation is[20]

$$S^{\mu}(x, z_\nu) = \pi^{\mu\nu}(x)[z_\nu - \tfrac{1}{2}A_\nu(x)] . \tag{5.8}$$

(b) We next examine E-dynamics with external current $j^\mu(x)$ with the LF

$$\mathscr{L} = -\tfrac{1}{4}f_{\mu\nu}f^{\mu\nu} - j^\nu(x)z_\nu , \tag{5.9}$$

which gives the canonical momenta

$$\pi^{\mu\nu} := \frac{\partial \mathscr{L}}{\partial v_{\mu\nu}} = -f^{\mu\nu} .$$

Again we calculate the HF, despite the singularity of the LT,

$$\mathscr{H} = -\tfrac{1}{4}\pi^{\mu\nu}\pi_{\mu\nu} + j^\nu(x)z_\nu .$$

We will find a solution of the HJ equation of the type (5.7), if we take for $\mathscr{L}^*$ the Lagrangian (5.9).

A symmetric solution is the following[20]:

$$S^{\mu}(x, z_\nu) = \pi^{\mu\nu}(x)[z_\nu - \tfrac{1}{2}A_\nu(x)] - \tfrac{1}{2}\partial^\mu g(x) , \tag{5.10}$$

in which $g(x)$ has to satisfy the equation

$$\Box g = j^\mu(x)A_\mu(x) .$$

In order to get a regular LT, we take the LF in the Lorentz gauge $\partial_\mu A^{\mu}(x) = 0$:

$$\mathscr{L} = -\tfrac{1}{2}v_{\mu\nu}v^{\mu\nu} - j^{\nu}(x)z_\nu \tag{5.11}$$

with the canonical momenta

$$\pi^{\mu\nu} := \frac{\partial \mathscr{L}}{\partial v_{\mu\nu}} = -v^{\mu\nu} .$$

The LT is obviously regular, and we obtain the HF

$$\mathscr{H} = \pi^{\mu\nu}v_{\mu\nu} - \mathscr{L} = -\tfrac{1}{2}\pi^{\mu\nu}\pi_{\mu\nu} + j^\nu(x)z_\nu .$$

This time we want to construct a solution of the HJ equation by the method of characteristics. For that purpose we set S^j, $j = 1,2,3$:

$$S^j(x, z^\nu) = -\partial^j A^{\nu}(x)[z_\nu - A_\nu(x)] ,$$

in which $A_\nu(x)$ are the given extremals with momenta $\pi_{\mu\nu}(x) = -\partial_\mu A_\nu(x)$.

According to the procedure of Sec. 4 we have to calculate the characteristics which belong to the reduced HJ equation (4.2) [see Eq. (4.3)] from the equations

$$\dot z^\nu = \frac{\partial \hat{\mathscr{H}}}{\partial \pi^0_\nu} = -g^{\nu\mu}\pi^0_\mu , \tag{5.12a}$$

$$\dot\pi^0_\nu = -\frac{\partial \hat{\mathscr{H}}}{\partial z^\nu} = -\Delta A_\nu(x^0,\mathbf{c}) - j_\nu(x^0,\mathbf{c}) = -\frac{\partial^2 A_\nu}{\partial t^2}(x^0,\mathbf{c}) . \tag{5.12b}$$

The last equality in Eq. (5.12b) is valid because of the field equations $\Box A_\nu(x) = j_\nu(x)$. Solutions of these equations which satisfy the condition (4.5) are the following:

$$z^\nu(t) = A^{\nu}(t,\mathbf{c}) - A^{\nu}(0,\mathbf{c}) + u^\nu, \quad u^\nu = \text{const}, \tag{5.13a}$$

$$\pi^0_\nu(t) = -\frac{\partial A_\nu}{\partial t}(t,\mathbf{c}) . \tag{5.13b}$$

The calculation of the line integral (4.8) and solving Eqs. (5.13a) for the constants u^ν are simple here, and we get

$$S^0(x, z^\nu) = \int_0^{x^0} dt\, \mathscr{L}^*(t,\mathbf{x}) - \partial^0 A_\nu(x)[z^\nu - A^{\nu}(x)] .$$

For these solutions S^{μ} the integrability conditions (3.1) are fulfilled, for we have in this case

$$\varphi_{\mu\nu}(x, z) := \left.\frac{\partial \mathscr{H}}{\partial \pi^{\mu\nu}}\right|_{\pi = \partial S/\partial z} = -g_{\mu\sigma}\frac{\partial S^\sigma}{\partial z^\nu} = g_{\mu\sigma}\partial^\sigma A_\nu(x) .$$

(iv) Next we want to couple a complex, scalar field to an external electromagnetic field $A_\mu(x)$. The LF is

$$\mathscr{L} = (\bar v_\mu - iqA_\mu(x)\bar z)(v^\mu + iqA^{\mu}(x)z) - m^2\bar z z . \tag{5.14}$$

On Σ_4 [see Eq. (2.1)] this LF is gauge-invariant in the following sense:

$$A^{\mu}(x) \to A^{\mu}(x) + \partial^\mu f(x) ,$$

$$z(x) \to e^{-iqf(x)}z(x), \quad \bar z(x) \to e^{iqf(x)}\bar z(x) .$$

We get the canonical momenta

$$\pi^\mu := \frac{\partial \mathscr{L}}{\partial v_\mu} = \bar v^\mu - iqA^{\mu}(x)\bar z,$$

$$\bar\pi^\mu := \frac{\partial \mathscr{L}}{\partial \bar v_\mu} = v^\mu + iqA^{\mu}(x)z .$$

Therefore, the LT is regular, and the HF is

$$\mathscr{H} = \pi^\mu\bar\pi_\mu - iqA^{\mu}(x)\pi_\mu z + iqA^{\mu}(x)\bar\pi_\mu\bar z + m^2\bar z z .$$

The weak embedding corresponding to Eq. (4.10) gives for extremals $\varphi(x)$, $\bar\varphi(x)$ with momenta $\pi^\mu(x)$, $\bar\pi^\mu(x)$

$$\begin{aligned} S^{\mu}(x, z, \bar z) = g^{\mu 0}\int_0^{x^0} dt\, \mathscr{L}^*(t,\mathbf{x}) + \pi^\mu(x)[z - \varphi(x)] \\ + \bar\pi^\mu(x)[\bar z - \bar\varphi(x)] + g^{\mu 0}h(x)[z - \varphi(x)] \\ \times[\bar z - \bar\varphi(x)] , \end{aligned} \tag{5.15}$$

in which $h(x)$ satisfies Eq. (5.5) with the general solution (5.6). We can try here a symmetric ansatz analogous to (5.3), too:

$$S^{\mu}(x, z, \bar z) = \partial^\mu f(x) + g^{\mu}_1(x)z + g^{\mu}_2(x)\bar z + h^{\mu}(x)\bar z z ,$$

with the following equations for the coefficients:

$$\partial_\mu h^{\mu} = -h^{\mu}h_\mu - m^2 , \tag{5.16a}$$

$$g^{\mu}_1(x) = \pi^\mu(x) - h^{\mu}(x)\bar\varphi(x) , \tag{5.16b}$$

$$g^{\mu}_2(x) = \bar\pi^\mu(x) - h^{\mu}(x)\varphi(x) , \tag{5.16c}$$

$$\Box f = -g^{\mu}_1(x)g_{2,\mu}(x) . \tag{5.16d}$$

Special solutions of Eq. (5.16a) are, for instance,

$$h^0(x) = \tfrac{1}{2}m\tan(-\tfrac{1}{2}mx^0 + c^0),$$

$$h^j(x) = \tfrac{1}{2}m\tanh(-\tfrac{1}{2}mx^j + c^j),$$

$$j = 1,2,3, \quad c^\mu = \text{const}.$$

The integrability conditions lead to the constraints

$$[\partial_0 A_j(x) - \partial_j A_0(x) - (i/q)\partial_j h(x)][z - \varphi(x)] = 0, \quad j = 1,2,3 ,$$

$$[\partial_j A_k(x) - \partial_k A_j(x)][z - \varphi(x)] = 0, \quad j,k = 1,2,3,\ j \neq k ,$$

and analogous equations for $\bar z - \bar\varphi(x)$.

(v) The example (iv) will now be extended by coupling the electromagnetic field to the matter fields $\varphi(x)$, $\overline{\varphi}(x)$, too. Instead of (5.14) we now have

$$\mathscr{L} = \overline{v}_\mu v^\mu - iq\, y^j \overline{z} v_j + iq\, y^j z \overline{v}_j + q^2 y_j\, y^j \overline{z} z - m^2 \overline{z} z - \tfrac{1}{4} f^{\mu\nu} f_{\mu\nu}\,, \tag{5.17}$$

in which $y^j \to A^j(x)$ on the extremals and $f_{\mu\nu} := v_{\mu\nu} - v_{\nu\mu}$, $v_{\mu 0} = v^{\mu 0} := 0$. The function (5.17) is the LF in the temporal gauge $A_0 = 0$. We can get the complete set of field equations in the temporal gauge from this LF, if we require as a secondary condition the conservation of the current[21]:

$$\partial_\mu j^\mu(x) = 0,\quad j^0(x) = iq[\overline{\varphi}(x)\,\partial^0 \varphi(x) - \varphi(x)\,\partial^0 \overline{\varphi}(x)]\,,$$
$$j^k(x) = iq[\overline{\varphi}(x)\,\partial^k \varphi(x) - \varphi(x)\,\partial^k \overline{\varphi}(x)] - 2q^2\, \overline{\varphi}(x) \varphi(x) A^k(x),\quad k = 1,2,3.$$

The above LF then leads to the complete theory in the $A_0 = 0$ – gauge. For the momenta we get

$$\pi^0 := \frac{\partial \mathscr{L}}{\partial v_0} = \overline{v}^{\,0},\quad \overline{\pi}^0 := \frac{\partial \mathscr{L}}{\partial \overline{v}_0} = v^0\,,$$
$$\pi^j := \frac{\partial \mathscr{L}}{\partial v_j} = \overline{v}^j - iq\, y^j \overline{z}\,,$$
$$\overline{\pi}^j := \frac{\partial \mathscr{L}}{\partial \overline{v}_j} = v^j + iq\, y^j z,\quad j = 1,2,3,$$
$$\pi^{\mu j} := \frac{\partial \mathscr{L}}{\partial v_{\mu j}} = -f^{\mu j},\quad \mu = 0,\ldots,3,\quad j = 1,2,3.$$

It is true that the LT is only regular with respect to $v^0,\ldots, v^3$, $\overline{v}^{\,0},\ldots, \overline{v}^3$, and v_{0j} but, nevertheless, we can calculate the HF

$$\mathscr{H} = \pi^\mu \overline{\pi}_\mu - iq\pi^j y_j z + iq\overline{\pi}^j y_j \overline{z} - \tfrac{1}{4}\pi^{kj}\pi_{kj} - \tfrac{1}{2}\pi^{0j}\pi_{0j} + m^2 \overline{z} z\,.$$

The ansatz (4.10) is again successful in this case, because only the regularity of the LT with respect to the zero components of the momenta is important. If $\overline{\varphi}(x)$, $\varphi(x)$, and $A^j(x)$ are extremals which are to be embedded, then the solution takes the form

$$\begin{aligned} S^\mu(x, z, \overline{z}, z_j) = {} & g^{\mu 0} \int_0^{x^0} dt\, \mathscr{L}^*(t,\mathbf{x}) + \pi^\mu(x)[z - \varphi(x)] \\ & + \overline{\pi}^\mu(x)[\overline{z} - \overline{\varphi}(x)] + \pi^{\mu k}(x)[\,y_k - A_k(x)] \\ & + h_{(1)}(x)[\overline{z} - \overline{\varphi}(x)]\,[z - \varphi(x)] \\ & + h^{\,k}_{(2)}(x)[z - \varphi(x)]\,[\,y_k - A_k(x)] \\ & + \overline{h}^{\,k}_{(2)}(x)[\overline{z} - \overline{\varphi}(x)]\,[\,y_k - A_k(x)] \\ & + \tfrac{1}{2} h^{\,kl}_{(3)}(x)[\,y_k - A_k(x)]\,[\,y_l - A_l(x)] \\ & + \tfrac{1}{2} h_{(4)}(x)[z - \varphi(x)]^2 + \tfrac{1}{2}\overline{h}_{(4)}(x)[\overline{z} - \overline{\varphi}(x)]^2\,, \end{aligned}$$

where $h^{\,kl}_{(3)}(x) = h^{\,lk}_{(3)}(x)$. The functions h can be ordered in a Hermitian 5×5 matrix

$$S := \begin{pmatrix} h_{(1)} & h_{(4)} & h^{\,1}_{(2)} & h^{\,2}_{(2)} & h^{\,3}_{(2)} \\ \overline{h}_{(4)} & h_{(1)} & \overline{h}^{\,1}_{(2)} & \overline{h}^{\,2}_{(2)} & \overline{h}^{\,3}_{(3)} \\ \overline{h}^{\,1}_{(2)} & h^{\,1}_{(2)} & \ddots & & \cdot^{\cdot^{\cdot}} \\ \overline{h}^{\,2}_{(2)} & h^{\,2}_{(2)} & & h^{\,kl}_{(3)} & \\ \overline{h}^{\,3}_{(2)} & h^{\,3}_{(2)} & \cdot^{\cdot^{\cdot}} & & \ddots \end{pmatrix}.$$

For the (Hermitian) matrix $\mathscr{H}_{zz}$ from Eq. (4.14), we find

$$\mathscr{H}_{zz} = \begin{pmatrix} m^2 & 0 & -iq\pi^1(x) & -iq\pi^2(x) & -iq\pi^3(x) \\ 0 & m^2 & iq\overline{\pi}^1(x) & iq\overline{\pi}^2(x) & iq\overline{\pi}^3(x) \\ iq\overline{\pi}^1(x) & -iq\pi^1(x) & & & \\ iq\overline{\pi}^2(x) & -iq\pi^2(x) & & 0 & \\ iq\overline{\pi}^3(x) & -iq\pi^3(x) & & & \end{pmatrix}.$$

We proved already in Eq. (4.15) that S fulfills the matrix Riccati equation

$$\dot{S} + S^2 + \mathscr{H}_{zz} = 0\quad (\mathscr{H}_{\pi^0\pi^0} = E_5!)\,.$$

It seems very difficult to find analytical solutions of this equation because of the explicit t dependence of $\mathscr{H}_{zz}$; in simple special cases a solution will be possible [for example, if $\varphi(x)$, $\overline{\varphi}(x)$, and $A^j(x)$ are independent of time].

(vi) We will now treat the free Dirac equation using once more the method of characteristics. At first we want to derive an LF which yields a regular LT. For that purpose we start with the common LF

$$\mathscr{L} = \tfrac{1}{2} i(\overline{z}\gamma^\mu v_\mu - \overline{v}_\mu \gamma^\mu z) - m\overline{z} z\,, \tag{5.18}$$

in which z is a complex, four-component column vector and $\overline{z} := z^+\gamma^0$ is the adjoint row vector. Obviously, the LT with respect to the LF (5.18) is not regular. To change this fact, we add to the expression (5.18) a total divergence of the form[22]

$$\overline{v}_\mu \sigma^{\mu\nu} v_\nu = \frac{d}{dx^\mu}(\overline{z}\sigma^{\mu\nu} v_\nu)\,, \tag{5.19}$$

where $\sigma^{\mu\nu} := \tfrac{1}{2} i[\gamma^\mu, \gamma^\nu]$ and

$$\frac{d}{dx^\mu} := \partial_\mu + \overline{v}_\mu \frac{\partial}{\partial \overline{z}} + v^\mu \frac{\partial}{\partial z} + \overline{v}_{\mu\nu}\frac{\partial}{\partial \overline{v}_\nu} + v_{\mu\nu}\frac{\partial}{\partial v_\nu}$$

with $\overline{v}_{\mu\nu} = \overline{v}_{\nu\mu}$, $v_{\mu\nu} = v_{\nu\mu}$. The new LF reads

$$\mathscr{L}' = \tfrac{1}{2} i(\overline{z}\gamma^\mu v_\mu - \overline{v}_\mu\gamma^\mu z) - m\overline{z} z - i\lambda \overline{v}_\mu \sigma^{\mu\nu} v_\nu\quad (\lambda > 0)\,. \tag{5.20}$$

The canonical momenta assume the form

$$\begin{aligned} \overline{p}^\mu &:= \frac{\partial \mathscr{L}'}{\partial v_\mu} = \frac{i}{2}\overline{z}\gamma^\mu - i\lambda \overline{v}_\nu \sigma^{\nu\mu}\,, \\ p^\mu &:= \frac{\partial \mathscr{L}'}{\partial \overline{v}_\mu} = -\frac{i}{2}\gamma^\mu z - i\lambda\sigma^{\mu\nu} v_\nu\,. \end{aligned} \tag{5.21}$$

With the help of the matrix

$$\tau_{\mu\nu} := \tfrac{2}{3} i g_{\mu\nu} E_4 - \tfrac{1}{3}\sigma_{\mu\nu}\,,$$

which obeys the equations

$$\tau_{\mu\nu}\sigma^{\nu\lambda}=\delta^{\lambda}_{\mu}E_4=\sigma^{\lambda\nu}\tau_{\nu\mu},\quad \gamma^{\mu}\tau_{\mu\nu}=\tau_{\nu\mu}\gamma^{\mu}=-\tfrac{1}{3}i\gamma_{\nu}, \tag{5.22}$$

we can solve Eqs. (5.21) for v_ν and $\bar{v}_\nu$:

$$\lambda\bar{v}_\nu=i\bar{p}^{\mu}\tau_{\mu\nu}-\tfrac{1}{6}i\bar{z}\gamma_\nu,$$
$$\lambda v_\nu=i\tau_{\nu\mu}p^{\mu}+\tfrac{1}{6}i\gamma_\nu z. \tag{5.23}$$

Now we are ready to calculate the HF,

$$\begin{aligned}\mathscr{H}&=\bar{p}^{\nu}v_\nu+\bar{v}_\nu p^{\nu}-\mathscr{L}'\\&=(1/\lambda)(i\bar{p}^{\nu}\tau_{\nu\mu}p^{\mu}+\tfrac{1}{6}i\bar{p}^{\nu}\gamma_\nu z-\tfrac{1}{6}i\bar{z}\gamma_\nu p^{\nu}+\tfrac{1}{3}\bar{z}z)+m\bar{z}z.\end{aligned}$$

If $\psi(x)$ and $\bar{\psi}(x)$ are extremals, we can make the following ansatz:

$$\begin{aligned}S^{j}(x,\bar{z},z)=&\tfrac{1}{2}i[\bar{\psi}(x)\gamma^{j}z-\bar{z}\gamma^{j}\psi(x)]\\&-i\lambda\,\partial_\nu\bar{\psi}(x)\sigma^{\nu j}[z-\psi(x)]\\&-i\lambda\,[\bar{z}-\bar{\psi}(x)]\sigma^{j\nu}\partial_\nu\psi(x),\end{aligned} \tag{5.24}$$
$$j=1,2,3$$

and solve the reduced HJ equation according to the formula (4.2a). The corresponding differential equations of the characteristics are

$$\begin{aligned}\dot{z}=&-\frac{2}{3\lambda}p^{0}+\frac{i}{6\lambda}\gamma_0 z-\frac{i}{2\lambda}\gamma_0\psi(t,\mathbf{c})+\partial_0\psi(t,\mathbf{c})\\&-\frac{2}{3}i\sigma^{0\nu}\partial_\nu\psi(t,\mathbf{c}),\end{aligned} \tag{5.25a}$$

$$\begin{aligned}\dot{\bar{z}}=&-\frac{2}{3\lambda}\bar{p}^{0}-\frac{i}{6\lambda}\bar{z}\gamma_0+\frac{i}{2\lambda}\bar{\psi}(t,\mathbf{c})\gamma^{0}+\partial_0\bar{\psi}(t,\mathbf{c})\\&-\frac{2}{3}i\,\partial_\nu\bar{\psi}(t,\mathbf{c})\sigma^{\nu0},\end{aligned}$$

$$\begin{aligned}\dot{p}^{0}=&\frac{i}{6\lambda}\gamma^{0}p^{0}-\Big(m+\frac{1}{3\lambda}\Big)z+\frac{1}{4\lambda}\psi(t,\mathbf{c})\\&+\frac{1}{6}\gamma_j\sigma^{j\nu}\partial_\nu\psi(t,\mathbf{c})+\frac{1}{2}i\gamma^{j}\partial_j\psi(t,\mathbf{c})\\&+i\lambda\sigma^{j\nu}\partial_j\partial_\nu\psi(t,\mathbf{c}),\end{aligned} \tag{5.25b}$$

$$\begin{aligned}\dot{\bar{p}}^{0}=&-\frac{i}{6\lambda}\bar{p}^{0}\gamma_0-\Big(m+\frac{1}{3\lambda}\Big)\bar{z}+\frac{1}{4\lambda}\bar{\psi}(t,\mathbf{c})\\&-\frac{1}{6}\partial_\nu\bar{\psi}(t,\mathbf{c})\sigma^{\nu j}\gamma_j-\frac{1}{2}i\,\partial_j\bar{\psi}(t,\mathbf{c})\gamma^{j}\\&+i\lambda\,\partial_j\partial_\nu\bar{\psi}(t,\mathbf{c})\sigma^{\nu j}.\end{aligned}$$

If we use the Dirac representation of the γ matrices in which γ^0 is diagonal and if we define the constants $m_0^2:=(2/3\lambda)(m+1/3\lambda)$, $\mathbf{u}:=(u^1,\ldots,u^4)^T$, $\mathbf{w}:=(u^5,\ldots,u^8)$, we obtain the following characteristics which satisfy Eqs. (4.5):

$$\begin{aligned}z(t)=&\cosh(m_0t)\Big[\cos\Big(\frac{t}{6\lambda}\Big)E_4\\&+i\sin\Big(\frac{t}{6\lambda}\Big)\gamma^{0}\Big][\mathbf{u}-\psi(0,\mathbf{c})]+\psi(t,\mathbf{c}),\end{aligned} \tag{5.25c}$$

$$\begin{aligned}\bar{z}(t)=&\cosh(m_0t)[\mathbf{w}-\bar{\psi}(0,\mathbf{c})]\Big[\cos\Big(\frac{t}{6\lambda}\Big)E_4\\&-i\sin\Big(\frac{t}{6\lambda}\Big)\gamma^{0}\Big]+\bar{\psi}(t,\mathbf{c}),\end{aligned}$$

$$\begin{aligned}p^{0}(t)=&-\frac{3\lambda}{2}m_0\sinh(m_0t)\Big[\cos\Big(\frac{t}{6\lambda}\Big)E_4\\&+i\sin\Big(\frac{t}{6\lambda}\Big)\gamma^{0}\Big][\mathbf{u}-\psi(0,\mathbf{c})]\\&-\frac{i}{2}\gamma^{0}\psi(t,\mathbf{c})-i\lambda\sigma^{0\nu}\partial_\nu\psi(t,\mathbf{c}),\end{aligned} \tag{5.25d}$$

$$\begin{aligned}\bar{p}^{0}(t)=&-\frac{3\lambda}{2}m_0\sinh(m_0t)[\mathbf{w}-\bar{\psi}(0,\mathbf{c})]\\&\times\Big[\cos\Big(\frac{t}{6\lambda}\Big)E_4-i\sin\Big(\frac{t}{6\lambda}\Big)\gamma^{0}\Big]\\&+\frac{i}{2}\bar{\psi}(t,\mathbf{c})\gamma^{0}-i\lambda\,\partial_\nu\bar{\psi}(t,\mathbf{c})\sigma^{\nu0}.\end{aligned}$$

The calculation of the line integral (4.8) and solving Eqs. (5.25c) for the u's and w's leads to the result

$$\begin{aligned}S^{0}(x,\bar{z},z)=&\int_0^{x^0}dt\,\mathscr{L}'^{*}(t,\mathbf{x})\\&-\frac{3\lambda}{2}m_0\tanh(m_0x^0)[\bar{z}-\bar{\psi}(x)][z-\psi(x)]\\&+[\tfrac{1}{2}i\bar{\psi}(x)\gamma^{0}-i\lambda\,\partial_\nu\bar{\psi}(x)\sigma^{\nu0}][z-\psi(x)]\\&+[\bar{z}-\bar{\psi}(x)][-\tfrac{1}{2}i\gamma^{0}\psi(x)-i\lambda\sigma^{0\nu}\partial_\nu\psi(x)].\end{aligned} \tag{5.26}$$

The LF (5.18) vanishes on the extremals $\psi(x)$, $\bar{\psi}(x)$; therefore, we can simplify the expression for $\mathscr{L}'^{*}(t,\mathbf{x})$ under the integral in Eq. (5.26), namely,

$$\mathscr{L}'^{*}(t,\mathbf{x})=\lambda\,[\partial^{\nu}\bar{\psi}(x)\,\partial_\nu\psi(x)-m^2\bar{\psi}(x)\psi(x)].$$

The integrability conditions (3.1) for the solutions (5.24) and (5.26) are in general only fulfilled on the extremals.

(vii) In order to couple the free Dirac field to an external electromagnetic field $A^{\mu}(x)$, we replace ∂_μ in the expression (5.18) by $\partial_\mu+iqA_\mu(x)$ (minimal coupling), and again we add the total divergence (5.19). This gives

$$\begin{aligned}\mathscr{L}'=&\tfrac{1}{2}i(\bar{z}\gamma^{\mu}v_\mu-\bar{v}_\mu\gamma^{\mu}z)-m\bar{z}z-qA_\mu(x)\bar{z}\gamma^{\mu}z\\&-i\lambda\bar{v}_\mu\sigma^{\mu\nu}v_\nu\quad(\lambda>0).\end{aligned} \tag{5.27}$$

The LF (5.27) is gauge-invariant on the extremals in the following sense: If we make a gauge transformation $A_\mu(x)\to A_\mu(x)+\partial_\mu f(x)$, $\psi(x)\to e^{-iqf(x)}\psi(x)$, $\bar{\psi}(x)\to e^{iqf(x)}\bar{\psi}(x)$, the LF which results from this transformation will differ from (5.27) only by a total divergence. After the gauge transformation we get

$$\mathscr{L}'\to\tilde{\mathscr{L}}'=\mathscr{L}'-\lambda q\frac{d}{dx^{\mu}}[f(x)\bar{z}\sigma^{\mu\nu}v_\nu-\bar{v}_\nu\sigma^{\nu\mu}f(x)z]$$

with d/dx^{μ} as in Eq. (5.19).

The canonical momenta are identical with those in Eq. (5.21) from which we get the HF with the help of Eqs. (5.23):

$$\begin{aligned}\mathscr{H}=&\frac{1}{\lambda}\Big(i\bar{p}^{\mu}\tau_{\mu\nu}p^{\nu}+\frac{i}{6}\bar{p}^{\mu}\gamma_\mu z-\frac{i}{6}\bar{z}\gamma_\mu p^{\mu}+\frac{1}{3}\bar{z}z\Big)\\&+m\bar{z}z+qA_\mu(x)\bar{z}\gamma^{\mu}z.\end{aligned}$$

The ansatz (5.24) works here, too. That leads to the following differential equations: Those in Eqs. (5.25a) remain the same; to Eqs. (5.25b) we have to add terms of the form $-qA_\mu(x)\gamma^\mu z$ and $-qA_\mu(x)\bar{z}\gamma^\mu$, respectively, on the right-hand sides. These equations are hard to solve, because the γ matrices cannot be diagonalized simultaneously. It is, however, possible to write down the form of a solution of the reduced HJ equation. It reads

$$S^0(x,\bar{z},z)=\int_0^{x^0} dt\,\mathscr{L}'^*(t,\mathbf{x})+[\tfrac{1}{2}i\bar{\psi}(x)\gamma^0-i\lambda\,\partial_\nu\bar{\psi}(x)\sigma^{\nu 0}][z-\psi(x)]$$
$$+[\bar{z}-\bar{\psi}(x)][-\tfrac{1}{2}i\gamma^0\psi(x)-i\lambda\sigma^{0\nu}\partial_\nu\psi(x)]+[\bar{z}-\bar{\psi}(x)]G(x)[z-\psi(x)]\,,$$

where the 4×4 matrix $G(x)$ satisfies the Riccati equation (4.15), that is to say,

$$\dot{G}-\frac{2}{3\lambda}G^2+\frac{i}{6\lambda}G\gamma^0-\frac{i}{6\lambda}\gamma^0 G+\Big(m+\frac{1}{3\lambda}\Big)E_4+qA_\mu(x)\gamma^\mu=0\,. \quad (5.28)$$

In special cases we can find solutions of Eq. (5.28), for example, $A_j\equiv 0, j=1,2,3$. We get the solution

$$G(x)=-\frac{3\lambda}{2}\begin{bmatrix}\epsilon\sqrt{|M_0|}\tanh(\epsilon\sqrt{|M_0|}x^0)E_2 & O_2\\ O_2 & \epsilon\sqrt{|M_1|}\tanh(\epsilon\sqrt{|M_1|}x^0)E_2\end{bmatrix},$$

in which $M_0(x):=(2/3\lambda)[m+1/3\lambda+qA_0(x)]$,
$M_1(x):=(2/3\lambda)[m+1/3\lambda-qA_0(x)]$, and $\epsilon=1$ or $\epsilon=i$ depending on the sign of $M_j, j=0,1$.

(viii) Finally we examine the nonabelian gauge theory (Yang–Mills) with gauge group SU(2). In this case we have 12 real gauge fields $A^{\ \mu}_{a}, \mu=0,\ldots,3, a=1,\ldots,3$. For reasons of convenience we take the temporal gauge [$A^{\ 0}_{a}=0$, $a=1,2,3$; see the arguments in example (v) and Ref. 23] and choose as LF

$$\mathscr{L}=-\tfrac{1}{4}f^{\mu\nu}_a f^a_{\mu\nu}\,,$$

where

$$f^{\mu\nu}_a:=v^{\mu\nu}_a-v^{\nu\mu}_a-g\epsilon_{abc}z^{\ \mu}_b z^\nu_c\,,$$
$$v^{\ \mu 0}_a=v^a_{\mu 0}:=0,\quad z^0_a=z^a_0:=0\,.$$

The canonical momenta are

$$\pi^{\ \mu j}_a:=\frac{\partial\mathscr{L}}{\partial v^a_{\mu j}}=-f^{\mu j}_a\,.$$

Therewith we get the HF, despite the singularity of the LT,

$$\mathscr{H}=-\tfrac{1}{2}\pi^{0j}_a\pi^a_{0j}-\tfrac{1}{4}\pi^{kj}_a\pi^a_{kj}+\tfrac{1}{2}g\epsilon_{abc}\pi^{kj}_a z^b_k z^c_j\,.$$

An ansatz of the type (4.10) [with extremals $A^{\ a}_j(x)$ and momenta $\pi^{\ \mu j}_a(x)$] yields

$$S^j(x,z^k_a)=\pi^{\ jk}_a(x)[z^a_k-A^{\ a}_k(x)],\quad j=1,2,3\,,$$
$$S^0(x,z^k_a)=\int_0^{x^0}dt\,\mathscr{L}^*(t,\mathbf{x})+\pi^{0k}_a(x)[z^a_k-A^{\ a}_k(x)]$$
$$+\tfrac{1}{2}h^{\ kl}_{bc}(x)[z^b_k-A^{\ b}_k(x)][z^l-A^{\ c}_l(x)]$$
$$\times[h^{\ kl}_{bc}(x)=h^{\ lk}_{cb}(x)]\,,$$

in which according to Eq. (4.15) the real symmetric 9×9 matrix $G:=(h^{\ lk}_{cb})$ has to fulfill the Riccati equation

$$\dot{h}^{\ kl}_{bc}+h^{\ kj}_{ba}h^{\ jl}_{ac}+g\epsilon_{abc}\pi^{kl}_a(x)=0$$

or in matrix notation

$$\dot{G}+G^2+g\epsilon_a\times\pi_a(x)=0,$$

where

$$\pi_a(x):=(\pi^{kl}_a(x)),\quad a=1,2,3\quad (\text{i.e., }\pi^T_a=-\pi_a)\,,$$
$$\epsilon_a\times\pi_a(x):=\begin{bmatrix}O_3 & \pi_3(x) & -\pi_2(x)\\ -\pi_3(x) & O_3 & \pi_1(x)\\ \pi_2(x) & -\pi_1(x) & O_3\end{bmatrix}.$$

B. Conserved currents associated with solutions of the HJ equation

We showed in Eq. (3.4) how we can construct conserved currents from solutions of the HJ equation. This shall be illustrated now by simple examples:

With the help of the ansatz (4.10) we found in Sec. 5A solutions of the HJ equation of the form

$$S^{\mu}(x,z)=g^{\mu 0}\int_0^{x^0}dt\,\mathscr{L}^*(t,\mathbf{x})+\pi^{\ \mu}_a(x)[z^a-f^a(x)]$$
$$+\tfrac{1}{2}g^{\mu 0}h_{ab}(x)[z^a-f^a(x)][z^b-f^b(x)]\,.$$

If we embed an extremal $f^a(x)$ which depends differentiably on a parameter α—that is, $f^a=f^a(x;\alpha)$—then we will get according to Eq. (3.4a) the current

$$G^{\mu}(x)=g^{\mu 0}\frac{\partial}{\partial\alpha}\int_0^{x^0}dt\,\mathscr{L}(x,f(x;\alpha),\partial f(x;\alpha))$$
$$-\pi^{\ \mu}_a(x)\frac{\partial}{\partial\alpha}f^a(x;\alpha)\,. \quad (5.29)$$

If we calculate the divergence of the expression (5.29), we obtain

$$\partial_\mu G^{\mu}(x)=\frac{\partial\mathscr{L}^*}{\partial z^a}\partial_\alpha f^a(x;\alpha)+\frac{\partial\mathscr{L}^*}{\partial v^a_\mu}\partial_\alpha\partial_\mu f^a(x;\alpha)$$
$$-\partial_\mu\pi^{\ \mu}_a(x)\,\partial_\alpha f^a(x;\alpha)-\pi^{\ \mu}_a(x)\,\partial_\alpha\partial_\mu f^a(x;\alpha)\,.$$

Because of the validity of the Euler–Lagrange equations for the functions f^a and because of

$$\pi^\mu_a(x)=\frac{\partial\mathscr{L}^*}{\partial v^a_\mu}$$

the functions in (5.29) are indeed the components of a conserved 4-current, and that is independent of the fact whether the current was derived from a solution of the HJ equation or not. For $m=2$, $n=1$, the current (5.29) was first derived in Ref. 7.

For extremals which vanish fast enough for $|\mathbf{x}|\to\infty$ the conserved charge

$$Q:=\int d^3\mathbf{x}\,G^0(x^0,\mathbf{x})=\int d^3\mathbf{x}\frac{\partial}{\partial\alpha}\int_0^{x^0}dt\,\mathscr{L}^*(t,\mathbf{x})-\int d^3\mathbf{x}\,\pi_a^0(x^0,\mathbf{x})\frac{\partial f^a}{\partial\alpha}(x^0,\mathbf{x})$$

exists. Because Q is time-independent, we can calculate Q while we consider the limit $x^0\to 0$, where we have to distinguish two cases:

(i) If the parameter α is related to the time coordinate x^0, the term involving the integral $\int_0^{x^0}dt\,\mathscr{L}^*(t,\mathbf{x})$ in general will give a finite contribution in the limit $x^0\to 0$. For instance, if α appears in the combination $x^0+\alpha$ (time translation), we get

$$Q=\int_{x^0=0}d^3\mathbf{x}\left[\mathscr{L}^*(x^0,\mathbf{x})-\pi_a^0(x^0,\mathbf{x})\frac{\partial f^a}{\partial x^0}(x^0,\mathbf{x})\right].$$

(This is obviously the energy conservation.)

(ii) If α is not related to x^0, we get

$$Q=-\int_{x^0=0}d^3\mathbf{x}\,\pi_a^0(x^0,\mathbf{x})\frac{\partial f^a}{\partial\alpha}(x^0,\mathbf{x}).$$

Instead of the current (5.29) we can take the conserved current

$$G'^{\mu}(x):=g^{\mu 0}\frac{\partial}{\partial\alpha}\int_{-\infty}^{x^0}dt\,\mathscr{L}(x,f(x;\alpha),\partial f(x;\alpha))-\pi_a^\mu(x)\frac{\partial f^a}{\partial\alpha}(x;\alpha),\qquad(5.29')$$

if the integral exists. For solutions with finite action we get the charge

$$Q'=\frac{\partial}{\partial\alpha}\int d^4x\,\mathscr{L}^*-\int_{x^0\to\infty}d^3\mathbf{x}\,\pi_a^0(x^0,\mathbf{x})\frac{\partial f^a}{\partial\alpha}(x^0,\mathbf{x}),$$

provided α is not related to x^0 and the different limits to be taken are interchangable. Before we treat some examples, we want to remark that the special role of the zero component in the expressions (5.29) and (5.29′) is completely arbitrary. We could obtain similar currents in which one of the space components plays a special role.

1. Plane waves in E-dynamics with $j_\mu(x)=0$ [see Eq. (5.11)]

Given $A^0(x)=0$, $A^j(x)=f^j(t-\mathbf{n}\cdot\mathbf{x})$, $j=1,2,3$, $n^2=1$, for instance,

$$\mathbf{n}=(n^1,n^2,n^3)=(\cos\alpha\sin\beta,\ \sin\alpha\sin\beta,\ \cos\beta).$$

The 4-potential A^μ which has been defined in this way is a solution of the homogeneous Maxwell equations in the Lorentz gauge and depends on the parameters α and β. We find that the LF (5.11) vanishes on these extremals, i.e., $\mathscr{L}^*(t,\mathbf{x})=0$ in the expression (5.29). For the current (5.29) we get after a short calculation

$$\begin{aligned}G^0(x)&=-\mathbf{E}^2(x)\sin\beta\,(x^1\sin\alpha-x^2\cos\alpha),\\ G^j(x)&=-n^j\mathbf{E}^2(x)\sin\beta\,(x^1\sin\alpha-x^2\cos\alpha),\quad j=1,2,3,\end{aligned}\qquad(5.30)$$

where $E^j=-\dot{f}^j=-\partial_0A^j$ is the electric field strength.

For plane waves of the above form the continuity equation

$$\partial_0(\mathbf{E}^2)+\mathrm{div}(\mathbf{n}\mathbf{E}^2)=0$$

is always valid; therefore, the current (5.30) is of the following type: Suppose $g^\mu(x)$ is a conserved current, $\partial_\mu g^\mu(x)=0$, and $f(\mathbf{x})$ is a function which does not depend on x^0 and which fulfills $g^j\partial_jf=0$; then $G^\mu:=g^\mu f$ is a conserved current too.

2. Plane waves in the Dirac theory

A plane wave solution of the free Dirac equation is[24]

$$\psi(x)=e^{-ik\cdot x}\frac{\not{k}+m}{\sqrt{2m(m+k_0)}}u,$$

where $k_0:=+\sqrt{\mathbf{k}^2+m^2}$, $u:=(1,0,0,0)^T$, i.e., $\psi(x)$ is a solution to positive energy. The bispinor ψ depends differentiably on the parameters $\mathbf{k}=(k_1,k_2,k_3)^T$, the space components of the 4-momentum k. The LF (5.20) vanishes on this extremal, and therefore the integral in Eq. (5.29) disappears. For example, we can take k_1 as a parameter, and we get the following conserved current:

$$\begin{aligned}G^\mu=&\left(x_1-\frac{k_1}{k_0}x_0\right)\frac{k^\mu}{m}+\frac{i}{2}\left[\frac{1}{\sqrt{2k_0(m+k_0)}}+\frac{\lambda m}{\sqrt{2m(m+k_0)}}\right]\left[e^{-ik\cdot x}\bar\psi(x)\gamma^\mu\left(\frac{k_1}{k_0}\gamma^0+\gamma^1\right)u\right.\\&\left.-e^{ik\cdot x}\bar u\left(\frac{k_1}{k_0}\gamma^0+\gamma^1\right)\gamma^\mu\psi(x)\right]\\&-\frac{i}{2}\frac{\lambda}{\sqrt{2m(m+k_0)}}\left[e^{-ik\cdot x}\bar\psi(x)\gamma^\mu\not{k}\left(\frac{k_1}{k_0}\gamma^0+\gamma^1\right)u\right.\\&\left.-e^{ik\cdot x}\bar u\left(\frac{k_1}{k_0}\gamma^0+\gamma^1\right)\not{k}\gamma^\mu\psi(x)\right],\end{aligned}$$

in which

$$\bar\psi(x)=e^{ik\cdot x}\bar u\frac{\not{k}+m}{\sqrt{2m(m+k_0)}},\quad \bar u:=(1,0,0,0)\gamma^0.$$

If we take k_2 or k_3 as parameters, we will get completely analogous formulas.

3. Plane waves in the nonabelian gauge theory [SU(2)]

Plane wave solutions of the Yang–Mills equations are obtained from the ansatz[25]

$$A^a_\mu=\delta^a_1\Phi(u)p_\mu+\delta^a_3\Psi(u)q_\mu,$$

where $u:=s^\mu x_\mu$, $p^\mu s_\mu=q^\mu s_\mu=q^\mu p_\mu=0$. The special choice $s_\mu:=\delta^3_\mu-\delta^0_\mu$, $p_\mu=q_\mu:=\delta^0_\mu-\delta^2_\mu$ and $\Phi(u)=x_1F(u)$, $\Psi(u)=x_2G(u)$ (F, G are arbitrary functions) yields

$$A^a_\mu(x)=(\delta^0_\mu-\delta^3_\mu)[\delta^a_1x_1F(u)+\delta^a_3x_2G(u)].$$

These are Coleman's plane waves.

If we make the choice

$$\Phi(u)=\Psi(u)\quad\text{and}\quad p^2=q^2,$$

the field equations require that $\Phi(u)$ has to satisfy the differential equation

$$\Phi''(u)-\frac{g^2p^2}{s^2}\Phi^3(u)=0\quad\left[\Phi'(u):=\frac{d\Phi}{du}(u)\right].\qquad(5.31)$$

A solution of this equation is Jacobi's elliptic function[26]

$$\Phi(u) = \mathrm{cn}(u, k),$$

in which $k^2 = \frac{1}{2}$ and $s^2 = -g^2p^2$. We now calculate the current (5.29) with s_0 as the parameter α. We have

$$F^a_{\mu\nu}(x) := \partial_\mu A^a_\nu(x) - \partial_\nu A^a_\mu(x) - g\epsilon^{abc}A^b_\mu A^c_\nu,$$
$$\mathscr{L}^* = -\tfrac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a = -\tfrac{1}{2}p^2[2(\Phi')^2 + g^2\Phi^4p^2],$$

and

$$\frac{\partial\mathscr{L}^*}{\partial s_0} = -2p^2[\Phi'\cdot\Phi''x^0s^2 + g^2x^0p^2\Phi'\cdot\Phi^3 + s_0(\Phi')^2],$$
$$-\pi^{\mu\nu}_a(x)\frac{\partial A^a_\nu}{\partial s_0}(x) = F^{\mu\nu}_a(x)\frac{\partial A^a_\nu}{\partial s_0}(x) = 2x^0(\Phi')^2s^\mu p^2.$$

Using Eq. (5.31), we get the conserved current

$$G^\mu(x) = -2p^2g^{\mu 0}\int_0^{x^0} dt\,[2tg^2p^2\Phi'\cdot\Phi^3 + s_0(\Phi')^2] + 2x^0p^2s^\mu(\Phi')^2.$$

4. Instanton solutions of the Euclidean SU(2)–Yang–Mills theory

We now want to investigate the so-called instanton solutions of the Yang-Mills field equations in Euclidean space–time [with metric $(++++)$], which can be characterized through topological quantum numbers.[27] The gauge fields and the related "field strengths" for an Euclidean SU(2)–Yang–Mills theory are

$$\mathscr{A}_\mu := gA^a_\mu\tau_a/2, \quad \mu = 1,\dots,4,$$
$$\mathscr{F}_{\mu\nu} := gF^a_{\mu\nu}\tau_a/2 = \partial_\mu\mathscr{A}_\nu - \partial_\nu\mathscr{A}_\mu + i[\mathscr{A}_\mu, \mathscr{A}_\nu], \quad \mu,\nu = 1,\dots,4$$

(τ_1, τ_2, τ_3 are the Pauli matrices). They obey the field equations

$$\partial_\mu\mathscr{F}_{\mu\nu} + i[\mathscr{A}_\mu, \mathscr{F}_{\mu\nu}] = 0. \tag{5.32}$$

The field equations for the components $F^a_{\mu\nu}$ are

$$\partial_\mu F^a_{\mu\nu} - g\epsilon^{abc}A^b_\mu F^c_{\mu\nu} = 0.$$

Instantons are solutions of Eqs. (5.32) with self-dual "field strengths" $\mathscr{F}_{\mu\nu}$, that is,

$$\mathscr{F}_{\mu\nu} \overset{!}{=} {}^*\mathscr{F}_{\mu\nu} := \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}\mathscr{F}_{\rho\sigma}.$$

For such solutions the topological quantum number r (often called the *Pontryagin index*) coincides up to a constant factor with the action integral, namely,

$$-\frac{8\pi^2}{g^2}r = -\frac{1}{4}\int d^4x\, F^a_{\mu\nu}F^a_{\mu\nu}.$$

We consider instanton solutions with Pontryagin index $r = 1$; they are

$$\mathscr{A}_\mu(x) = -\frac{2(x_\nu - a_\nu)}{\lambda^2 + (x-a)^2}\rho_{\mu\nu}, \tag{5.33}$$

in which

$$x^2 := x_\mu x_\mu, \quad \rho_{\mu\nu} := (1/4i)(s_\mu\bar{s}_\nu - s_\nu\bar{s}_\mu),$$
$$(s_\mu) := (i\boldsymbol{\tau}, E_2), \quad (\bar{s}_\mu) := (-i\boldsymbol{\tau}, E_2).$$

Because of $\rho_{\mu\nu} = {}^*\rho_{\mu\nu}$ the "field strengths"

$$\mathscr{F}_{\mu\nu}(x) = +\frac{4\lambda^2}{[\lambda^2 + (x-a)^2]^2}\rho_{\mu\nu}$$

are self-dual.

The solutions (5.33) contain five parameters: λ and $a = (a_1,\dots,a_4)$. The Euclidean current associated with the parameter λ is

$$G_\mu(x) = \delta_{\mu 4}\frac{\partial}{\partial\lambda}\int_0^{x_4} d\tau\,\mathscr{L}^*(\tau, \mathbf{x}) - \pi^a_{\mu\nu}(x)\frac{\partial A^a_\nu}{\partial\lambda}(x), \tag{5.34}$$

where

$$\mathscr{L}^* = -\tfrac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu} = -\frac{48}{g^2}\frac{\lambda^4}{[\lambda^2 + (x-a)^2]^4}$$

and

$$-\pi^a_{\mu\nu}(x)\frac{\partial A^a_\nu}{\partial\lambda}(x) = F^a_{\mu\nu}(x)\frac{\partial A^a_\nu}{\partial\lambda}(x) = -\frac{48}{g^2}\frac{\lambda^3(x_\mu - a_\mu)}{[\lambda^2 + (x-a)^2]^4}.$$

The current (5.34) takes the form

$$G_\mu(x) = -\frac{192}{g^2}\lambda^3\delta_{\mu 4} \times \int_0^{x_4} d\tau\,\frac{(\mathbf{x}-\mathbf{a})^2 + (\tau - a_4)^2 - \lambda^2}{[\lambda^2 + (\mathbf{x}-\mathbf{a})^2 + (\tau - a_4)^2]^5} - \frac{48}{g^2}\lambda^3\frac{x_\mu - a_\mu}{[\lambda^2 + (x-a)^2]^4}.$$

The corresponding charge has the value

$$Q := \int d^3x\, G_4(\mathbf{x}, x_4) = \frac{6\pi^2}{g^2}\frac{\lambda^3a_4}{(\lambda^2 + a_4^2)^{5/2}}.$$

The charge associated with a current of the form (5.29′) is $Q' = 0$.

C. Construction of new extremals from solutions of the HJ equation

We now want to illustrate the use of a complete integral for the construction of a solution of the field equation by a simple example: For that purpose we take the Klein–Gordon field from Eq. (5.1). The ansatz

$$S^\mu(x, z) := \tfrac{1}{2}h^\mu(x)z^2$$

leads to the differential equation

$$\partial_\mu h^\mu = -h^\mu h_\mu - m^2.$$

A solution of this equation is

$$h^0 = \tfrac{1}{2}m\tan(-\tfrac{1}{2}mx^0 + \alpha^0),$$
$$h^j = \tfrac{1}{2}m\tanh(-\tfrac{1}{2}mx^j + \alpha^j),$$
$$j = 1,2,3, \quad \alpha^\mu = \text{const}.$$

The integrability conditions here are

$$(\partial_\nu h_\mu - \partial_\mu h_\nu)z \overset{!}{=} 0$$

and are apparently fulfilled. Furthermore, we have

$$\frac{\partial S^0}{\partial \alpha^0} = \frac{m}{4}\frac{z^2}{\cos^2(-\frac{1}{2}mx^0+\alpha^0)},$$
$$\frac{\partial S^j}{\partial \alpha^j} = \frac{m}{4}\frac{z^2}{\cosh^2(-\frac{1}{2}mx^j+\alpha^j)}, \quad j=1,2,3.$$

Thus, the value of the functional determinant (3.5)

$$\left|\frac{\partial^2 S^\mu}{\partial z \partial \alpha^\nu}\right| = \left(\frac{m}{2}\right)^4 \times \frac{z^4}{\cos^2(-\frac{1}{2}mx^0+\alpha^0)\Pi_{j=1}^3 \cosh^2(-\frac{1}{2}mx^j+\alpha^j)}$$

is in general unequal to zero. Therewith all assumptions are given in order to carry out the method of solving.

Only the diagonal elements $\widetilde{G}^0{}_0,\ldots,\widetilde{G}^3{}_3$ of the functions $\widetilde{G}^\mu{}_\nu$ from Eq. (3.8) are unequal to zero; therefore $g^\mu{}_\nu \equiv 0$ $(\mu \neq \nu)$. Eventually we get four equations for the remaining components of $g^\mu{}_\nu$ $(\mu=\nu)$, namely,

$$\widetilde{G}^0{}_0 = \frac{\partial S^0}{\partial \alpha^0} = \frac{m}{4}\frac{z^2}{\cos^2(-\frac{1}{2}mx^0+\alpha^0)} \overset{!}{=} g^0{}_0(x),$$
$$\widetilde{G}^j{}_j = \frac{\partial S^j}{\partial \alpha^j} = \frac{m}{4}\frac{z^2}{\cosh^2(-\frac{1}{2}mx^j+\alpha^j)} \overset{!}{=} g^j{}_j(x), \quad j=1,2,3 \tag{5.35}$$

with the additional conditions (3.12):

$$\partial_0 g^0{}_0(x)=0, \quad \ldots, \quad \partial_3 g^3{}_3(x)=0.$$

We choose the functions $g^\mu{}_\nu(x)$ $(\mu=\nu)$ to be

$$g^0{}_0 = \left(\frac{m}{4}\right)^3 \prod_{j=1}^3 \cosh^2\left(-\frac{m}{2}x^j+\alpha^j\right),$$
$$g^j{}_j = \left(\frac{m}{4}\right)^3 \cos^2\left(-\frac{m}{2}x^0+\alpha^0\right) \times \prod_{\substack{i=1\\ i\neq j}}^3 \cosh^2\left(-\frac{m}{2}x^i+\alpha^i\right), \quad j=1,2,3.$$

We are now able to solve (5.35) uniquely for z with the result

$$z=z(x)=\left(\frac{m}{4}\right)\cos\left(-\frac{m}{2}x^0+\alpha^0\right) \times \prod_{j=1}^3 \cosh\left(-\frac{m}{2}x^j+\alpha^j\right). \tag{5.36}$$

Indeed, the expression (5.36) solves the Klein–Gordon equation

$$(\Box + m^2)z(x)=0.$$

ACKNOWLEDGMENTS

I am indebted to Professor H. A. Kastrup for suggesting the above investigation of the theory of De Donder and Weyl and for many fruitful discussions on this subject. Furthermore, I thank him for a careful reading of the manuscript. Thanks are due to Dr. M. Magg and to Dr. M. Rinke for some helpful suggestions. Finally I am grateful to my friends K.-P. Dörpelkus, M. Morgenstern, and A. Nüsser for critical and stimulating discussions.

[1]Th. DeDonder, *Theorie invariantive du calcul des variations* (Gauthier-Villars, Paris, 1935), nouv. éd.

[2]H. Weyl, Ann. Math. **36**, 607 (1935).

[3]H. A. Kastrup, "Canonical Theories of Dynamical Systems in Physics," Aachen preprint, March 1982, Phys. Rep. (to be published).

[4]C. Caratheodory, *Calculus of Variations and Partial Differential Equations of the First Order* (Holden-Day, San Francisco, 1967), Part II, Secs. 227–232.

[5]H. Boerner, Math. Ann. **112**, 187 (1936).

[6]I. M. Gel'fand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, NJ, 1963), Sec. 33.

[7]H. A. Kastrup, "Canonical Theories of Dynamical Systems in Physics," Aachen preprint, 1982, Chap. 4.

[8]H. Rund, *The Hamilton–Jacobi Theory in the Calculus of Variations* (Krieger, Huntington, NY, 1973), Chap. 4.2.

[9]R. Klötzler, *Mehrdimensionale Variationsrechnung* (Birkhäuser, Basel, 1970).

[10]H. Rund, *The Hamilton–Jacobi Theory in the Calculus of Variations* (Krieger, Huntington, NY, 1973).

[11]I. M. Gel'fand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, NJ, 1963), Chap. 7.

[12]H. A. Kastrup, "Canonical Theories of Dynamical Systems in Physics," Aachen preprint, 1982, Chaps. 4,6–8.

[13]E. Kamke, *Differentialgleichungen, Lösungsmethoden und Lösungen* (Akademische Verlagsgemeinschaft, Leipzig, 1959), Bd. II, Chap. 12.9.

[14]R. Klötzler, *Mehrdimensionale Variationsrechnung* (Birkhäuser, Basel, 1970), Sec. 24.

[15]W. T. Reid, *Riccati Differential Equations* (Academic, New York, 1972), Chap. 1.2.

[16]J. Harnad, P. Winternitz, and R. L. Anderson, "Superposition Principles for Matrix Riccati Equations," CRMA Preprint 1024, Montreal, 1981.

[17]C. Caratheodory, *Calculus of Variations and Partial Differential Equations of the First Order* (Holden-Day, San Francisco, 1967), Part II, Sec. 314.

[18]R. Klötzler, *Mehrdimensionale Variationsrechnung* (Birkhäuser, Basel, 1970), Sec. 21.

[19]Reference 18, Sec. 20.

[20]M. Rinke, private communication, 1982.

[21]M. Creutz, Ann. Phys. **117**, 471 (1979).

[22]S. Gasiorowicz, *Elementary Particle Physics* (Wiley, New York, 1966), Chap. 5.

[23]L. D. Faddeev and A. A. Slavnov, *Gauge Fields* (Benjamin/Cummings, London, 1980), Chap. 3.2.

[24]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980), Chap. 2.2.

[25]C. H. Oh and R. Teh, "Non-Abelian Plane Waves," CERN Preprint TH 3276, 1982.

[26]A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. II, Chap. 13.16.

[27]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Phys. Lett. B **59**, 85 (1975).

Causal transformations of Wightman localizations

D. P. L. Castrigiano
Institut für Mathematik der Technischen Universität München, Arcisstrasse 21, D-8000 München 2, Federal Republic of Germany

(Received 9 March 1982; accepted for publication 25 March 1983)

Causal transformations of Wightman localizations are studied without assuming further symmetries (e.g., Galilean or Lorentz invariance). They are characterized completely by a condition which imposes analyticity and restricted growth on the matrix elements and implies a close connection between the helicity components. This condition is applied to obtain causal time evolutions. Spin-0 and spin-$\frac{1}{2}$ systems and, generally, all simple systems of definite spin are considered in detail. The factorizing causal time evolutions are determined completely.

PACS numbers: 11.90. + t, 03.65. − w

I. INTRODUCTION

The Wightman localization[1] is used to describe the localizability of quantum systems, e.g., relativistic or nonrelativistic massive particles. It consists of a continuous unitary representation U of the universal covering group ISU(2) of the Euclidean group, and of a projection valued measure E on the coordinate space $\mathbb{R}^3$, both acting on a Hilbert space $\mathscr{H}$ and satisfying the covariance property

$$U(g)E(\Delta)U(g)^{-1} = E(g\cdot\Delta), \tag{1}$$

where $g = (\mathbf{b},B)\in\mathrm{ISU}(2)$, $\Delta\subset\mathbb{R}^3$ a Borel set, $g\cdot\Delta := \{\mathbf{b} + B\cdot\mathbf{x}\,|\,\mathbf{x}\in\Delta\}$, $B\cdot\mathbf{x} := \Lambda(B)\mathbf{x}$, and $\Lambda:\mathrm{SU}(2)\to\mathrm{SO}(3)$ the universal covering homomorphism. In the one-particle case its physical interpretation is rather obvious. In the general case it describes a pointlike and Euclidean covariant characteristic of the quantum system. (For example, one could think of its center of mass.)

The time evolution of a quantum system is usually described by a continuous unitary representation V of the additive group of real numbers. Let us assume that, as in the case of the nonhomogeneous Galilei group or the Poincaré group, V commutes with U, i.e.,

$$V(t)U(g) = U(g)V(t). \tag{2}$$

Localizability and time evolution are not independent of each other but are connected by causality. We give the following natural description of it:

Definition 1: The time evolution V is called causal if

$$V(t)E(\Delta)V(-t)\leqslant E(\Delta_t) \tag{3}$$

for every t and Δ, where

$$\Delta_t := \{\mathbf{y}\,|\,\exists\ \mathbf{x}\in\Delta : |\mathbf{y}-\mathbf{x}|\leqslant|t|\}.$$

The interpretation of (3) is clear. A system, localized in Δ at time zero, is expected to be found within Δ_t at time $t>0$, since it cannot move faster than light (velocity of light = 1). Equations (1)–(3) give a simple mathematical description of Einstein causality for quantum systems. However, it contradicts other physical requirements. First, it does not allow positivity of energy except in the trivial case where the time evolution V commutes with the localization E. Indeed, take $t>0$ and Δ, e.g., a closed ball. Then $V(\tau)E(\Delta)V(-\tau)(1-E(\Delta_t)) = 0$ for all $|\tau|<t$. By a theorem of Borchers,[2] this remains valid for all τ if the generator of V, the energy operator, is semibounded. In this case the limit $t\to 0$ shows that $V(\tau)E(\Delta) = E(\Delta)V(\tau)$. Now it is a merely technical step to extend this equation to all Borel sets Δ.

Statements similar to this one, and others which also involve conditions on the energy-momentum spectrum are proven in Refs. 3–5. Acausal features, predicted according to the preceding considerations, are studied in Ref. 6 for a free relativistic massive particle.

A second defect of the above concept is that the Wightman localization does not exist for relativistic massless particles with nonzero helicity (see Ref. 1).

There is a more sophisticated approach to causality, namely, to look for representations in Hilbert space of the causal logic $\mathscr{L}$.[7,8] $\mathscr{L}$ is the lattice of the space-time domains of determinancy (= diamonds) for spacelike regions in the case of a limited velocity of propagation. (Its structure is well known from hyperbolic partial differential equations.) However, since the Cauchy data have to be prescribed for spacelike regions, it is possible to return to a formulation of causality involving only spacelike hypersurfaces. Indeed, let F be a representation of $\mathscr{L}$, i.e., a normalized σ-orthoadditive mapping of $\mathscr{L}$ into the lattice of the projections on $\mathscr{H}$, which is covariant under Poincaré transformations by means of a continuous unitary representation W of the universal covering group $\widetilde{\mathscr{P}}$ of the Poincaré group. Define U and V as the restriction of W to Euclidean transformations and time translations respectively, and define E by $E(\Delta) := F(\Delta^{\perp\!\!\perp})$, where $\Delta^{\perp\!\!\perp}$ denotes the causal closure of Δ. Then (1) and (2) hold trivially, and (3) follows from $V(t)E(\Delta)V(-t) = F((t+\Delta)^{\perp\!\!\perp})$, since $t+\Delta\subset(\Delta_t)^{\perp\!\!\perp}$. Thus the difficulties mentioned persist.

A possible way to solve both problems is to substitute the Wightman localization for the more general Euclidean system of covariance (ESC) by replacing the projection valued measure by a positive operator valued one, see Ref. 9 and literature cited there. The reasonable physical features of such systems are shown generally for massless particles in Ref. 10 and particularly for the photon in Ref. 11. Let us also see that the difficulty of negative energy can be avoided when ESC are used. Suppose a Wightman localization (U,E) and a

 0022-2488/84/041116-06$02.50

nontrivial causal time evolution V are given. Then according to the foregoing considerations, the energy operator is not semibounded. However, there exists a spectral projection P of it for a semibounded interval such that $PV(t)P$ does not commute with $PE(\Delta)P$. Clearly P commutes with V, and since V commutes with U, it commutes with U. Let $\widetilde{V}$ and $\widetilde{U}$ denote the reduction of V and U, respectively, by P, and let $R(\Delta)$ be defined by $PE(\Delta)P$ taken as an operator in the range of P. Then $(\widetilde{U},R)$ is an ESC, and $\widetilde{V}$ is a unitary time evolution which commutes with $\widetilde{U}$, does not commute with R, and is causal in the sense that

$$\widetilde{V}(t)R(\Delta)\widetilde{V}(-t)\leqslant R(\Delta_t), \tag{4}$$

and the generator of $\widetilde{V}$ is semibounded, as desired.

This method of constructing "causal ESC with semibounded energy" states one of the reasons to study Wightman localizations with causal time evolutions. (However, it is an open question whether all ESC with causal time evolution can be obtained by that method.) On the other hand, one can still hope to find a reasonable interpretation for the negative energies. Last but not least, these investigations can help clarify the relation between relativity and causality in quantum theory.

The following notion, which is apparently weaker than that of a causal time evolution, turns out to be very useful. Its analysis is the main purpose of this paper.

Definition 2: A bounded operator T commuting with U is called a causal transformation if there is a compact ball Ω of positive radius such that for each state ψ localized within Ω, i.e., $E(\Omega)\psi=\psi$, there is a compact region Δ such that the transformed state $T\psi$ is localized within Δ, i.e., $E(\Delta)T\psi = T\psi$. Our main result is the following:

Theorem: A bounded operator T commuting with U is a causal transformation if and only if there are entire functions f_{jkl} of one complex variable such that $z\to f_{jkl}(z^2)$ are uniformly exponentially bounded and

$$\langle k,s|T_{sp}|j,s\rangle = \sum_l (-1)^{j-s}\sqrt{2l+1}\begin{pmatrix} j & k & l\\ -s & s & 0\end{pmatrix}p^l f_{jkl}(p^2), \tag{5}$$

where $|j,s\rangle$ and $|k,s\rangle$ are spin eigenstates of helicity s, p is the absolute value of the momentum, $\left(\begin{smallmatrix} j & k & l\\ -s & s & 0\end{smallmatrix}\right)$ is the Wigner 3-j symbol, and T_{sp} is the reduction of T with respect to the helicity basis.

Equation (5) imposes analyticity and restricted growth on the matrix elements with respect to their dependence on momentum. It also connects closely the helicity components with each other. Of course, causal transformations do not mix up boson and fermion states since they commute with Euclidean transformations. But in addition, the transition probability (5) between two boson (fermion) spin eigenstates $|j\rangle$ and $|k\rangle$ of equal helicity vanishes for small linear momenta p like $p^{|j-k|}$.

In Sec. III we apply the Theorem to study causal time evolutions for some physical interesting cases, namely, (a) spin-0 systems with arbitrary finite multiplicity, (b) spin-$\frac{1}{2}$ systems with multiplicity one or two, and (c) factorizing causal time evolutions which include all simple systems with definite spin. In case (a) it turns out that there are only trivial unitary causal transformations, i.e., they commute with E. One remembers the difficulties of incorporating the Klein–Gordon equation to Hilbert space formalism so that the time evolution is unitary. Here we get a no-go theorem even without assuming relativistic time evolution. In (b), the case of multiplicity one mainly yields the relativistic time evolution of the neutrino–antineutrino system. In the case of multiplicity two the Dirac energy operator and the energy operator of the doubled neutrino–antineutrino system are the only generators for a relativistic causal time evolution. In (c) we encounter the situation that there is a proportional coupling of the limiting velocity to the absolute value of the helicity, i.e., another kind of dispersion relation. These time evolutions, like some time evolutions occurring in (b), are causal but not relativistic. Because of their Euclidean invariance, they could serve to describe quantum systems, for example quasiparticles, in a homogeneous medium.

II. PROOF OF THE THEOREM

According to Mackey's imprimitivity theorem the Wightman localizations are obtained by inducing the representations of SU(2) to ISU(2)

$$\bigoplus_j \nu_j D^{(j)}\uparrow U =: \bigoplus_j \nu_j \overline{D}^{(j)}, \tag{6}$$

where the spin j runs through $\{0,\frac{1}{2},1,\ldots\}$, $\nu_j\in\{0,1,2,\ldots,\aleph_0\}$ are the multiplicities ($\leqslant\aleph_0$ since we restrict ourselves to separable spaces) of the mutually inequivalent representations $D^{(j)}$ of SU(2), and the induced representations $\overline{D}^{(j)}$ act on $\mathscr{H}_j := L^2(\mathbb{R}^3,\mathbb{C}^{2j+1})$ by $(\overline{D}^{(j)}(\mathbf{b},B)g)(\mathbf{x}) = D^{(j)}(B)g(B^{-1}(\mathbf{x}-\mathbf{b}))$. The corresponding covariant PV-measure E is the canonical one, i.e., $E(\Delta)$ multiplies each component by the characteristic function 1_Δ of Δ. This is the coordinate space representation.

We use the notation $|j,\iota\rangle\otimes g = (|j,\iota,s\rangle\otimes g_s)_{s\in[j]}$, where the label $\iota\in\{0,1,\ldots,\nu_j\}$ accounts for the multiplicity, $[j] := \{-j,-j+1,\ldots,j-1,j\}$, and $g\in\mathscr{H}_j$, to indicate the various mutually orthogonal subspaces and components.

Now we pass to the helicity representation. Each subrepresentation $\overline{D}^{(j)}$ of U is unitarily transformed by the Fourier transformation $\mathscr{F}$ which acts separately on each spin component and, subsequently, by the helicity transformation

$$(X_j f)(\mathbf{p}) := D^{(j)}(B(\mathbf{p})^{-1}f(\mathbf{p}),$$

where $f\in\mathscr{H}_j$ and $\mathbb{R}^3\ni\mathbf{p}\to B(\mathbf{p})\in SU(2)$ is any measurable map satisfying

$$B(\mathbf{p})\cdot\mathbf{e}_3 = \mathbf{p}, \mathbf{e}_3 := \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

One takes, for instance, the helicity section

$$B(\mathbf{p}) := \begin{pmatrix} a_+ & -b^*a_-\\ ba_- & a_+\end{pmatrix},\quad a_\pm := \left(\frac{p\pm p_3}{2p}\right)^{1/2},$$

$$b := \frac{p_1+ip_2}{|p_1+ip_2|}.$$

Since the Wigner rotation $R(\mathbf{p},B) := B(\mathbf{p})^{-1}BB(B^{-1}\cdot\mathbf{p})$ leaves $\mathbf{e}_3$ fixed, it is diagonal so that $D^{(j)}(R(\mathbf{p},B)) = \mathrm{diag}(\kappa(\mathbf{p},B)^{2j},\ldots,\kappa(\mathbf{p},B)^{-2j})$, where $\kappa(\mathbf{p},B) := R(\mathbf{p},B)_{11}$.

Therefore the transformed representation $\overline{D}^{(j)}$ splits into the direct sum of $2j+1$ subrepresentations characterized by the exponent $2s, s\in[j]$. The primary component of U corresponding to a particular value of $s\in\mathbb{Z}/2$ acts on the subspace $\mathscr{V}_s$ generated by the orthonormal set $\{|j,\iota,s\rangle \mid [j]\ni s, 1\leqslant\iota\leqslant\nu_j\}$. The dimension of $\mathscr{V}_s$ is $\Sigma_{[j]\in s}\nu_j$. Let X be the unitary transformation acting on each subspace $|j,\iota\rangle\otimes\mathscr{H}_j$ by X_j. Then, in the helicity representation, one gets $U^h:=X\mathscr{F}U\mathscr{F}^{-1}X^{-1}$ acting on $\oplus_s L^2(\mathbb{R}^3,\mathscr{V}_s)$ by

$$(U^h(\mathbf{b},B)F)_s(\mathbf{p})=\exp(-i\langle\mathbf{b},\mathbf{p}\rangle)\kappa(\mathbf{p},B)^{2s}F_s(B^{-1}\cdot\mathbf{p}). \quad (7)$$

One infers from (7) that a bounded (unitary) operator T commutes with U if, and only if, there exist uniformly bounded (unitary) operators T_{sp} in $\mathscr{V}_s$ such that in the helicity representation

$$(T^hF)_s(\mathbf{p})=T_{sp}F_s(\mathbf{p}). \quad (8)$$

T commutes with (U,E) if and only if, after altering T_{sp} on some null set, T_{sp} does not depend on p, say $T_{sp}=T_s$, and $\langle k,\kappa,r|T_r|j,\iota,r\rangle=\delta_{kj}\langle k,\kappa,s|T_s|j,\iota,s\rangle\,\forall\ r,s\in[k]\cap[j]$. Returning to the coordinate space representation one gets

$$T|j,\iota\rangle\otimes g=\sum_k\sum_{1\leqslant\kappa\leqslant\nu_k}|k,\kappa\rangle\otimes\mathscr{F}^{-1}A^{j\iota}_{k\kappa}\mathscr{F}g, \quad (9)$$

where $A^{j\iota}_{k\kappa}$ is a complex $(2k+1)\times(2j+1)$ matrix valued function on $\mathbb{R}^3$, which maps $\mathscr{H}_j$ into $\mathscr{H}_k$ by pointwise multiplication. It is built out of the matrix elements

$$a^{j\iota}_{k\kappa,s}(p):=\langle k,\kappa,s|T_{sp}|j,\iota,s\rangle \quad (10)$$

as follows. Define the pseudodiagonal $(2k+1)\times(2j+1)$ matrix $(N^{j\iota}_{k\kappa})_{rs}(p):=\delta_{rs}a^{j\iota}_{k\kappa,s}(p), r\in[k], s\in[j]$. Then

$$A^{j\iota}_{k\kappa}(\mathbf{p}):=D^{(k)}(B(\mathbf{p}))N^{j\iota}_{k\kappa}(p)D^{(j)}(B(\mathbf{p})^{-1}). \quad (11)$$

In the following, we use the multi-index ρ for (j,ι,k,κ).

$$A^\rho(B\cdot\mathbf{p})=D^{(k)}(B)A^\rho(\mathbf{p})D^{(j)}(B)^{-1}\forall\ B\in SU(2),\quad \mathbf{p}\in\mathbb{R}^3, \quad (12)$$

is a necessary and sufficient condition that A^ρ is of the form (11). In this case the following relation holds:

$$A^\rho(p\mathbf{e}_3)=N^\rho(p). \quad (13)$$

Now we turn to causal transformations (see Def. 2).

Lemma 1: Let T be a causal transformation. Then there exists a compact ball $\widetilde{\Omega}\supset\Omega$ concentric with Ω such that $(1-E(\widetilde{\Omega}))TE(\Omega)=0$.

Proof: Assume the contrary. Then there is an increasing sequence (Δ_n) of compact sets tending to the whole space, and a sequence (ψ_n) in $E(\Omega)\mathscr{H}$ such that $E(\Delta_n)T\psi_n=T\psi_n$ and $E(\Delta_n)T\psi_{n+1}\neq T\psi_{n+1}$. Apply Gram–Schmidt orthonormalization to get (ψ_n) orthonormal and assume without loss of generality that $\|T\|=1$. Then

$$\beta_1:=1, \beta_{n+1}:=\|(1-E(\Delta_n))T\psi_{n+1}\|\in]0,1],$$

$$(\alpha_n):=\left(3^{-n}\beta_n^{-1}\prod_{\nu=1}^n\beta_\nu\right)\in l^2,$$

and

$$\psi:=\sum_{n=1}^\infty\alpha_n\psi_n\in E(\Omega)\mathscr{H}.$$

However,

$$\|(1-E(\Delta_k))T\psi\|=\|\sum_{n=k+1}^\infty\alpha_n\|(1-E(\Delta_k))T\psi_n\|$$
$$\geqslant\alpha_{k+1}\|(1-E(\Delta_k))T\psi_{k+1}\|-\sum_{n=k+2}^\infty\alpha_n$$
$$\geqslant\alpha_{k+1}\beta_{k+1}-\tfrac{1}{2}\alpha_{k+1}\beta_{k+1}>0,\quad\forall\ k.$$

This contradicts T being a causal transformation. ■

Lemma 2: T is a causal transformation if and only if A^ρ has an analytic continuation to an exponentially bounded entire function. More precisely, there exists an entire function Φ^ρ on $\mathbb{C}^3$ into the Banach space of all complex $(2k+1)\times(2j+1)$ matrices such that $\Phi^\rho|\mathbb{R}^3=A^\rho$ almost everywhere independently of ρ and $\|\Phi^\rho(\mathbf{z})\|\leqslant c_\rho\exp[\delta(|z_1|+|z_2|+|z_3|)]\,\forall\ \mathbf{z}\in\mathbb{C}^3$, where c_ρ, $\delta\geqslant0$ are constants. δ can be chosen to be the difference between the radii of Ω and $\widetilde{\Omega}$ (cf. Lemma 1).

For the proof, one uses the coordinate space representation where T is given by (9) and $E(\Delta)$ is the multiplication operator by 1_Δ. Then the main steps are to apply the generalized Paly–Wiener theorem[12] in both directions.

In particular, since altering $p\mapsto T_{sp}$ on a null set does not alter T, we may assume a^ρ_s (10) to be analytic. Then N^ρ and, by continuity, A^ρ are analytic. Let us denote their respective entire continuations by the same letter.

For further analysis of the A^ρ's we introduce a Hilbert space $\mathscr{A}^j_k$, in short $\mathscr{A}$, of entire functions similar to that of Bargmann.[13] It consists of all entire functions Φ on $\mathbb{C}^3$ into the Banach space of all complex $(2k+1)\times(2j+1)$ matrices such that $\int d\mu(\mathbf{z})\mathrm{tr}(\Phi(\mathbf{z})^+\Phi(\mathbf{z}))<\infty$, where $\mathbf{z}=(z_1,z_2,z_3)=(x_1+iy_1,\ldots)$ and $d\mu(\mathbf{z}):=\pi^{-3}\times\exp(-|z_1|^2-|z_2|^2-|z_3|^2)dx_1\,dy_1\,dx_2\,dy_2\,dx_3\,dy_3$. The inner product is given by $\langle\Phi,\psi\rangle=\int d\mu(\mathbf{z})\mathrm{tr}(\Phi(\mathbf{z})^+\psi(\mathbf{z}))$. Clearly, $\mathscr{A}$ contains all entire functions which are exponentially bounded and, in particular, the set $\mathscr{P}$ of all polynomials. $\mathscr{P}$ is dense in $\mathscr{A}$. $\mathscr{A}$ carries the continuous unitary representation ξ of SU(2)

$$(\xi(B)\Phi)(\mathbf{z}):=D^{(k)}(B)\Phi(B^{-1}\cdot\mathbf{z})D^{(j)}(B)^{-1}. \quad (14)$$

Because of (12) the A^ρ's are ξ-invariant. The projection π in $\mathscr{A}$ onto the subspace $\mathscr{A}_0$ of the ξ-invariant elements is given by $\pi=\int\xi(B)dB$ where dB denotes the Haar measure on SU(2) (see, e.g., Ref. 14). In order to evaluate this integral, we need some facts on homogeneous polynomials.

Lemma 3: Every homogeneous polynomial q of the variables $X_1,\ldots,X_m$, $m\geqslant2$, and degree $l\in\{0,1,2,\ldots\}$ has a unique expansion of the form

$$q=\sum_{\nu=0}^{\langle l/2\rangle}d^\nu H_{l-2\nu}, \quad (15)$$

where $\langle l/2\rangle$ is the greatest integer $\leqslant l/2$, $d(X_1,\ldots,X_m):=(X_1)^2+\cdots+(X_m)^2$, and H_k is a homogeneous harmonic polynomial of degree k.

Proof: In order to prove uniqueness assume $0=\Sigma_{\nu=\nu_0}^{\langle l/2\rangle}d^\nu H_{l-2\nu}$, where $0\leqslant\nu_0\leqslant\langle l/2\rangle$ and $H_{l-2\nu_0}\neq0$. Then $H_{l-2\nu_0}=dq$, where q is homogeneous of degree $k\geqslant0$. Applying the Laplace operator n times to this equation, where $n\leqslant1+k/2$, and using the Euler equation each time one gets $\alpha_n\Delta^{n-1}q+d\Delta^nq=0$, where $\alpha_1=2m+4k$ and $\alpha_n=\alpha_{n-1}+2m+4(k-2(n-1))$, $\langle1+k/2\rangle$ is the lowest integer $n \geqslant 0$ such that $\Delta^n q = 0$. However, it contradicts the foregoing equation. This proves uniqueness and also the linear independence of $(d^\nu H_{1-2\nu})$. The dimension of the linear space of all homogeneous harmonic polynomials of m variables and of degree k is equal to $d(k,m) := (2k+m-2)[(k+m-3)!/(k!(m-2)!)]$, see Ref. 15; Sec. 11.2(2). Therefore the combinations (15) form a linear subspace of dimension

$$\sum_{\nu=0}^{\langle l/2 \rangle} d(l-2\nu,\ m) = \binom{l+m-1}{l}.$$

This is exactly the dimension of the space of all homogeneous polynomials of m variables and of degree l, see Ref. 15; Sec. 11.2(5). ■

In the case $m = 3$ we take $(H_l^\lambda)_{\lambda \in [l]}$ [see Ref. 15; Sec. 11.5.1(2)] as a basis of the space of all homogeneous harmonic polynomials of degree l. H_l^λ transform properly under SU(2):

$$n_l^\lambda H_l^\lambda(B\cdot\mathbf{p}) = \sum_{\mu \in [l]} D_{\lambda\mu}^{(l)}(B) n_l^\mu H_l^\mu(\mathbf{p}), \tag{16}$$

where $n_l^\lambda = (-1)^\lambda((l-\lambda)!(l+\lambda)!)^{1/2}$. Furthermore, $n_l^\lambda H_l^\lambda(p\mathbf{e}_3) = (-2)^l l! p^l \delta_{0\lambda}$. Now, we are in a position to determine the invariant elements of $\mathscr{A}_k^j$.

Lemma 4: Let l run through $\{|k-j|, |k-j|+1, \ldots, k+j-1, k+j\}$, let $r \in [k]$, $s \in [j]$, and let $\left(\begin{smallmatrix} j & r & l \\ -s & r & 0 \end{smallmatrix}\right)$ denote the Wigner $3-j$ symbol. It is $\Phi \in \mathscr{A}_0$ if and only if there are entire functions $f_l: \mathbb{C} \to \mathbb{C}$ such that

$$\int d\mu(\mathbf{z}) |z_1^2 + z_2^2 + z_3^2|^l |f_l(z_1^2 + z_2^2 + z_3^2)|^2 < \infty$$

and

$$\Phi(\mathbf{p}) = D^{(k)}(B(\mathbf{p}))\Phi(p\mathbf{e}_3) D^{(j)}(B(\mathbf{p})^{-1}),$$

where $p > 0$ and

$$\Phi_{rs}(p\mathbf{e}_3) = \sum_l (-1)^{j-s}\sqrt{2l+1}\begin{pmatrix} j & k & l \\ -s & r & 0 \end{pmatrix} p^l f_l(p^2). \tag{17}$$

The f_l are uniquely determined by Φ since

$$p^l f_l(p^2) = \sum_s (-1)^{j-s}\sqrt{2l+1}\begin{pmatrix} j & k & l \\ -s & s & 0 \end{pmatrix} \Phi_{ss}(p\mathbf{e}_3), \tag{18}$$

where $s \in [k] \cap [j]$. Moreover, Φ is exponentially bounded if and only if the $z \mapsto f_l(z^2)$ are exponentially bounded, and Φ is a polynomial if and only if the f_l are polynomials.

Proof: In order to determine $\pi\mathscr{P}$ it suffices to consider polynomials Φ of the form $\Phi_{\kappa\sigma}(\mathbf{z}) = \delta_{\kappa\alpha}\delta_{\sigma\beta}(z_1^2 + z_2^2 + z_3^2)^n n_l^\lambda H_l^\lambda(\mathbf{z})$, where $\alpha \in [k]$, $\beta \in [j]$, $\lambda \in [l]$, $l, n \in \{0,1,2,\ldots\}$, since according to Lemma 3 fol. they generate $\mathscr{P}$ by finite linear combinations. Because of (16) fol.

$$(\xi(B)\Phi)_{rs}(p\mathbf{e}_3) = (-2)^l l! D_{\beta s}^{(j)}(B^{-1}) D_{r\alpha}^{(k)}(B) D_{\lambda 0}^{(l)}(B^{-1}) p^{2n+1}$$

and hence

$$(\pi\Phi)_{rs}(p\mathbf{e}_3) = (-2)^l l! \begin{pmatrix} j & k & l \\ \beta & -\alpha & \lambda \end{pmatrix} (-1)^{\alpha+s} \begin{pmatrix} j & k & l \\ s & -r & 0 \end{pmatrix} p^{2n+1}$$

[see Ref. 16; (4.6.2)]. This proves (17) in the case of $\Phi \in \mathscr{P}$, where the f_l are polynomials. The inversion formula (18) follows from Ref. 16 (3.7.8).

Since each subspace consisting of all polynomials of same degree is invariant under ξ, $\pi\mathscr{P} \subset \mathscr{P}$ holds. Hence $\pi\mathscr{P} \subset \mathscr{P} \cap \mathscr{A}_0$ and $\mathscr{P} \cap \mathscr{A}_0$ is dense in $\mathscr{A}_0$.

Consider the case that f_m is the constant polynomials $\delta_{lm} 1$. Then $\varphi^{(l)}$,

$$\varphi_{rs}^{(l)}(\mathbf{p}) := p^l \sum_{\kappa,\sigma} D_{r\kappa}^{(k)}(B(\mathbf{p})) \times \sqrt{2l+1}(-1)^{j-\sigma}\begin{pmatrix} j & k & l \\ -\sigma & \kappa & 0 \end{pmatrix} D_{\sigma s}^{(j)}(B(\mathbf{p})^{-1}), \tag{19}$$

are homogeneous polynomials of degree l belonging to $\mathscr{A}_0$ and, therefore, every Φ of the form

$$\Phi(\mathbf{z}) = \sum_l \varphi^{(l)}(\mathbf{z}) f_l(z_1^2 + z_2^2 + z_3^2), \tag{20}$$

where f_l are entire functions satisfying

$$\int d\mu(\mathbf{z}) |z_1^2 + z_2^2 + z_3^2|^l |f_l(z_1^2 + z_2^2 + z_3^2)|^2 < \infty,$$

belongs to $\mathscr{A}_0$. We have to see that a limit in $\mathscr{A}$ of functions of the form (20) is still of this form. By means of

$$\operatorname{tr}(\varphi^{(l)}(\mathbf{z}^*)^+ \varphi^{(m)}(\mathbf{z})) = (z_1^2 + z_2^2 + z_3^2)^l \delta_{lm}, \tag{21}$$

which follows from Ref. 16; (3.7.8), Eq. (20) is inverted

$$(z_1^2 + z_2^2 + z_3^2)^l f_l(z_1^2 + z_2^2 + z_3^2) = \operatorname{tr}(\varphi^{(l)}(\mathbf{z}^*)^+ \Phi(\mathbf{z})). \tag{22}$$

This indeed shows that (20) is maintained under limits, since convergence in $\mathscr{A}$ implies pointwise convergence of the Taylor coefficients and pointwise convergence.

The last assertion is verified directly using (20), (22). ■

The Theorem is proved combining Lemmas 2 and 4.

III. APPLICATIONS

We determine the possible causal time evolutions for some physical systems. The systems are characterized by the values of the multiplicities ν_j. The class of factorizing causal time evolutions is determined completely. This includes the determination of the causal time evolutions for all simple systems of definite spin.

A. Spin-0 systems ($\nu_j = 0$ if $j \neq 0$)

We consider the case $\nu_0 < \infty$. Then in the helicity representation, a causal transformation T is a $\nu_0 \times \nu_0$ matrix valued function depending quadratic on $p > 0$ and admitting an entire exponentially bounded continuation, see (5).

Lemma 5: If T is unitary then T is constant.

Proof: Pick $\iota, \kappa \in \{1, 2, \ldots, \nu_0\}$ and consider $f(z) := a_0^\varrho(z) a_0^\varrho(z^*)^*, z \in \mathbb{C}$, cf. (10). f is an entire exponentially bounded function satisfying (i) $f(z) = f(-z)$, (ii) $f(z^*)^* = f(z)$, and, because of (i) and the unitarity of T, (iii) $|f(x)| \leqslant 1$, $|f(iy)| \leqslant 1\ \forall\ x, y \in \mathbb{R}$. The Phragmen–Lindelöf theorem (Ref. 17; VI, 4.2) for the sector $\{z \in \mathbb{C} | 0 < \arg z < \pi/2\}$, combined with (i), (ii), yields f bounded by 1. By the Liouville theorem f is a constant c. If $c = 0$ then $a_0^\varrho = 0$. If $c \neq 0$ then a_0^ϱ is an entire function without zeros and, therefore, $a_0^\varrho = \exp g$, where g is entire. By Hadamard's factorization theorem

(Ref. 17; XI, 3.4) the exponential boundedness of a_0^p implies $g(z) = \alpha z + \beta$. Since a_0^p is even, $\alpha = 0$ follows. ■

By (8) fol., this implies that there are only trivial causal time evolutions for spin-0 systems.

B. Spin-$\frac{1}{2}$ systems ($\nu_j = 0$ if $j \neq \frac{1}{2}$)

We consider the cases $\nu_{1/2} = 1$ and $\nu_{1/2} = 2$.

In the case $\nu_{1/2} = 1$, $0 < p \mapsto T_{sp}$ is a scalar valued function admitting an entire exponentially bounded continuation, and $T_{-1/2,z} = T_{1/2,-z}$, see (5). If T is unitary, then $|T_{sp}| = 1$ $\forall$ $p > 0$, see (8), and by the factorization theorem $T_{sp} = \exp[i(\epsilon\alpha p + \beta)]$, where α,β are real constants and $\epsilon := \operatorname{sgn} s$. Therefore, the causal time evolution V is given by $V_{sp}(t) = \exp[i(\epsilon\alpha p + \beta)t]$. β fixes the value of the energy at $p = 0$ and gives rise to the trivial part of the time evolution. $|\alpha|$ is the velocity of propagation. $\beta = 0$ and $\alpha = 1$ yields the relativistic time evolution of the neutrion-antineutrino system.

In the case $\nu_{1/2} = 2$, $0 < p \mapsto T_{sp}$ is a 2×2 matrix valued function admitting an entire exponentially bounded continuation, and $T_{-1/2,z} = T_{1/2,-z}$, see (5). Let $t \mapsto V(t)$ be a unitary group of causal transformations.

Lemma 6: There is an entire 2×2 matrix valued function $z \mapsto H_{sz}$ such that $(H_{sz})^+ = H_{sz^*}$ and $V_{sz}(t) = \exp(itH_{sz})$. Of course, $H_{-1/2,z} = H_{1/2,-z}$.

Proof: We omit the index s. By analyticity, V_z is a one parameter group for all $z \in \mathbb{C}$. Let iH_z be its generator. Fix $w \in \mathbb{C}$. There is $t > 0$ so small that $\|V_z(t) - 1\| < 1$ for all z in some neighborhood of w. There $itH_z = \ln V_z(t)$, where $\ln X := -\sum_{n=1}^{\infty}(1/n)(1 - X)^n$, which proves analyticity. $(H_z)^+ = H_{z^*}$ follows from $(H_p)^+ = H_p, p > 0$, by analyticity. ■

Applying a preceding argument, $\det V_{sp}(t) = \exp[i(\epsilon\tilde{\alpha}p + \tilde{\beta})t]$ follows, where $\tilde{\alpha},\tilde{\beta}$ are real constants and $\epsilon = \operatorname{sgn} s$. Hence

$$H_{1/2,p} = (\tilde{\alpha}p + \tilde{\beta})1_2 + \begin{bmatrix} a(p) & b(p) \\ b(p)^* & -a(p) \end{bmatrix},$$

$$H_{-1/2,p} = H_{1/2,-p}, \tag{23}$$

where a,b are entire and $a(z^*)^* = a(z)$. Now the consequences of the exponential boundedness of $z \mapsto V_{sz}(t)$ have to be studied.

Considering the eigenvalues of $V_{sz}(t)$ it follows that $\operatorname{Im}(\lambda)$ is linearly bounded, where $\lambda(z)$ is any square root of $a(z)^2 + b(z)b(z^*)^*$. Hence $\operatorname{Re}(-\lambda^2)$ is quadratically bounded from above. By the Borel–Caratheodory theorem[18] and the Liouville theorem this implies that λ^2 is a quadratic polynomial.

So the spectrum of the energy is given by

$$\tilde{\alpha}p + \tilde{\beta} \pm \sqrt{(\alpha p + \beta)^2 + \gamma^2}, \tag{24}$$

where α, β, and γ are real constants.

$$V_{1/2,(\cdot)}(t) = e^{i[\tilde{\alpha}(\cdot) + \tilde{\beta}]t}\begin{pmatrix} \cos\lambda t + ia\dfrac{\sin\lambda t}{\lambda} & ib\dfrac{\sin\lambda t}{\lambda} \\ i\bar{b}\dfrac{\sin\lambda t}{\lambda} & \cos\lambda t - ia\dfrac{\sin\lambda t}{\lambda} \end{pmatrix},$$

$$V_{-1/2,z} = V_{1/2,-z}, \tag{25}$$

where $\bar{b}(z) := b(z^*)^*$, shows that a and b must be exponentially bounded. Moreover, by the Paley–Wiener theorem, V is a causal time evolution if and only if $\delta(a) = \delta(b) = 0$, where

$$\delta(f) := \inf\{\delta > 0 \mid \exists\ c > 0: |f(z)| \leqslant c\exp(\delta|z|)\ \forall z \in G\}$$

for any function $f: G \to \mathbb{C}, G \subset \mathbb{C}$. Otherwise V describes a propagation with a limiting velocity which is superposed by an instantaneous one of a finite limiting range. It could serve as a model for macrocausality in the absence of microcausality. An example is given by $a(z) = \cos z$, $b(z) = e^{irz}\sin z$ for r real. The example $\tilde{\alpha} = \tilde{\beta} = 0$, $a(z) = z$, and $b(z) = me^{irz}$, $m > 0, r \in \mathbb{R}\setminus\{0\}$, is remarkable, since it is noncausal and, nevertheless, it yields a relativistic time evolution, namely the Dirac one. (However, by a unitary transformation leaving U and V invariant, E can be replaced by a unique $\tilde{E}$ such that V becomes causal, cf. Lemma 7 fol.). Here, microacausality manifests itself by an instantaneous propagation beyond the light cone and within a cone around about it at a fixed distance. The particular time the system leaves the light cone depends on its initial state.

Lemma 7: If $\delta(a) = 0$, then a is linear and there is $r \in \mathbb{R}$ such that $b\exp[ir(\cdot)]$ is linear. Clearly, if $\delta(a) > 0$ then $\delta(b) > 0$.

Proof: In the sequel, we refer to the fact (*) that any analytic function on the upper half-plane which is bounded on the real axis and satisfies $\delta(f) = 0$, is bounded (cf. Ref. 17; VI, 4.4).

We already know that $a^2 + b\bar{b}$ is either constant or equal to $z \mapsto \eta(z - w)(z - w^*)$, where $\eta > 0$ and $w \in \mathbb{C}$. Particularly, it is $\delta(a^2) = \delta(b\bar{b}) = 0$. In the first case (*) applies to a^2 and $b\bar{b}$. Therefore a is constant and, by the argument at the end of the proof of Lemma 5, b is a constant times $\exp[ir(\cdot)]$ with r real. In the second case we assume first that w is real. Then $a(w) = 0 = b(w)$, and $z \mapsto a(z)/(z - w)$ and $z \mapsto b(z) \times \exp[-ir(\cdot)]/(z - w)$ are constant by the argument used in the first case. Now let $\operatorname{Im}(w) > 0$. Then (*) applies to $f(z) := [b(z)\bar{b}(z) - b(w)\bar{b}(w)]/(z - w)(z - w^*)$ for $\operatorname{Im}(z) \geqslant 0$, which shows that $b\bar{b}$ is bounded quadratically on the upper half-plane. The same result follows analogously for the lower half-plane. Thus by the Liouville theorem, $b\bar{b}$ is a polynomial of degree $\leqslant 2$. $b\bar{b}$ is either constant or equal to $z \mapsto \eta(z - w)(z - w^*)$, where $\eta > 0$ and $w \in \mathbb{C}$. In the first case, b is a constant times $\exp[ir(\cdot)]$, as we already know. The second case reduces to the first one by considering $b(z)/(z - w)$ instead of $b(z)$. Moreover, a^2 is a polynomial of degree $\leqslant 2$. Hence, by repeating the foregoing argument, a is linear. ■

Thus V is causal if and only if a and b are linear. It is causal and relativistic if and only if, additionally, the spectrum (24) is relativistic, i.e., equal to $\pm(p^2 + m^2)^{1/2}, m \geqslant 0$. If $m > 0$, up to a unitary equivalence transformation leaving (U,E) invariant, the Diract system is unique. If $m = 0$, in addition there is the doubled neutrino–antineutrino system. The first is given by $\tilde{\alpha} = \tilde{\beta} = 0$ and $\lambda(z)^2 = z^2$, the latter by $\tilde{\alpha} = \pm 1$ and $\tilde{\beta} = a = b = 0$. The other cases yield causal and not relativistic time evolutions. Because of their Euclidean invariance they could serve to describe motions relative to a homogeneous medium.

C. Factorizing time evolutions

In this section we specifically discuss the unitary one parameter groups V of factorizing causal transformations

$V_{sp}(t) = V_s(t)\exp[it\varphi_s(p)]$, where V_s is a unitary group acting on $\mathscr{V}_s$, and φ_s: $]0,\infty[\to\mathbb{R}$ is measurable. Clearly, the time evolutions of a simple spin-j system, i.e., $\nu_k = \delta_{kj}$, are of this kind.

Lemma 8: Let V factorize. Then (i) each V_s leaves the spin eigenspaces invariant and its reduction to them does not depend on s, cf. (8) fol., and (ii) there is a real constant δ such that $\varphi_s(p) = 2\delta sp$ almost everywhere. Conversely, every V of this kind is causal.

Proof: It follows from (5) and the orthogonality relation [Ref. 16; (3.7.8)] that $\Sigma_s |a_s^{t\varphi}(p)|^2 = \Sigma_l p^{2l} |f^t_{\rho l}(p^2)|^2$. Since the left-hand side does not depend on p, it vanishes identically if $l>0$, i.e., $j\neq k$. Moreover, according to (5), we may assume that $p\mapsto\exp[it\varphi_s(p)]$ admits an exponentially bounded entire continuation. By the argumentation at the end of the proof of Lemma 5, $\exp[it\varphi_s(p)] = \exp(i\delta_{ts}\, p)$ and hence $\delta_{ts} = t\delta_s$, where δ_s is real. Therefore $a_s^{t\varphi}(0) = (-1)^{2j}(2j+1)^{-1/2} f^t_{\rho 0}(0)\delta_{jk}$ which is independent of s. This proves (i).

Let t and j be fixed and let C denote the orthogonal $(2j+1)\times(2j+1)$ matrix

$$C_{ls} := (-1)^{j-s}\sqrt{2l+1}\begin{pmatrix} j & j & l \\ -s & s & 0\end{pmatrix}.$$

According to (5) it remains to be solved that $p^l f_l(p^2) = \Sigma_s C_{ls}\exp(it\delta_s p)$.

Obviously, the right-hand side, denoted by $\varphi_1(p)$, admits an entire exponentially bounded continuation. Since $C_{l,-s} = (-1)^l C_{ls}$, $\delta_{-s} = -\delta_s$ and $\varphi_l(-z) = (-1)^l\varphi_l(z)$ follow. Therefore $\varphi_l(z)$ is of the form $z^l f_l(z^2)$ if and only if $z=0$ is a zero of φ_l of multiplicity l. This means that δ_s solve the equations

$$\sum_{s\in[j]} C_{ls}(\delta_s)^\lambda = 0, \quad \forall\, l\in\{1,\dots,2j\}, \lambda\in\{0,1,\dots,l-1\}. \tag{26}$$

If $j\in\{0,\frac{1}{2}\}$ then $\delta_s = 2\delta s$ (see below) because of $\delta_{-s} = -\delta_s$. Let j be $\geqslant 1$. Then the case $\lambda = 1$ in (26) means that the vector (δ_s) is orthogonal to the lines $l = 2,\dots,2j$ of C. Therefore $\delta_s = \xi C_{0s} + \eta C_{1s}$ where ξ,η are some real coefficients. Since $\delta_{-s} = -\delta_s$ and $C_{l,-s} = (-1)^l C_{ls}$, $\xi = 0$ follows and hence $\delta_s = \eta C_{1s} = 2\delta s$, where δ is a real constant not depending on s [e.g., Ref. 19; (C.27)]. Thus it remains to be verified that $\delta_s = s$ satisfies (26). First, we note that (26) is true in the case of $\lambda = 0$. Indeed, this follows from the relation in Ref. 16; (3.7.13), solving for C_{ls}. Secondly, we need the recursion relation in Ref. 19;(C.20): $sC_{ls} = d_{l+1}C_{l+1,s} + d_l C_{l-1,s}$, where d_l is a constant not depending on s. Now (26) is proved by induction on λ. ■

We encounter the situation that the limiting velocity $2|\delta|\,|s|$ is coupled to helicity. In particular, this means that a factorizing causal time evolution is not relativistic except in the case of a multiple neutrino-antineutrino system, and therefore the neutrino-antineutrino system is the only simple spin j system admitting a relativistic causal time evolution.

ACKNOWLEDGMENTS

The author is indebted to Dr. U. Mutze and W. Satzger for many valuable discussions.

[1] A. S. Wightman, "On the Localizability of Quantum Mechanical Systems," Rev. Mod. Phys. **34**, 845–872 (1962).

[2] H. J. Borchers, Commun. Math. Phys. **4**, 315–323 (1967), Theorem, III.1.

[3] G. C. Hegerfeldt, "Remark on Causality and Particle Localization," Phys. Rev. D **10**, 3320–3321 (1974); B. K. Skagerstam, "Some Remarks concerning the Question of Localization of Elementary Particles," Int. J. Theor. Phys. **15**, 213–230 (1976); G. C. Hegerfeldt and N. M. Ruijsenaars, "Remarks on Causality, Localization, and Spreading of Wave Packets," Phys. Rev. D **22**, 377–384 (1980).

[4] B. Jancewicz, "Operator Density Current and Relativistic Localization Problem," J. Math. Phys. **18**, 2487–2494 (1977), Sec. 1.

[5] S. Schlieder, "Zum kausalen Verhalten eines relativistischen quanten-mechanischen Systems," in *Quanten und Felder*, edited by H. P. Dürr (Vieweg, Braunschweig, 1971).

[6] S. N. M. Ruijsenaars, "On Newton-Wigner Localization and Superluminal Propagation Speeds," Ann. Phys. (NY) **137**, 33–43 (1981).

[7] W. Cegla and A. Z. Jadczyk, "Causal Logic of Minkowski Space," Commun. Math. Phys. **57**, 213–217 (1977); W. Cegla and A. Z. Jadczyk, "States on Causal Logic," Lett. Math. Phys. **3**, 109–112 (1979).

[8] W. Cegla and B. Jancewicz, "Representation of Relativistic Causality Structure by an Operator Density Current," Rep. Math. Phys. **11**, 53–63 (1977); A. Borowiec and A. Jadczyk, "Covariant Representation of the Causal Logic," Lett. Math. Phys. **3**, 255–257 (1979).

[9] U. Cattaneo, "On Mackey's Imprimitivity Theorem," Comment. Math. Helv. **54**, 629–641 (1979).; D. P. L. Castrigiano, R. W. Henrichs, "Systems of Covariance and Subrepresentations of Induced Representations," Lett. Math. Phys. **4**, 169–175 (1980).; S. T. Ali, "A Geometrical Property of POV-Measures and Systems of Covariance" to appear in *Lecture Notes in Mathematics* (Springer, Berlin) (to be published).

[10] D. P.L. Castrigiano, "On Euclidean Systems of Covariance for Massless Particles," Lett. Math. Phys. **5**, 303–309 (1981).; "Position Operators as Integrals with respect to Euclidean Systems of Covariance," Lett. Math. Phys. **6**, 261–269 (1982).

[11] K. Kraus, "Position Observables of the Photon," in *The Uncertainty Principle and Foundations of Quantum Mechanics*, edited by W. C. Price and S. S. Chissick (Wiley, London, 1977), pp. 293–320.

[12] B. A. Fuks, *Introduction to the Theory of Analytic Functions of Several Complex Variables* (Am. Math. Soc., Providence, RI, 1963), Th. 26.11.

[13] V. Bargmann, "On the Representations of the Rotation Group," Rev. Mod. Phys. **34**, 829–845 (1962).

[14] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Heidelberg, 1976), Sec. 9.2. Th.1.

[15] Erdélyi *et al.*, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. 2.

[16] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957).

[17] J. B. Conway, *Functions of One Complex Variable* (Springer, New York, 1978).

[18] E. C. Titchmarsh, *The Theory of Functions* (Oxford U.P., New York, 1952), Sec. 5.5.

[19] A. Messiah, *Quantum Mechanics II* (North-Holland, Amsterdam, 1964), Appendix C, Sec. 5.

Bradyons and tachyons in an expansion of the spherical wave $\exp[i(\omega/c)(r-ct)]/r$ into cylindrical waves

J. L. Agudin[a] and A. M. Platzeck[a]
Observatorio Astronómico, Universidad Nacional de La Plata, La Plata, Argentina

(Received 8 March 1983; accepted for publication 6 May 1983)

A resolution of the spherical wave $\exp[i(\omega/c)(r-ct)]/r$ into (evanescent plus outgoing) cylindrical waves, valid everywhere except at $\rho=0$ is given. This expansion, being the $\exp[i(\omega/c)(r-ct)]/r$ solution of an inhomogeneous wave equation, turns out to be wholly equivalent to the expansion of the corresponding source into uniformly moving δ-line densities, which includes bradyonic as well as tachyonic contributions. When applied to the electromagnetic potentials produced by a charged particle in an arbitrary one-dimensional motion, it provides an alternative derivation of the decomposition of the fields of the charge into their bradyonic and tachyonic parts.

PACS numbers: 14.80.Pb, 11.10. – z

To our knowledge, the only attempt to represent a spherical wave as a sum of cylindrical waves goes back to Sommerfeld.[1] In solving the problem of a vertical dipole located at the surface of a plane, finitely conducting earth, Sommerfeld succeeded in decomposing the primary excitation $\exp[i(\omega/c)(r-ct)]/r$ into two sets of cylindrical waves, one valid for $z<0$ and the other for $z>0$, i.e.,

$$\frac{e^{i(\omega/c)(r-ct)}}{r}=\int_0^\infty \frac{J_0(\lambda\rho)\exp\left[\pm\sqrt{\lambda^2-(\omega^2/c^2)}z\right]e^{-i\omega t}}{\sqrt{\lambda^2-\omega^2/c^2}}\lambda\, d\lambda. \quad (1)$$

Sommerfeld's expansion, although adequate to solve the problem at hand, has the drawback of having the region of nonvalidity (the plane $z=0$), not invariant under z-translations. It is the aim of the present paper to show that, starting from well-known integral representations of cylindrical functions, a resolution of the spherical wave $\exp[i(\omega/c)(r-ct)]/r$ into (evanescent plus outgoing) cylindrical waves, valid everywhere except at $\rho=0$, can be obtained, in which the total cylindrical symmetry (rotation plus translation) is preserved. This expansion, being $\exp[i(\omega/c)(r-ct)]/r$ solution of an *inhomogeneous* wave equation, turns out to be wholly equivalent to the expansion of the *corresponding source* into uniformly moving δ-line densities, which includes bradyonic as well as tachyonic contributions. When applied to the (Fourier analyzed in time) electromagnetic potentials produced by a charged particle in an arbitrary one-dimensional motion, it provides an alternative derivation of the decomposition of the fields of the charge into their bradyonic and tachyonic parts recently given by the present writers.[2]

Let us consider the two following definite integrals[3]: $\rho>0$,

$$\int_0^\infty \frac{\sin\left[(\omega/c)\sqrt{\rho^2+u^2}\right]}{\sqrt{\rho^2+u^2}}\cos ku\, du=\begin{cases}(\pi/2)\,J_0(\sqrt{\omega^2/c^2-k^2}\rho) & \text{for } \omega/c>k>0,\\ 0 & \text{for } k>\omega/c>0,\end{cases}$$

$$\int_0^\infty \frac{\cos\left[(\omega/c)\sqrt{\rho^2+u^2}\right]}{\sqrt{\rho^2+u^2}}\cos ku\, du=\begin{cases}-(\pi/2)N_0(\sqrt{\omega^2/c^2-k^2}\rho) & \text{for } \omega/c>k>0,\\ K_0(\sqrt{k^2-\omega^2/c^2}\rho) & \text{for } k>\omega/c>0.\end{cases} \quad (2)$$

From these expressions and using $H_0^{\binom{1}{2}}=J_0\pm iN_0$, it is rather straightforward to write $\rho>0$, $k\gtrless 0$,

$$\int_{-\infty}^{\infty}\frac{e^{i(\omega/c)\sqrt{\rho^2+u^2}}}{\sqrt{\rho^2+u^2}}e^{-iku}\,du=\begin{cases}2K_0(\sqrt{k^2-\omega^2/c^2}\rho) & \text{for } \omega^2/c^2<k^2,\ \ \omega\gtrless 0,\\ i\pi H_0^{(1)}(\sqrt{\omega^2/c^2-k^2}\rho) & \text{for } \omega^2/c^2>k^2,\ \ \omega>0,\\ -i\pi H_0^{(2)}(\sqrt{\omega^2/c^2-k^2}\rho) & \text{for } \omega^2/c^2>k^2,\ \ \omega<0.\end{cases} \quad (3)$$

By multiplying both members of Eq. (3) by e^{ikz}, integrating from $k=-\infty$ to $k=\infty$, taking into account the properties of the δ-function, and introducing the time factor $e^{-i\omega t}$, we readily arrive at $\rho>0$, $\omega\gtrless 0$,

$$\frac{e^{i(\omega/c)(r-ct)}}{r}=\frac{1}{\pi}\int_{k^2>\omega^2/c^2}K_0\left(\sqrt{k^2-\frac{\omega^2}{c^2}}\rho\right)e^{i(kz-\omega t)}dk\pm\frac{i}{2}\int_{k^2<\omega^2/c^2}H_0^{\binom{1}{2}}\left(\sqrt{\frac{\omega^2}{c^2}-k^2}\rho\right)e^{i(kz-\omega t)}\,dk, \quad (4)$$

where $r=\sqrt{\rho^2+z^2}$.

[a] Fellow of CONICET, Argentina.

 0022-2488/84/041122-03$02.50

Let us discuss this result. For all $k^2 > (\omega/c)^2$, and taking $\kappa\rho = \sqrt{k^2 - \omega^2/c^2}\rho \gg 1$, each term of this expansion behaves asymptotically as

$$K_0\left(\sqrt{k^2 - \frac{\omega^2}{c^2}}\rho\right)e^{i(kz-\omega t)} \simeq \frac{e^{-\kappa\rho}}{\sqrt{\kappa\rho}}e^{i(kz-\omega t)}, \qquad (5a)$$

which is an *evanescent* cylindrical wave traveling with the phase velocity $v_{\rm ph} = \omega/k$ along the z axis. For all $k^2 < (\omega/c)^2$, and taking $k_\rho\,\rho = \sqrt{\omega^2/c^2 - k^2}\rho \gg 1$, each term of this expansion behaves asymptotically as

$\omega \gtrless 0$,

$$H_0^{\binom{1}{2}}\left(\sqrt{\frac{\omega^2}{c^2} - k^2}\rho\right)e^{i(kz-\omega t)} \simeq \frac{e^{\pm i(k_\rho\rho \pm kz - |\omega| t)}}{\sqrt{k_\rho\,\rho}}, \qquad (5b)$$

which is an *outgoing* cylindrical wave traveling with the phase velocity $v_{\rm ph} = \omega/\sqrt{k_\rho^2 + k^2} = c$ in the direction of the "asymptotic wave vector"

$$\mathbf{k}_a = k_\rho\,\check{\rho} \pm k\check{z}. \qquad (6)$$

Hence what we have obtained through Eq. (4) is an expansion of the *outgoing* spherical wave into *evanescent plus outgoing* cylindrical waves, valid everywhere except at $\rho = 0$.

Now, the spherical wave satisfies

$$\Box\frac{e^{i(\omega/c)(r-ct)}}{r} \equiv \left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\frac{e^{i(\omega/c)(r-ct)}}{r} = -4\pi\delta(\mathbf{r})e^{-i\omega t}, \qquad (7)$$

and thus we may ask for the expansion of the inhomogeneity $-4\pi\delta(\mathbf{r})e^{-i\omega t}$ that matches with the expansion of $\exp[i(\omega/c)(r-ct)]/r$ given above. By applying a D'Alembert operator to both sides of Eq. (4), using Eq. (7), and taking into account that[4]

$$\Box\left[K_0\left(\sqrt{k^2 - \frac{\omega^2}{c^2}}\rho\right)e^{i(kz-\omega t)}\right] = -\frac{2\delta(\rho)}{\rho}e^{i(kz-\omega t)}, \qquad (8a)$$

$$\Box\left[H_0^{\binom{1}{2}}\left(\sqrt{\frac{\omega^2}{c^2} - k^2}\rho\right)e^{i(kz-\omega t)}\right] = \pm i\frac{4}{\pi}\frac{\delta(\rho)}{\rho}e^{i(kz-\omega t)}, \qquad (8b)$$

we readily arrive at

$$-4\pi\delta(\mathbf{r})e^{-i\omega t} = -\frac{2}{\pi}\frac{\delta(\rho)}{\rho}\int_{k^2>\omega^2/c^2} e^{i(kz-\omega t)}\,dk - \frac{2}{\pi}\frac{\delta(\rho)}{\rho}\int_{k^2<\omega^2/c^2} e^{i(kz-\omega t)}\,dk, \qquad (9)$$

which gives the "source" $-4\pi\delta(\mathbf{r})e^{-i\omega t}$ as a sum of δ-line "source densities" $[-2\delta(\rho)/\pi\rho]e^{i(kz-\omega t)}$ moving at the constant velocities $v_{k\omega} = \omega/k$ along the z axis. The first integral, which extends over those k satisfying $k^2 > \omega^2/c^2$, comprises only the "bradyonic" components of the source, i.e., those δ-line source densities moving uniformly with $|v_{k\omega}| < c$. The elementary waves excited by these bradyonic components are the evanescent cylindrical waves $K_0(\sqrt{k^2 - \omega^2/c^2}\rho)e^{i(kz-\omega t)}$, cf. Eq. (8a). The second integral, which extends over those k satisfying $k^2 < \omega^2/c^2$, comprises only the "tachyonic" components of the source, i.e., those δ-line source densities with $|v_{k\omega}| > c$. The elementary waves excited by these tachyonic components are the outgoing cylindrical waves $H_0^{\binom{1}{2}}(\sqrt{\omega^2/c^2 - k^2}\rho)e^{i(kz-\omega t)}$, cf. Eq. (8b). By comparing Eqs. (4) and (9), and taking into account Eq. (7), we can conclude that our expansion of the spherical wave $\exp[i(\omega/c)(r-ct)]/r$ into evanescent plus outgoing cylindrical waves is wholly equivalent to the expansion of the corresponding source $-4\pi\delta(\mathbf{r})e^{-i\omega t}$ into bradyonic plus tachyonic δ-line source densities.[5] This equivalence now makes evident why the expansion given by Eq. (4) is not valid along the z axis: each of its terms is not a solution of the homogeneous wave equation, but of the inhomogeneous one, the source of which is singular at $\rho = 0$.

In order to illustrate the usefulness of the present expansion, we consider the electromagnetic potentials Φ and $\mathbf{A} = (0, 0, A_z)$ produced by a particle of charge q which is moving arbitrarily along the z axis. Within the Lorentz gauge, the scalar potential $\Phi(\mathbf{r}, t)$ satisfies

$$\Box\,\Phi(\mathbf{r}, t) = -4\pi\rho(\mathbf{r}, t) = -4\pi q\delta(x)\delta(y)\delta[z - z(t)], \qquad (10)$$

and there is a similar equation for $A_z(\mathbf{r}, t)$. By Fourier analyzing in time both $\Phi(\mathbf{r}, t)$ and $\rho(\mathbf{r}, t)$, we readily arrive at

$$\Phi(\mathbf{r}, t) = \int_{-\infty}^{\infty} d\omega \iiint_{-\infty}^{\infty} dx'\,dy'\,dz'\,G_\omega(\mathbf{r}, \mathbf{r}')\rho_\omega(\mathbf{r}')e^{-i\omega t}, \qquad (11)$$

where

$$\rho_\omega(\mathbf{r}) = \frac{q}{2\pi}\delta(x)\delta(y)\int_{-\infty}^{\infty}\delta[z - z(t')]e^{i\omega t'}dt' \qquad (12)$$

and $G_\omega(\mathbf{r}, \mathbf{r}')$ is the Green's function of Helmholz's equation, i.e.,

$$\nabla^2 G_\omega(\mathbf{r}, \mathbf{r}') + (\omega^2/c^2)G_\omega(\mathbf{r}, \mathbf{r}') = -4\pi\delta(\mathbf{r} - \mathbf{r}'). \qquad (13)$$

There are three different Green's functions, solution of Eq. (13), which are the outgoing spherical wave $\exp[i(\omega/c)|\mathbf{r} - \mathbf{r}'|]/|\mathbf{r} - \mathbf{r}'|$, the ingoing spherical wave $\exp[-i(\omega/c)|\mathbf{r} - \mathbf{r}'|]/|\mathbf{r} - \mathbf{r}'|$, and the standing spherical wave $\cos[(\omega/c)|\mathbf{r} - \mathbf{r}'|]/|\mathbf{r} - \mathbf{r}'|$.[6] If we are looking for the *retarded* scalar potential, we must choose that Green's function which, inserted in Eq. (11), gives outgoing waves, i.e.,

$$G_\omega(\mathbf{r}, \mathbf{r}') = e^{i(\omega/c)|\mathbf{r}-\mathbf{r}'|}/|\mathbf{r} - \mathbf{r}'|. \qquad (14)$$

By introducing Eqs. (12) and (14) into Eq. (11) and performing the integration over x' and y', we obtain

$$\Phi(\mathbf{r}, t) = \frac{q}{2\pi}\iiint_{-\infty}^{\infty} d\omega\,dz'\,dt'\,\delta[z' - z(t')]\,\frac{\exp i(\omega/c)[\sqrt{\rho^2 + (z-z')^2} - c(t-t')]}{\sqrt{\rho^2 + (z-z')^2}}, \qquad (15)$$

which gives the (retarded) potential $\Phi(\mathbf{r}, t)$ as the sum of the outgoing spherical waves $\exp i(\omega/c)[\sqrt{\rho^2 + (z-z')^2} - c(t-t')]/\sqrt{\rho^2 + (z-z')^2}$ excited by the charged particle at every point z' along the z axis.

Now a different representation of this potential can be found if we replace each outgoing *spherical* wave in Eq. (15) by its expansion into evanescent plus outgoing *cylindrical* waves given by Eq. (4). Thus we obtain

$$\begin{aligned}\Phi(\mathbf{r},t) = {}& \frac{q}{2\pi^2}\int_{-\infty}^{\infty} d\omega \iint_{-\infty}^{\infty} dz'\,dt' \int_{k^2>\omega^2/c^2} dk\, K_0\left(\sqrt{k^2-\frac{\omega^2}{c^2}}\rho\right) e^{i[k(z-z')-\omega(t-t')]}\delta[z'-z(t')] \\ & - \frac{iq}{4\pi}\int_{-\infty}^{0} d\omega \iint_{-\infty}^{\infty} dz'\,dt' \int_{k^2<\omega^2/c^2} dk\, H_0^{(2)}\left(\sqrt{\frac{\omega^2}{c^2}-k^2}\rho\right) e^{i[k(z-z')-\omega(t-t')]}\delta[z'-z(t')] \\ & + \frac{iq}{4k\pi}\int_{0}^{\infty} d\omega \iint_{-\infty}^{\infty} dz'\,dt' \int_{k^2<\omega^2/c^2} dk\, H_0^{(1)}\left(\sqrt{\frac{\omega^2}{c^2}-k^2}\rho\right) e^{i[k(z-z')-\omega(t-t')]}\delta[z'-z(t')].\end{aligned} \tag{16}$$

Due to the dissimilar asymptotic behavior of the cylindrical functions K_0 and $H_0^{\binom{1}{2}}$, we see that $\Phi(\mathbf{r},t)$ splits naturally into two terms, i.e.,

$$\Phi(\mathbf{r},t) = \Phi^B(\mathbf{r},t) + \Phi^T(\mathbf{r},t), \tag{17}$$

where $\Phi^B(\mathbf{r},t)$ contains only (all) the *evanescent* cylindrical waves and $\Phi^T(\mathbf{r},t)$ contains only (all) the *outgoing* cylindrical waves.[7] From Eq. (16), and calling, as in Ref. 2,

$$\wp(k,\omega) = \frac{1}{(2\pi)^2}\iint_{-\infty}^{\infty} dz'\,dt'\,\delta[z'-z(t')]e^{-i(kz'-\omega t')}, \tag{18}$$

we can finally write

$$\Phi^B(\mathbf{r},t) = 2q\int_{-\infty}^{\infty} dk \int_{-|k|c}^{|k|c} d\omega\, \wp(k,\omega) K_0\left(\sqrt{k^2-\frac{\omega^2}{c^2}}\rho\right)e^{i(kz-\omega t)}, \tag{19a}$$

$$\begin{aligned}\Phi^T(\mathbf{r},t) = {}& -i\pi q\int_{-\infty}^{\infty} dk \int_{-\infty}^{-|k|c} d\omega\, \wp(k,\omega) H_0^{(2)}\left(\sqrt{\frac{\omega^2}{c^2}-k^2}\rho\right)e^{i(kz-\omega t)} \\ & + i\pi q\int_{-\infty}^{\infty} dk \int_{|k|c}^{\infty} d\omega\, \wp(k,\omega) H_0^{(1)}\left(\sqrt{\frac{\omega^2}{c^2}-k^2}\rho\right)e^{i(kz-\omega t)},\end{aligned} \tag{19b}$$

which coincides with the resolution of the scalar potential of an accelerated charge into its bradyonic and tachyonic parts given in Ref. 2.[8] By carrying out a similar analysis for the vector potential $A_z(\mathbf{r},t)$, we get $A_z = A_z^B + A_z^T$, where $A_z^B[A_z^T]$ is obtained from Eq. (19a) [Eq. (19b)] by replacing $\wp(k,\omega)$ with $(\omega/ck)\wp(k,\omega)$. From $\mathbf{E}^{B,T} = -\nabla\Phi^{B,T} - (1/c)\partial\mathbf{A}^{B,T}/\partial t$ and $\mathbf{H}^{B,T} = \nabla\wedge\mathbf{A}^{B,T}$, it is straightforward to arrive at the decomposition of the fields of the charge into their bradyonic and tachyonic parts [cf. Ref. 2, Eqs. (8) and (9)]. Thus our development of $\exp[i(\omega/c)(r-ct)]/r$ into (evanescent plus outgoing) cylindrical waves given by Eq. (4) appears, at least for one-dimensional motions, as a useful device capable of uncovering the bradyonic and tachyonic components of the fields produced by the moving charge.

[1] A. Sommerfeld, Ann. Physik **28**, 665 (1909). See also J. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), pp. 573–576; G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U. P., Cambridge, 1960), p. 416, Eq. (4).

[2] J. L. Agudín and A. M. Platzeck, Phys. Lett. A **83**, 423 (1981). See also J. L. Agudín and A. M. Platzeck, "Tachyons and the radiation of an accelerated charge," Phys. Rev. D **26**, 1923 (1982).

[3] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980), p. 472, Eqs. 3.876(1, 2). Here we call ρ, k, ω/c, and u what Gradshteyn and Ryzhik call a, b, p, and x.

[4] See, for example, Ref. 2, Eqs. (5), (6), and (7).

[5] We note in passing that the separation of the source into its bradyonic and tachyonic parts is a covariant one, since a bradyonic (tachyonic) component remains always bradyonic (tachyonic) under any Lorentz transformation. This also applies to the expansion of the spherical wave into *evanescent* plus *outgoing* cylindrical waves. We exclude here the so-called "Superluminal Lorentz Transformations," for which the words bradyon and tachyon do not have an absolute meaning but only a relative one. It is worthwhile to note here that all these "superluminal theories" have been recently shown to be nonlinear, see J. L. Agudín and A. M. Platzeck, Phys. Lett. A **90**, 173 (1982).

[6] In saying that $\exp[\pm i(\omega/c)|\mathbf{r}-\mathbf{r}'|]/|\mathbf{r}-\mathbf{r}'|$, $\cos[(\omega/c)|\mathbf{r}-\mathbf{r}'|]/|\mathbf{r}-\mathbf{r}'|$ are spherical waves, one must bear in mind that the time factor $e^{-i\omega t}$ is always present.

[7] According to what has been discussed in Ref. 5, this separation is manifestly covariant.

[8] The unprejudiced reader could ask himself what result would have been obtained if, instead of our expansion, we had inserted Sommerfeld's representation in Eq. (15). The answer is that Sommerfeld's formula could hardly be applied here, since in replacing each spherical wave in Eq. (15) by its representation into cylindrical waves given by Eq. (1), we would have the whole space filled with planes at which the expansion is not valid.

The reduced Dirac Green function for the Coulomb potential

D. J. Hylton
Gibbs Laboratory, Physics Department, Yale University, New Haven, Connecticut 06520

(Received 19 July 1983; accepted for publication 23 September 1983)

The radial reduced Dirac Green function for the Coulomb potential is calculated for all bound states. The result is written in a form suitable for numerical calculations. The nonrelativistic limit agrees with the known radial reduced Schrödinger Green function for the Coulomb potential. Also obtained are the derivatives with respect to the energy of the relativistic and nonrelativistic radial solutions evaluated at any eigenvalue.

PACS numbers: 31.15. + q, 03.65.Ge

1. INTRODUCTION

In the standard Rayleigh–Schrödinger perturbation theory, the second-order energy shift or the first-order correction to the wavefunction can be evaluated with the aid of the reduced Green function[1,2]

$$G^R(\mathbf{x}_2, \mathbf{x}_1, E_n) = \sum_{k \neq n}{}' \frac{\Psi_k(\mathbf{x}_2)\Psi_k^+(\mathbf{x}_1)}{E_k - E_n}, \tag{1.1}$$

where the prime on the summation indicates an integration over the continuous part of the spectrum and $\Psi_k(\mathbf{x})$ is a normalized eigenfunction of the Hamiltonian with eigenvalue E_k. If one solves the Dirac equation for $\Psi_k(\mathbf{x})$ and E_k, then (1.1) defines the reduced Dirac Green function (RDGF), which is a 4×4 matrix.

The RDGF calculated as an expansion over the complete set of eigenfunctions of angular momentum is useful in applications. In the case where the perturbing potential commutes with angular momentum, this expansion is particularly useful since only one term will contribute provided the unperturbed wavefunctions are eigenfunctions of angular momentum. In Sec. 2, the Dirac Green function (DGF) for the Coulomb potential is constructed as such an expansion,[3,4] and the relation to the familiar sum over eigenfunctions is established.[1,2,5] In Sec. 3, the radial RDGF is explicitly calculated from the radial DGF and is written in the form of an integral representation. Other calculations either have written the RDGF in terms of an infinite series of Laguerre polynomials[6] or have calculated the reduced Green function for the second order Dirac equation in similar form.[7] In Ref. 7 the ground state was considered, and in Ref. 6 the RDGF was explicitly written down for the special cases $j = n - \frac{1}{2}$.

As a check of the present results, the nonrelativistic limit is obtained in Sec. 4 and found to be equal to the results of previous calculations of the radial reduced Schrödinger Green function for the Coulomb potential.[8,9] This limit yields a new way of writing the nonrelativistic result in terms of the derivative with respect to energy of the radial solutions. As another check, and also to verify that the results are suitable for numerical calculations, the second-order energy shift for a perturbing potential $\delta V = -\epsilon\alpha/r$ was numerically calculated. The numerical results agree with the exact results, and a summary of the calculation is given in Appendix B.

2. THE DIRAC GREEN FUNCTION

The DGF, $G(\mathbf{x}_2, \mathbf{x}_1, z)$, satisfies the equation[1]

$$[H(\mathbf{x}_2) - zI]G(\mathbf{x}_2, \mathbf{x}_1, z) = \delta^3(\mathbf{x}_2 - \mathbf{x}_1)I, \tag{2.1}$$

where I is the 4×4 identity matrix and $H(\mathbf{x})$, the Dirac Hamiltonian, is[10]

$$H(\mathbf{x}) = \boldsymbol{\alpha}\cdot\hat{x}(p_x + (i/x)\beta K) + V(\mathbf{x}) + \beta, \tag{2.2}$$

where

$$p_x = -\frac{i}{x}\frac{d}{dx}x, \quad K = \beta(\boldsymbol{\sigma}\cdot\mathbf{L} + 1), \tag{2.3}$$

and $\mathbf{L}$ is the orbital angular momentum operator. The units chosen are those for which $\hbar = c = m_e = 1$. The representation for the Dirac matrices is the same as in Ref. 11, and σ_i are the usual Pauli matrices.

The DGF, expanded over the complete set of spin-angular eigenfunctions of K and $J_3 = L_3 + \frac{1}{2}\sigma_3$, $\chi_\kappa^\mu(\hat{x})$ is[3,4,12,13]

$$G(\mathbf{x}_2,\mathbf{x}_1,z) = \sum_{\kappa,\mu}\begin{pmatrix} G_\kappa^{11}(x_2,x_1,z)\chi_\kappa^\mu(\hat{x}_2)\chi_\kappa^{\mu^+}(\hat{x}_1) & -iG_\kappa^{12}(x_2,x_1,z)\chi_\kappa^\mu(\hat{x}_2)\chi_{-\kappa}^{\mu^+}(\hat{x}_1) \\ iG_\kappa^{21}(x_2,x_1,z)\chi_{-\kappa}^\mu(\hat{x}_2)\chi_\kappa^{\mu^+}(\hat{x}_1) & G_\kappa^{22}(x_2,x_1,z)\chi_{-\kappa}^\mu(\hat{x}_2)\chi_{-\kappa}^{\mu^+}(\hat{x}_1)\end{pmatrix}. \tag{2.4}$$

The spin angular eigenfunctions are written in (A3). The coefficients of the expansion, $G_\kappa^{ij}(x_2, x_1, z)$, which are the components of the radial DGF, are μ-independent. This is because the effect of the Hamiltonian with a radial potential on $\chi_\kappa^\mu(\hat{x})$ is μ-independent (see Appendix A).

Inserting (2.4) into (2.1) gives

$$\begin{pmatrix} 1 + V(x_2) - z & -\frac{1}{x_2}\frac{d}{dx_2}x_2 + \frac{\kappa}{x_2} \\ \frac{1}{x_2}\frac{d}{dx_2}x_2 + \frac{\kappa}{x_2} & -1 + V(x_2) - z\end{pmatrix}\begin{pmatrix} G_\kappa^{11}(x_2,x_1,z) & G_\kappa^{12}(x_2,x_1,z) \\ G_\kappa^{21}(x_2,x_1,z) & G_\kappa^{22}(x_2,x_1,z)\end{pmatrix} = \frac{\delta(x_2 - x_1)}{x_2 x_1}I, \tag{2.5}$$

 0022-2488/84/041125-08$02.50

which is to be solved for $G_\kappa^{ij}(x_2,x_1,z)$. The solution can be constructed in terms of the solutions to the homogeneous equation

$$\begin{pmatrix} 1+V(x)-z & -\frac{1}{x}\frac{d}{dx}x+\frac{\kappa}{x} \\ \frac{1}{x}\frac{d}{dx}x+\frac{\kappa}{x} & -1+V(x)-z \end{pmatrix}\begin{pmatrix} \psi_1(x,z) \\ \psi_2(x,z) \end{pmatrix}=0, \tag{2.6}$$

as[2,3]

$$G_\kappa^{ij}(x_2,x_1,z)=\frac{1}{K(z)}\,[\psi_i^L(x_2,z)\psi_j^G(x_1,z)\theta(x_1-x_2) + \psi_i^G(x_2,z)\psi_j^L(x_1,z)\theta(x_2-x_1)], \tag{2.7}$$

where

$$K(z)=x^2[\psi_2^L(x,z)\psi_1^G(x,z)-\psi_2^G(x,z)\psi_1^L(x,z)]. \tag{2.8}$$

In (2.7) and (2.8), the superscript L means the solution regular at $x=0$, while the superscript G means the solution regular as $x\to\infty$. Using (2.6), one can show $dK(z)/dx$ vanishes.

For the Coulomb potential, $V(x)=-Z\alpha/x$, the solutions to (2.6) are[13]

$$\begin{aligned}
\psi_1^L(x,z)&=\frac{1+z}{x^{3/2}}\times\Big[(\lambda-\nu)M_{\nu-1/2,\lambda}(2cx)-\Big(\kappa-\frac{\gamma}{c}\Big)M_{\nu+1/2,\lambda}(2cx)\Big],\\
\psi_2^L(x,z)&=\frac{c}{x^{3/2}}\times\Big[(\lambda-\nu)M_{\nu-1/2,\lambda}(2cx)+\Big(\kappa-\frac{\gamma}{c}\Big)M_{\nu+1/2,\lambda}(2cx)\Big],\\
\psi_1^G(x,z)&=\frac{1+z}{x^{3/2}}\times\Big[\Big(\kappa+\frac{\gamma}{c}\Big)W_{\nu-1/2,\lambda}(2cx)+W_{\nu+1/2,\lambda}(2cx)\Big],\\
\psi_2^G(x,z)&=\frac{c}{x^{3/2}}\times\Big[\Big(\kappa+\frac{\gamma}{c}\Big)W_{\nu-1/2,\lambda}(2cx)-W_{\nu+1/2,\lambda}(2cx)\Big],
\end{aligned} \tag{2.9}$$

where

$$\gamma=Z\alpha,\quad \lambda=(\kappa^2-\gamma^2)^{1/2},\qquad c=(1-z^2)^{1/2},\quad \mathrm{Re}(c)>0,\quad \nu=\gamma z/c, \tag{2.10}$$

and $M_{\alpha,\beta}(x)$ and $W_{\alpha,\beta}(x)$ are Whittaker functions.[14] From the small x forms of the functions in (2.9),[14] one evaluates (2.7) in the limit $x\to 0$ to obtain

$$K(z)=4(1+z)c^2\Gamma(1+2\lambda)/\Gamma(\lambda-\nu),\quad \lambda>\tfrac{1}{2}. \tag{2.11}$$

Equations (2.7)–(2.11) determine $G_\kappa^{ij}(x_2,x_1,z)$ for the Coulomb potential, which can be related to the familiar sum over normalized eigenfunctions.[1] Since $G_\kappa^{ij}(x_2,x_1,z)$ is analytic everywhere in the complex z-plane except for branch points at $z=\pm 1$ [due to the $(1-z^2)^{1/2}$ dependence] and poles at

$$z=z_m=(\lambda+m)/[\gamma^2+(\lambda+m)^2]^{1/2},\quad m=0,1,2,\cdots, \tag{2.12}$$

the zeros of $K(z)$, one has from Cauchy's theorem

$$G_\kappa^{ij}(x_2,x_1,z)=\frac{1}{2\pi i}\int_c dy\,\frac{G_\kappa^{ij}(x_2,x_1,y)}{y-z},\qquad z\neq z_m \text{ and } z\notin\{[1,\infty),(-\infty,-1]\}. \tag{2.13}$$

The contour C circles the point z in the positive sense as shown in Fig. 1(a). The contour is deformed to the contour C' shown in Fig. 1(b). Since $\mathrm{Re}(c)>0$, $G_\kappa^{ij}(x_2,x_1,z)$ falls off exponentially for large z for fixed x_1 and x_2. Therefore, the contributions from the semicircles at infinity vanish, and (2.13) reduces to

$$\begin{aligned}
G_\kappa^{ij}(x_2,x_1,z)=\sum_m\frac{\psi_i^{m,\kappa}(x_2)\psi_j^{m,\kappa}(x_1)}{z_m-z}+\lim_{\epsilon\to 0^+}\Big\{\frac{1}{2\pi i}\int_1^\infty dE\\
\times\frac{G_\kappa^{ij}(x_2,x_1,E+i\epsilon)-G_\kappa^{ij}(x_2,x_1,E-i\epsilon)}{E-z}\\
+\frac{1}{2\pi i}\int_{-\infty}^{-1}dE\\
\times\frac{G_\kappa^{ij}(x_2,x_1,E+i\epsilon)-G_\kappa^{ij}(x_2,x_1,E-i\epsilon)}{E-z}\Big\}.
\end{aligned} \tag{2.14}$$

The first term in (2.14) is the sum of the residues of the poles, where $\psi_i^{m,\kappa}(x)$ is the normalized radial bound state wavefunction. The wavefunctions obtained in this way agree up to an undetermined phase with the standard results written in Appendix A. From the definitions[14]

$$M_{\alpha,\beta}(z)=e^{-z/2}z^{\beta+1/2}M(\beta-\alpha+\tfrac{1}{2},2\beta+1,z), \tag{2.15}$$

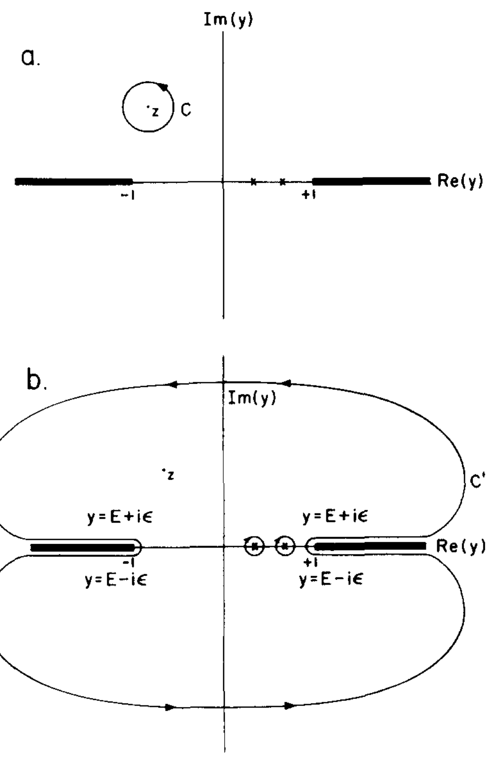


FIG. 1. (a) The contour C in Eq. (2.13). The x's denote bound state poles only two of which are shown. The bold lines denote cuts corresponding to the condition $\mathrm{Re}(c)>0$. (b) The new contour C'. This contour is used to obtain Eq. (2.14).

$$M(a,b,z)=\sum_{j=0}^{\infty}\frac{(a)_j z^j}{(b)_j\, j!}, \tag{2.16}$$

$$(a)_j=\frac{\Gamma(a+j)}{\Gamma(a)}, \tag{2.16$'$}$$

and the identities[14]

$$W_{\alpha,\beta}(z)=\frac{\Gamma(-2\beta)}{\Gamma(\frac{1}{2}-\beta-\alpha)}M_{\alpha,\beta}(z)+\frac{\Gamma(2\beta)}{\Gamma(\frac{1}{2}+\beta-\alpha)}M_{\alpha,-\beta}(z), \tag{2.17}$$

$$M(a,b,z)=e^z M(b-a,b,-z), \tag{2.18}$$

along with

$$c_{\pm}=pe^{\mp i\pi/2},\quad E>1, \tag{2.19a}$$

$$c_{\pm}=pe^{\pm i\pi/2},\quad E<-1, \tag{2.19b}$$

where $p=(E^2-1)^{1/2}$ and c_+ and c_- are the values of c above and below the real axis respectively; one finds for $E>1$

$$\psi_i^L(x,E+i\epsilon)=-e^{-i\pi(\lambda+1/2)}\frac{\lambda-i\gamma E/p}{\kappa+i\gamma/p}\psi_i^L(x,E-i\epsilon) \tag{2.20}$$

and

$$\begin{aligned}\psi_j^G(x,E\pm i\epsilon)&=\frac{-1}{\kappa\mp i\gamma/p}\frac{\Gamma(-2\lambda)}{\Gamma(-\lambda\mp i\gamma E/p)}\psi_j^L(x,E\pm i\epsilon)\\&\quad+A_j e^{\pm ipx}/x^{-3/2}(2px)^{-\lambda+1/2}\\&\quad\times e^{\mp i\pi(-\lambda+1/2)/2}\frac{\Gamma(2\lambda)}{\Gamma(\lambda\mp i\gamma E/p)}\\&\quad\times\left[\frac{\kappa\pm i\gamma/p}{\lambda\mp i\gamma E/p}M(-\lambda+1\mp i\gamma E/p,1-2\lambda,\mp 2ipx)\right.\\&\quad\left.+(-1)^{j+1}M(-\lambda\mp i\gamma E/p,1-2\lambda,\mp 2ip\gamma)\right],\end{aligned} \tag{2.21}$$

where $A_1=(1+E)$ and $A_2=pe^{\mp i\pi/2}$. Similar expressions are obtained for $E<-1$. From (2.20) and (2.21), one has

$$\begin{aligned}G_\kappa^{ij}(x_2,x_1,z)&=\sum_m\frac{\psi_i^{m,\kappa}(x_2)\psi_j^{m,\kappa}(x_1)}{z_m-z}\\&\quad+\int_1^\infty dE\,\frac{\psi_i^\kappa(x_2,E)\psi_j^{\kappa *}(x_1,E)}{E-z}\\&\quad+\int_{-\infty}^{-1}dE\,\frac{\psi_i^\kappa(x_2,E)\psi_j^{\kappa *}(x_1,E)}{E-z},\end{aligned} \tag{2.22}$$

where $\psi_i^\kappa(x\ E)$ are the continuum radial wavefunctions, which agree with the standard results as given in Appendix A. It should be noted that the second term in (2.21) cancels within the first integrand of (2.14), and similarly for the second integrand using the analogous expression for $E<-1$.

This explicitly relates the radial DGF for the Coulomb potential in (2.7) to the sum over eigenfunctions in (2.22). Furthermore, since (2.22) satisfies (2.5), this also demonstrates the completeness of the radial eigenfunctions.

3. THE RADIAL RDGF FOR THE COULOMB POTENTIAL

From (1.1) and (2.22), the components of the radial RDGF are

$$G_\kappa^{ijR}(x_2,x_1,E_n)=\lim_{z\to E_n}\left[G_\kappa^{ij}(x_2,x_1,z)-\frac{\psi_i^{m,\kappa}(x_2)\psi_j^{m,\kappa}(x_1)}{E_n-z}\right], \tag{3.1}$$

where

$$m=n-|\kappa|, \tag{3.2}$$

and E_n is the eigenvalue as given in Appendix A. To evaluate (3.1), one expands G_κ^{ij} in a Laurent series in powers of $(z-E_n)$ up to and including zero order. The pole cancels the second term on the right-hand side of (3.1), and the zero order term is the RDGF. It is convenient to define

$$\lambda-\nu=-m+\delta \tag{3.3}$$

and expand about $\delta=0$, since $(\lambda-\nu)=-m$ when $z=E_n$. Then, in terms of δ, one has

$$z=\frac{\lambda+m-\delta}{[(\lambda+m-\delta)^2+\gamma^2]^{1/2}}=E_n-\frac{\gamma^2}{N^3}\delta+O(\delta^2), \tag{3.4}$$

$$c=\frac{\gamma}{[(\lambda+m-\delta)^2+\gamma^2]^{1/2}}=c_0+c_1\delta+O(\delta^2), \tag{3.5}$$

where

$$N=[(\lambda+m)^2+\gamma^2]^{1/2},\quad c_0=\gamma/N,\quad c_1=\gamma(\lambda+m)/N^3. \tag{3.6}$$

The expansion of the homogenous solutions (2.9) is considered first. From (2.16) and the identity $\Gamma(w)\Gamma(1-w)=\pi/\sin\pi w$, it is straightforward to derive for k a nonnegative integer

$$\begin{aligned}M_1(k,b,w)&\equiv\left.\frac{d}{d\delta}M(\delta-k,b,w)\right|_{\delta=0}\\&=k!\sum_{j=0}^{k}\frac{(-1)^j w^j}{(k-j)!\,j!\,(b)_j}\\&\quad\times[\psi(1+k-j)-\psi(1+k)]\\&\quad+k!\sum_{j=k+1}^{\infty}\frac{(-1)^k(j-k-1)!\,w^j}{j!\,(b)_j},\end{aligned} \tag{3.7}$$

where b is not a negative integer and $\psi(w)$ is the logarithmic derivative of $\Gamma(w)$.[14] One can write an integral representation for the second term in (3.7)

$$\sum_{j=k+1}^{\infty}\frac{(j-k-1)!\,w^j}{j!\,(b)_j}=\frac{1}{(b)_k}\int_0^1 dt\,\frac{(1-t)^{b+k-1}}{t^{k+1}}\times\left[e^{wt}-\sum_{j=0}^{k}\frac{w^j t^j}{j!}\right],\quad \mathrm{Re}(b)>-k, \tag{3.8}$$

which can be verified by expanding the expression in the square brackets in a power series in t. Using (3.4)–(3.7) along with the identities[14]

$$\begin{aligned}k!\sum_{j=0}^{k}\frac{(-1)^j w^j}{(k-j)!(b)_j}\frac{j}{j!}&=w\frac{d}{dw}M(-k,b,w)\\&=-kM(1-k,b,w)+kM(-k,b,w),\end{aligned} \tag{3.9}$$

$$(k-1)!\sum_{j=0}^{k-1}\frac{(-1)^j w^j}{(k-j-1)!(b)_j}\frac{j}{j!}$$
$$= w\frac{d}{dw}M(1-k,b,w)$$
$$=(b+k-1)M(-k,b,w)-(b+k-w-1)M(1-k,b,w) \tag{3.10}$$

to evaluate the derivative with respect to the argument of $M_{\alpha,\beta}(z)$, one has

$$M_{\nu\pm 1/2,\lambda}(2cx)=e^{-c_0x}(2c_0x)^{\lambda+1/2}\{M(m\pm,1+2\lambda,2c_0x)$$
$$+\delta[A_\pm M(-m,1+2\lambda,2c_0x)$$
$$+B_\pm M(1-m,1+2\lambda,2c_0x)$$
$$+M_1(-m\pm,1+2\lambda,2c_0x)]\}+O(\delta^2), \tag{3.11}$$

where

$$m_\pm=\tfrac12-m\mp\tfrac12,\quad A_+=(c_1/c_0)(\lambda+m+\tfrac12-c_0x),$$
$$A_-=(c_1/c_0)(2\lambda+m),$$
$$B_+=-mc_1/c_0,\quad B_-=(c_1/c_0)(-\lambda-m+\tfrac12+c_0x). \tag{3.11$'$}$$

For $m=0$, $M_1(-1,b,w)$, which appears in (3.11), is defined to be zero. This is just a convenience since it will turn out that in this case the first-order correction in (3.11) does not enter in the final result for the RDGF.

A similar treatment is done for the W function, where[14]

$$W_{\alpha,\beta}(w)=e^{-w/2}w^{\beta+1/2}U(\tfrac12+\beta-\alpha,1+2\beta,w). \tag{3.12}$$

An integral representation is used for the U function[14]

$$U(\delta-k,b,w)$$
$$=\frac{1}{\Gamma(\delta-k)}\int_0^\infty dt\,e^{-wt}t^{\delta-k-1}(1+t)^{b+k-\delta-1},$$
$$\mathrm{Re}(\delta)>k,\quad \mathrm{Re}(w)>0. \tag{3.13}$$

Equation (3.13) must be analytically continued to the neighborhood of $\delta=0$. This can be done by adding and subtracting the first $(k+1)$ terms in the power series expansion of $(1+t)^{b+k-\delta-1}$ in the integrand to get

$$U(\delta-k,b,w)=\sum_{j=0}^{k}\frac{\Gamma(\delta-k+j)\Gamma(b+k-\delta)}{\Gamma(\delta-k)\Gamma(b+k-j-\delta)}\frac{w^{k-j-\delta}}{j!}$$
$$+\frac{1}{\Gamma(\delta-k)}\int_0^\infty e^{-wt}t^{\delta-k-1}\Big[(1+t)^{b+k-\delta-1}$$
$$-\sum_{j=0}^{k}\frac{\Gamma(b+k-\delta)}{\Gamma(b+k-\delta-j)}\frac{t^j}{j!}\Big]dt,$$
$$\mathrm{Re}(\delta)>-1. \tag{3.14}$$

Again using $\Gamma(w)\Gamma(1-w)=\pi/\sin\pi wz$, one obtains

$$\frac{d}{d\delta}U(\delta-k,b,w)\Big|_{\delta=0}=k!\sum_{j=0}^{k}\frac{(-1)^{k+j}\Gamma(b+k)}{\Gamma(b+j)(k-j)!}\frac{w^j}{j!}$$
$$\times[\psi(j+1)-\psi(k+1)+\psi(j+b)-\psi(b+k)-\ln w]$$
$$+(-1)^k k!\int_0^\infty e^{-wt}t^{-k-1}\Big[(1+t)^{b+k-1}$$
$$-\sum_{j=0}^{k}\frac{\Gamma(b+k)}{\Gamma(b+k-j)}\frac{t^j}{j!}\Big]dt, \tag{3.15}$$

where in the first term the replacement $k-j\to j$ was made. Using (3.4) and (3.5) with the identities (3.9) and (3.10) and

$$\Gamma(\delta-k)=\frac{(-1)^k}{k!}\Big[\frac{1}{\delta}+\psi(k+1)+O(\delta)\Big], \tag{3.16}$$

one has

$$\Gamma(\lambda-\nu+\tfrac12\mp\tfrac12)W_{\nu\pm1/2,\lambda}(2cx)$$
$$=e^{-c_0x}(2c_0x)^{\lambda+1/2}$$
$$\times\Big\{\frac{[m+1\pm(1-m)]}{2m!\Gamma(2\lambda+1)}\frac{\Gamma(2\lambda-m\pm+1)}{\delta}$$
$$\times M(m\pm,1+2\lambda,2c_0x)$$
$$+W_1(-m\pm,1+2\lambda,2c_0x)$$
$$+\frac{[m+1\pm(1-m)]\Gamma(2\lambda-m\pm+1)}{2m!\Gamma(2\lambda+1)}$$
$$\times[A_\pm M(-m,1+2\lambda,2c_0x)$$
$$+B_\pm M(1-m,1+2\lambda,2c_0x)]\Big\}+O(\delta), \tag{3.17}$$

where

$$W_1(k,b,w)=\frac{\Gamma(b+k)}{\Gamma(b)}\sum_{j=0}^{k}\frac{(-1)^j w^j}{(b)_j(k-j)!j!}$$
$$\times[\psi(j+b)-\psi(k+b)+\psi(j+1)-\ln w]$$
$$+\int_0^\infty e^{-wt}t^{-k-1}\Big[(1+t)^{k+b-1}$$
$$-\sum_{j=0}^{k}\frac{\Gamma(b+k)}{\Gamma(b+k-j)}\frac{t^j}{j!}\Big]dt, \tag{3.18}$$

which is the zero order term in the expansion of $\Gamma(\delta-k)U(\delta-k,b,w)$. In (3.18), if the upper limit of the sum is smaller than the lower limit, then the sum is zero.

Using (3.4), (3.5), (3.11), and (3.17), one finds

$$\frac{-1}{\Gamma(1+2\lambda)}\sqrt{\frac{\Gamma(2\lambda+m+1)}{4m!N(N-\kappa)}}\frac{\psi_j^L(x,z)}{\sqrt{1+z}}$$
$$=\psi_j^{m,\kappa}(x)+\frac{\gamma^2}{N^3}F_j(x)\delta+O(\delta^2), \tag{3.19}$$

$$-\sqrt{\frac{m!N(N-\kappa)}{4\Gamma(2\lambda+m+1)}}\frac{\Gamma(\lambda-\nu)}{c^2\sqrt{1+z}}\psi_j^G(x,z)$$
$$=\frac{N^3/\gamma^2}{\delta}\psi_j^{m,\kappa}(x)+G_j(x)$$
$$-\frac{3N(\lambda+m)}{2\gamma^2}\psi_j^{m,\kappa}(x)+O(\delta), \tag{3.20}$$

where

$$F_j(x)=-N'\sqrt{1-s_jE_n}\,x^{\lambda-1}e^{-c_0x}(N^3/\gamma^2)$$
$$\times[A_j(x)M(1-m,1+2\lambda,2c_0x)$$
$$+L_j(x)M(-m,1+2\lambda,2c_0x)$$
$$-mM_1(m-1,1+2\lambda,2c_0x)$$
$$+s_j(\kappa-N)M_1(m,1+2\lambda,2c_0x)], \tag{3.21}$$

$$G_j(x)=[3N(\lambda+m)/2\gamma^2]\psi_j^{m,\kappa}(x)$$
$$-N'\sqrt{1-s_jE_n}\,x^{\lambda-1}e^{-c_0x}(N^3/\gamma^2)$$
$$\times\{P_j(x)M(1-m,1+2\lambda,2c_0x)$$

$$+ R_j(x)M(-m,1+2\lambda,2c_0x)$$
$$+ [m!\Gamma(1+2\lambda)/\Gamma(2\lambda+m+1)]$$
$$\times[-(2\lambda+m)W_1(m-1,1+2\lambda,2c_0x)$$
$$+ s_j(\kappa-N)W_1(m,1+2\lambda,2c_0x)]\}, \qquad (3.22)$$

where N' is the eigenfunction normalization given in Appendix A and

$$s_j = (-1)^j,$$
$$A_j(x) = 1 + m(c_1/c_0)[\lambda + m - \tfrac{1}{2} - c_0x - s_j(\kappa - N)] - s_jm(1+s_jE_n)/2N, \qquad (3.23)$$
$$L_j(x) = s_j[E_n + (\kappa - N)(c_1/c_0)(\lambda + m + \tfrac{1}{2} - c_0x)] - m(c_1/c_0)(2\lambda+m) + (\kappa-N)(1+s_jE_n)/2N,$$
$$P_j(x) = -1 + m(c_1/c_0)[\lambda + m - \tfrac{1}{2} - c_0x - s_j(\kappa-N)] + mE_n/(\kappa+N) - s_jm(1-3s_jE_n)/2N,$$
$$R_j(x) = s_j(\kappa-N)(c_1/c_0)(\lambda+m+\tfrac{1}{2}-c_0x) - m(c_1/c_0)(2\lambda+m) + (\kappa-N)(1-3s_jE_n)/2N.$$

From (3.19) and (3.20), it is clear that $F_j(x)$ and $G_j(x)$ are related to the derivative with respect to energy of the regular and irregular solutions of (2.6) evaluated at the eigenvalue. The additional term in (3.20), proportional to the eigenfunction, can be removed by changing the z dependence of the normalization coefficient of the solutions.

Finally, from (2.7), (2.11), (3.1), (3.19), (3.20), and

$$\frac{1}{E_n - z} = \frac{N^3/\gamma^2}{\delta} - \frac{3N(\lambda+m)}{2\gamma^2} + O(\delta), \qquad (3.24)$$

the radial RDGF for $x_1 > x_2$ is

$$G_\kappa^{ij^R}(x_2,x_1,E_n) = \psi_j^{m,\kappa}(x_1)F_i(x_2) + \psi_i^{m,\kappa}(x_2)G_j(x_1). \qquad (3.25)$$

The radial RDGF for $x_1 < x_2$ can be obtained from the symmetry relation

$$G_\kappa^{ij}(x_2,x_1,z) = G_\kappa^{ji}(x_1,x_2,z), \qquad (3.26)$$

which is evident from (2.7).

4. NONRELATIVISTIC REDUCTION

As a check of (3.25), the nonrelativistic limit is taken here. This is accomplished by dropping all terms of order γ compared to unity (γx is not considered small compared to unity).[11] The expansions are

$$\begin{aligned}
\lambda &= |\kappa| + O(\gamma^2),\\
N &= n + O(\gamma^2),\\
c_0 &= \gamma/n + O(\gamma^3),\\
c_1 &= \gamma/n^2 + O(\gamma^3),\\
(\lambda+m) &= n + O(\gamma^2),\\
E_n &= 1 - \gamma^2/2n^2 + O(\gamma^4),
\end{aligned} \qquad (4.1)$$

and

$$\begin{aligned}
\kappa &= l, \quad \kappa > 0,\\
\kappa &= -(l+1), \quad \kappa < 0.
\end{aligned} \qquad (4.2)$$

In this limit, G^{12^R} and G^{21^R} are of order γ smaller than G^{11^R} while G^{22^R} is of order γ^2 smaller than G^{11^R}. So G^{11^R} is the dominant component and produces the radial reduced Schrödinger Green function in the nonrelativistic limit.

A few of the less direct steps in taking this limit are outlined below. Since λ approaches a nonnegative integer, the integral in (3.18) can be carried out in the nonrelativistic limit. For $k = m-1$ and $b = 1 + 2|\kappa|$, one has

$$\int_0^\infty e^{-wt}t^{-m}\Big[(1+t)^{m+2|\kappa|-1} - \sum_{j=0}^{m-1}\frac{(m+2|\kappa|-1)!}{(m+2|\kappa|-1-j)!}\frac{t^j}{j!}\Big]dt$$
$$= \int_0^\infty e^{-wt}\sum_{j=0}^{2|\kappa|-1}\frac{(n+|\kappa|-1)!}{(j+n-|\kappa|)!}\frac{t^j}{(2|\kappa|-1-j)!}dt$$
$$= \sum_{j=0}^{2|\kappa|-1}\frac{(n+|\kappa|-1)!j!w^{-(j+1)}}{(j+n-|\kappa|)!(2|\kappa|-1-j)!}. \qquad (4.3)$$

A similar result follows for $k = m$. In (3.21), the two infinite series arising from the definition of $M_1(k,b,z)$ in (3.7) can be combined with the aid of the identity

$$-m!\sum_{j=m}^{\infty}\frac{(-1)^{m-1}(j-m)!(2(\gamma/n)x)^j}{j!(1+2|\kappa|)_j}$$
$$+ m!(n-\kappa)\sum_{j=m+1}^{\infty}\frac{(-1)^m(j-m-1)!(2(\gamma/n)x)^j}{j!(1+2|\kappa|)_j}$$
$$= \Big(-2\frac{\gamma}{n}x\Big)^m(2|\kappa|)!m!\sum_{j=0}^{\infty}\Big(2\frac{\gamma}{n}x\Big)^j \times\frac{j!+(n-\kappa)(j-1)!(1-\delta_{0j})}{(n+|\kappa|+j)!(n-|\kappa|+j)!}. \qquad (4.4)$$

Differences of ψ functions can be combined using the recurrence relation[14]

$$\psi(w) = \psi(w-1) + 1/(w-1) \qquad (4.5)$$

to produce finite sums.

The cases $\kappa < 0$ and $\kappa > 0$ reduce to the same result for a given l, but terms combine differently. After a lengthly algebraic calculation, F_1 and G_1 in the nonrelativistic limit become

$$F_1^{\rm NR}(x) = -\frac{n^3}{\gamma^2}y^l\Big(\frac{2\gamma}{n}\Big)^{3/2}e^{-y/2} \times\Big(\frac{(n+l)!}{2n(n-l-1)!}\Big)^{1/2}s_k\Big\{\sum_{j=1}^{n-l-1}B_jy^j$$
$$+ \frac{(-y)^{n-l}}{(n+l+1)!(n-l)}\,{}_2F_2(1,1,n+l+2,n-l+1,y)$$
$$+ \frac{y}{n}\frac{n-l-1}{(2l+2)!}M(l-n+2,2l+3,y)$$
$$+ \frac{1}{(2l+1)!}\Big(\frac{y}{2n} - \frac{2l+3}{2n}\Big)M(l-n+1,2l+2,y)$$
$$+ \frac{\theta(\kappa)}{(2l+1)!(n-l)}M(l-n+1,2l+2,y)\Big\} \qquad (4.6)$$

and

$$G_1^{\rm NR}(x) = -\frac{n^3}{\gamma^2} y^l \left(\frac{2\gamma}{n}\right)^{3/2} e^{-y/2} \left(\frac{(n+l)!}{2n(n-l-1)!}\right)^{1/2} s_\kappa$$
$$\times \left\{ -\sum_{j=1}^{2l+1} \frac{(n-l-1)!(j-1)! y^{-j}}{(n-l-1+j)!(2l+1-j)!} \right.$$
$$+ \sum_{j=0}^{n-l-2} C_j y^j - \frac{1}{(2l+1)!}\left[\psi(n-l) - \ln y - \frac{1}{2n}\right.$$
$$\left. + \frac{2l+3}{2n} - \frac{y}{2n}\right] M(l-n+1, 2l+2, y)$$
$$+ \frac{(n-l-1)}{(2l+2)!}\frac{y}{n} M(l-n+2, 2l+3, y)$$
$$\left. - \frac{\theta(\kappa)}{(2l+1)!(n-l)} M(l-n+1, 2l+2, y)\right\}, \qquad (4.7)$$

where[15]

$$s_\kappa = \frac{|\kappa|}{\kappa}, \quad y = \frac{2\gamma}{n}x,$$
$${}_2F_2(1,1,a,b,w) = \sum_{j=0}^{\infty} \frac{w^j j!}{(a)_j (b)_j},$$
$$B_j = \frac{(-1)^j}{j!} \frac{(n-l-1)!}{(2l+j+1)!(n-l-j-1)!} \times \sum_{k=0}^{j-1} \frac{1}{n-l-k-1}, \qquad (4.8)$$
$$C_j = \frac{(-1)^j}{j!} \frac{(n-l-1)!}{(2l+j+1)!(n-l-j-1)!} \times \sum_{k=j+1}^{n-l-1} \frac{2k+2l+1}{k(2l+k+1)}.$$

The last terms in (4.6) and (4.7) are proportional to the nonrelativistic wavefunction; and although they are κ-dependent, they will cancel upon substitution in the nonrelativistic analog of (3.25). Therefore, a κ-independent radial reduced Schrödinger Green function will result. Equations (4.6)–(4.8) together with the nonrelativistic limit of the eigenfunctions

$$\psi^{\rm NR}(x) = s_\kappa \lim_{\gamma\to 0} \psi_1^{m,\kappa}(x), \qquad (4.9)$$

where $\psi^{\rm NR}(x)$ is the normalized nonrelativistic radial eigenfunction,[11] determine the nonrelativistic limit of (3.25). With the aid of the identity[14]

$$M(-\beta, \alpha+1, x) = [\beta!\alpha!/(\beta+\alpha)!] L^\alpha{}_\beta(x), \qquad (4.10)$$

this limit can be shown to agree with the result of the calculations of Hameka[9] and Sherstyuk.[8]

5. CONCLUSION

The radial RDGF for a Coulomb potential for a general state is given in (3.25). For any state with $m = 0$ ($1S_{1/2}, 2P_{3/2}, 3D_{5/2}, \cdots$), one has

$$F_i(x) = \psi_i^{0,\kappa}(x) \frac{n^3}{\gamma^2}\left[\frac{\lambda}{n^2}\left(\lambda + \frac{1}{2} - \frac{\gamma}{n}x\right) + \frac{s_i}{2n} + I_i'\right], \qquad (5.1)$$

$$G_i(x) = \psi_i^{0,\kappa}(x) \frac{n^3}{\gamma^2}\left[\frac{\lambda}{n^2}\left(\lambda + \frac{1}{2} - \frac{\gamma}{n}x\right)\right.$$
$$\left. + \frac{s_i}{2n} + \psi(1) - \ln\left(\frac{2\gamma}{n}x\right) + I_i\right], \qquad (5.2)$$

where

$$I_i = e^{2(\gamma/n)x} \int_1^\infty dt\, e^{-(2\gamma/n)xt} \left[s_i \frac{\lambda}{n} t^{2\lambda-1} + \frac{t^{2\lambda}-1}{t-1}\right] \qquad (5.3)$$

and

$$I_i' = \int_0^1 dt \left[t^{2\lambda} \frac{e^{2(\gamma/n)x(1-t)} - 1}{1-t} - s_i \frac{\lambda}{n} t^{2\lambda-1} e^{2(\gamma/n)x(1-t)}\right]. \qquad (5.4)$$

The integral in (5.4) may be converted to an infinite sum via (3.8). For any state with $m = 1$ ($2S_{1/2}, 2P_{1/2}, 3P_{3/2}, 3D_{3/2}, \ldots$), one has

$$F_i(x) = g_i(x)\left\{\frac{\kappa - N_1}{2N_1} + 1 - \frac{s_i}{2N_1} + \frac{E_n}{N_1}\right.$$
$$\times \left[s_i(\kappa - N_1)\left(\lambda + 1 - \frac{1}{2}y_1\right) + s_i N_1 - \lambda - 1 - \frac{1}{2}y_1\right]$$
$$- \frac{y_1}{2\lambda+1}\left[\frac{\kappa - N_1}{2N_1} + \frac{E_n}{N_1} \times \left(s_i(\kappa - N_1)\left(\lambda + 2 - \frac{1}{2}y_1\right)\right.\right.$$
$$\left.\left.\left. + s_i N_1 - 2\lambda - 1\right) - s_i(\kappa - N_1)\right] + J_i(x)\right\}, \qquad (5.5)$$

$$G_i(x) = \frac{3N_1(\lambda+1)}{2\gamma^2}\psi_i^{1,\kappa}(x)$$
$$+ g_i(x)\left\{\frac{\kappa - N_1}{2N_1} - 1 - \frac{s_i}{2N_1}\right.$$
$$+ [1 - s_i(\kappa - N_1)][\ln y_1 - \psi(1)] \qquad (5.6)$$
$$+ \frac{E_n}{N_1}\left[s_i(\kappa - N_1)\left(\lambda - 1 - \frac{1}{2}y_1\right) + \frac{N_1}{\kappa + N_1} - \lambda + 1 - \frac{1}{2}y_1\right]$$
$$- \frac{y_1}{1+2\lambda}\frac{E_n}{N_1} \times \left[s_i(\kappa - N_1)\left(\lambda - \frac{1}{2}y_1\right) - 2\lambda - 1 + \frac{\kappa - N_1}{2E_n}\right]$$
$$\left. + \frac{s_i}{\kappa + N_1}[1 + y_1\psi(2) - y_1 \ln y_1] + J_i'(x)\right\},$$

where

$$N_1 = [(\lambda+1)^2 + \gamma^2]^{1/2}, \quad y_1 = (2\gamma/N_1)x,$$
$$g_i(x) = -\frac{N_1^3}{\gamma^2}\left(\frac{\Gamma(2\lambda+2)(1 - s_i E_n)}{4N_1(N_1 - \kappa)}\right)^{1/2} \times \frac{(2\gamma/N_1)^{\lambda+1/2} x^{\lambda-1} e^{-y_1/2}}{\Gamma(1+2\lambda)}, \qquad 5.7)$$
$$J_i(x) = \int_0^1 dt \frac{(1-t)^{2\lambda}}{t} \times \left[1 - e^{y_1 t} + \frac{s_i(\kappa - N_1)}{1+2\lambda}\frac{(1-t)}{t}(1 + y_1 t - e^{y_1 t})\right], \qquad (5.8)$$

$$J_i'(x) = \int_0^\infty dt\, e^{-y_i t}\left[\frac{1-(1+t)^{2\lambda}}{t} - \frac{s_i}{(\kappa+N_1)}\frac{(1+t)^{2\lambda+1}-1-(1+2\lambda)t}{t^2}\right]. \tag{5.9}$$

The integral in (5.8) may be written as an infinite sum via (3.8).

Also obtained are $F_i(x)$ and $G_i(x)$ in (3.19)–(3.22), which are the derivative with respect to energy of the regular and irregular solutions of the homogeneous equation (2.6), evaluated at the eigenvalue. The nonrelativistic analogs are given in (4.6)–(4.8).

ACKNOWLEDGMENTS

The author wishes to thank Dr. Peter J. Mohr for suggesting the problem and for helpful guidance and discussions. This research was supported by the National Science Foundation, Grant No. PHY80-26549.

APPENDIX A: COULOMB EIGENFUNCTIONS

The Coulomb eigenfunctions are solutions of the equation

$$[H(\mathbf{x}) - E_n I]\Psi(\mathbf{x}) = 0, \tag{A1}$$

where $H(\mathbf{x})$ is given in (2.2) with $V(x) = -\gamma/x$. The normalized solutions are[10]

$$\Psi(\mathbf{x}) = \begin{pmatrix} f_1(x)\chi_\kappa^\mu(\hat{x}) \\ i f_2(x)\chi_{-\kappa}^\mu(\hat{x}) \end{pmatrix}, \tag{A2}$$

where

$$\chi_\kappa^\mu(\hat{x}) = \begin{pmatrix} -\dfrac{\kappa}{|\kappa|}\left(\dfrac{\kappa+\frac{1}{2}-\mu}{2\kappa+1}\right)^{1/2} Y^{\mu-1/2}_{|\kappa+1/2|-1/2}(\hat{x}) \\ \left(\dfrac{\kappa+\frac{1}{2}+\mu}{2\kappa+1}\right)^{1/2} Y^{\mu+1/2}_{|\kappa+1/2|-1/2}(\hat{x}) \end{pmatrix}. \tag{A3}$$

The normalized radial bound-state wavefunctions are[10,11]

$$\begin{aligned} f_i(x) \equiv \psi_i^{m,\kappa}(x) &= N'\sqrt{1-s_iE_n}\, x^{\lambda-1}e^{-(\gamma/N)x} \\ &\times [mM(-m+1, 1+2\lambda, 2(\gamma/N)x) \\ &- s_i(\kappa-N)M(-m, 1+2\lambda, 2(\gamma/N)x)], \end{aligned} \tag{A4}$$

where

$$N' = \frac{(2/\gamma N)^{\lambda+1/2}}{\Gamma(2\lambda+1)}\left(\frac{\Gamma(2\lambda+m+1)}{4m!N(N-\kappa)}\right)^{1/2}, \tag{A5}$$

and

$$E_n = \frac{\lambda+m}{N}, \tag{A6}$$

where m, given in (3.3), is any nonnegative integer. The continuum radial wavefunctions ($E>1$ or $E<-1$) are[10]

$$\begin{aligned} f_j(x) \equiv \psi_j^\kappa(x,E) &= N_2\sqrt{1-s_jE}\, e^{-ipx}x^{\lambda-1} \\ &\times\left[M\left(\lambda+1+i\frac{\gamma E}{p}, 1+2\lambda, 2ipx\right)\right. \\ &\left. + s_j\frac{\kappa+i\gamma/p}{\lambda+i\gamma E/p}M\left(\lambda+i\frac{\gamma E}{p}, 1+2\lambda, 2ipx\right)\right], \end{aligned} \tag{A7}$$

where

$$\begin{aligned} N_2 &= \frac{|\Gamma(\lambda+1-i\gamma E/p)|e^{\pi\gamma E/2p}(2p)^\lambda}{2\Gamma(1+2\lambda)\sqrt{\pi p}}, \\ p &= (E^2-1)^{1/2}, \\ &\arg(\sqrt{1-s_jE}) = 0 \quad\text{or}\quad \pi/2. \end{aligned} \tag{A8}$$

Some important properties of the wavefunctions are

$$\begin{aligned} K\Psi(\mathbf{x}) &= -\kappa\Psi(\mathbf{x}), \\ J_3\Psi(\mathbf{x}) &= \mu\Psi(\mathbf{x}), \\ \boldsymbol{\sigma}\cdot\hat{x}\chi_\kappa^\mu(\hat{x}) &= -\chi_{-\kappa}^\mu(\hat{x}). \end{aligned} \tag{A9}$$

APPENDIX B: NUMERICAL CALCULATION

As a check of the results in (3.25), the radial RDGF was used to numerically calculate the second-order perturbation energy for a perturbing potential

$$\delta V = -\epsilon\alpha/r. \tag{B1}$$

The exact result can be obtained by replacing Z by $(Z+\epsilon)$ in the exact expression for the bound state energy in (A6). This is expanded to second order in ϵ to get

$$E_n' = E_n + E_n^{(1)}\epsilon + E_n^{(2)}\epsilon^2 + O(\epsilon^3), \tag{B2}$$

where

$$E_n^{(1)} = -(Z\alpha^2/\lambda N^3)(\kappa^2+\lambda m) \tag{B3}$$

and

$$\begin{aligned} E_n^{(2)} = &-\tfrac{1}{2}[(\alpha^2/\lambda N^3)(\kappa^2+\lambda m) + (Z\alpha^2)^2\kappa^2/\lambda^3N^3 \\ &+ (Z\alpha^2)^2(3m/\lambda^2N^5)(\kappa^2+\lambda m)]. \end{aligned} \tag{B4}$$

From perturbation theory, one has

$$\epsilon^2E_n^{(2)} = -\int_0^\infty dx_1\int_0^\infty dx_2\, x_1^2x_2^2\,\delta V(x_1)\delta V(x_2) \times\sum_{ij}\psi_i^{m,\kappa}(x_2)G_\kappa^{ij^R}(x_2,x_1,E_n)\psi_j^{m,\kappa}(x_1). \tag{B5}$$

To numerically calculate $G^{ij^R}(x_2,x_1,E_n)$, everything is straightforward with the possible exception of the evaluation of $M_1(k,b,z)$ and $W_1(k,b,z)$. The range of variables $10^{-6} \leqslant 2c_0x_i \leqslant 100$, $i=1,2$, is adequate to calculate (B5) to six figures. In this range, the infinite series in (3.7) can be evaluated to give the result

$$\sum_{j=k+1}^\infty \frac{(j-k-1)!w^j}{(b)_j j!} = \Gamma(b)w^{k+1}\left[\sum_{j=0}^M \frac{w^j j!}{(j+k+1)!\Gamma(j+k+b+1)} + R_M\right], \tag{B6}$$

where

$$R_M < T_{M+1}(1-w/M)^{-1}, \quad M>w, \tag{B6'}$$

and T_{M+1} is the $(M+1)$th term of the series on the right-hand side of (B6). The integral occuring in the expansion for $W_1(k,b,z)$ in (3.23) was evaluated using a 12-point Laguerre numerical integration method.[14] However, the integrand must be calculated in double precision because of cancellations. In the case $m=0$, the evaluation of $M(1,b,w)$ was required. This was done using a 10-point Legendre numerical integration method to evaluate the integral representation[14]

$$M(1,b,w) = (b-1)\int_0^1 e^{wt}(1-t)^{b-2}\,dt, \quad b>1. \tag{B7}$$

The integral in (B5) was numerically calculated to six figures by applying Gauss type numerical integration techniques. The calculated value for $E_n^{(2)}$ agrees with the exact value from (B4).

[1]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953).

[2]B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956).

[3]E. H. Wichmann and N. M. Kroll, Phys. Rev. **101**, 843 (1956).

[4]G. E. Brown and G. W. Schaefer, Proc. R. Soc. (London), Ser. A **233**, 527 (1956).

[5]A. Maquet, Phys. Rev. A **15**, 1088 (1977).

[6]N. L. Manakov and S. A. Zapryagaev, Phys. Lett. A **58** , 23 (1976).

[7]B. A. Zon, N. L. Manakov, and L. P. Rapoport, Sov. Nucl. Phys. **15**, 508 (1972).

[8]A. L. Sherstyuk, Opt. Spectrosc. **30**, 191 (1971).

[9]H. F. Hameka, J. Chem. Phys. **47**, 2728 (1967); erratum in J. Chem. Phys. **48**, 4810 (1968).

[10]M. E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).

[11]H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Springer-Verlag, New York, 1957); the normalized radial Dirac eigenfunctions, as given in (14.37) on p. 69, were used with a positive instead of a negative sign on the lower component.

[12]P. J. Mohr, Ann. Phys. **88**, 26 (1974).

[13]M. Gyulassy, Nucl. Phys. A **244**, 497 (1975).

[14]M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, NBS Applied Mathematical Series (Natl. Bur. Stds., Washington, DC, 1966).

[15]B. R. Johnson and J. O. Hirschfelder, J. Math. Phys. **20**, 2484 (1979).

Simple calculation of Löwdin's alpha-function

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

(Received 27 July 1983; accepted for publication 30 September 1983)

A function of the type $f(R)Y_L^M(\Theta,\Phi)$ around a specific center B is expanded about another center A as $f(R)Y_L^M(\Theta,\Phi) = \Sigma_{l=|M|}^{\infty}(1/r)\alpha_l(fLM|a,r)Y_l^M(\vartheta,\varphi)$, where $Y_L^M(\Theta,\Phi)$ and $Y_l^M(\vartheta,\varphi)$ are the complex spherical harmonics around the respective centers, which have been assumed to be quantized in the direction of A to B. The distances from the respective centers are R and r, and the separation between the two centers is a. The expansion coefficient $\alpha_l(fLM|a,r)$, which is usually called Löwdin's α-function, can be expressed in a remarkably simple form by using Silverstone and Moats' formula. Owing to the simplicity, the properties of the coefficients $b_{Kk}(LM|l)$ appearing in the α-function are manifested, and the recurrence formulas for evaluating $b_{Kk}(LM|l)$ are derived. Furthermore, when $r\approx 0$, the asymptotic expression for the α-function is given in a compact form.

PACS numbers: 31.15. + q, 71.10. + x, 02.30.Mv, 02.30.Bi

1. INTRODUCTION

For calculation on electronic structures of molecules and solids, a function of the type $f(R)Y_L^M(\Theta,\Phi)$ is frequently used as an atomic orbital. Here $f(R)$ is a specific radial function, and $Y_L^M(\Theta,\Phi)$ a complex spherical harmonic around a center B. Then any tractable expansion formula of the function in terms of spherical harmonics $Y_l^m(\vartheta,\varphi)$ around a displaced center A, if available, will be very useful to compute molecular integrals between atomic orbitals about separate centers. A number of such formulas have been presented by several investigators,[1–11] though they cannot be said to be easy to deal with. However, only one of them seems relatively compact and will be able to lead to some simple expression for the expansion when both $Y_L^M(\Theta,\Phi)$ and $Y_l^m(\vartheta,\varphi)$ are quantized in the direction of center A to center B. This is the one derived by Silverstone and Moats.[9]

In the above special case the expansion coefficient $\alpha_l(fLM|a,r)$ is usually called Löwdin's α-function[2] [see Eq. (2.1)]. Here a is the separation between the two centers A and B, and r the distance from A. The expressions for the α-function have been given in the conventional manner by expanding a product of two associated Legendre functions with power series of r, R, and a, but they are in considerably sophisticated forms.[1–4,6,8,10] On the other hand, it was suggested by Silvertone and Moats[9] that a simpler expression for the α-function is derivable from their expansion formula obtained by means of the Fourier-transform convolution technique.[5] No explicit expression, however, has been so far given.

The aim of this paper is to actually demonstrate that an expression for the α-function which makes its calculation remarkably easy can be obtained from Silverstone and Moats' formula. In this context, first, in next section the way to lead to the expression will be briefly described, and the coefficients $b_{Kk}(LM|l)$ [see Eq. (2.6)] appearing in the expression will be defined in a suitable form for their numerical calculation. Second, the property of $b_{Kk}(LM|l)$ will be manifested in Sec. 3, and then the recurrence formulas for evaluating them will be derived in Sec. 4. Third, in Sec. 5, when $r\approx 0$, the asymptotic expression for the α-function will be given in a compact form. Fourth, in Sec. 6 some general notes will be stated on computing molecular integrals by the use of the present formula.

2. DERIVATION OF THE α-FUNCTION WITH A TRACTABLE STRUCTURE

At the outset the α-function $\alpha_l(fLM|a,r)$ is defined by

$$f(R)Y_L^M(\Theta,\Phi) = \sum_{l=|M|}^{\infty}\left(\frac{1}{r}\right)\alpha_l(fLM|a,r)Y_l^M(\vartheta,\varphi), \tag{2.1}$$

where $f(R)$ is the radial part of a specific function to be expanded, and $Y_L^M(\Theta,\Phi)$ and $Y_l^M(\vartheta,\varphi)$ are the complex spherical harmonics in the respective polar coordinate systems shown in Fig. 1.

When the center of the function to be expanded is placed on the positive z-axis in the Cartesian coordinate system, Silverstone and Moats' formula[9] is reduced to a formula expressing $\alpha_l(fLM|a,r)$. It may be written as

$$\begin{aligned}\alpha_l(fLM|a,r) = {} & \frac{(-1)^M}{4\pi}[(2L+1)(2l+1)]^{1/2} \\ & \times 2\sum_{\lambda=|L-l|}^{L+l} C(Ll\lambda;M-M) \\ & \times C(Ll\lambda;00)v_{l\lambda L}(r,a),\end{aligned} \tag{2.2}$$

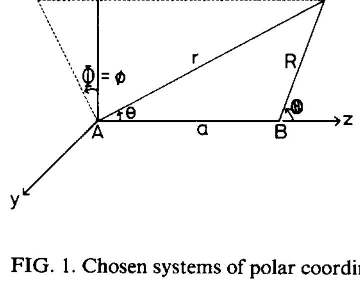


FIG. 1. Chosen systems of polar coordinates.

where $C(Ll\lambda;M\,{-M})$ are the Clebsch–Gordan coefficients,[12] and

$$v_{l\lambda L}(r,a) = 2\pi(-1)^l \sum_{K=0}^{(L+l+\lambda)/2} \sum_{k=0}^{(L+l+\lambda)/2-K} D_{l\lambda LKk} \times \left(\frac{r}{a}\right)^{2k-l} \int_{|r-a|}^{r+a} dR \left(\frac{R}{a}\right)^{2K-L+1} f(R), \tag{2.3}$$

with

$$D_{l\lambda LKk} = \left[(2K)!!(2K-2L-1)\overline{!!}(2k)!!(2k-2l-1)\overline{!!} \times (L+l+\lambda-2K-2k)!! \times (L+l-\lambda-2K-2k-1)\overline{!!}\right]^{-1}. \tag{2.4}$$

The symbol !! indicates a usual double factorial with a nonnegative argument, while $\overline{!!}$ indicates one with a non-negative or negative argument and, to be definite,

$$(2p-1)\overline{!!} = \begin{cases} (2p-1)!! & \text{if } p \text{ is a non-negative integer.} \\ 1/[(-1)^{-p}(-2p-1)!!] & \text{otherwise.} \end{cases} \tag{2.5}$$

Note that $(-1)!! \equiv 1$.

The coefficients irrespective of r and a in Eq. (2.2) can be expressed in a form depending on only two indices K and k by changing the order of the summations. Then we denote each of the coefficients by the product of two symbols $\gamma(LM\,|l)$ and $b_{Kk}(LM\,|l)$. Here $b_{Kk}(LM\,|l)$ has been defined as

$$b_{Kk}(LM\,|l) \equiv \beta(LM\,|l) \times 2 \sum_{\lambda=\lambda_{\min}}^{L+l} C(Ll\lambda;M\,{-M})C(Ll\lambda;00)D_{l\lambda LKk}, \tag{2.6}$$

where

$$\beta(LM\,|l) = (-1)^{L+l-M} \frac{2^{L+l}}{(2L-1)!!(2l-1)!!} \times [(L-M)!(L+M)!(l-M)!(l+M)!]^{1/2} \tag{2.7}$$

and

$$\lambda_{\min} = \max\{|L-l|, 2(K+k)-(L+l)\}. \tag{2.8}$$

Then $\gamma(LM\,|l)$ may be written as

$$\gamma(LM\,|l) = (-1)^L \frac{(2L-1)!!(2l-1)!!}{2^{L+l+1}} \times \left[\frac{(2L+1)(2l+1)}{(L-M)!(L+M)!(l-M)!(l+M)!}\right]^{1/2}. \tag{2.9}$$

Since $\alpha_l(fLM\,|a,r)$ has been defined in the same manner as $\alpha_l(NLM\,|a,r)$ by Sharma,[8] the coefficient $\gamma(LM\,|l)b_{Kk}(LM\,|l)$ is identical with Sharma's $b_\nu(slLM)$[8] (in Ref. 8, s and ν are used in place of our K and k, respectively). Here it should be noted that the coefficient is expressed in a remarkably simpler form than Sharma's $b_\nu(slLM)$. It may be attributed to the use of Silverstone and Moats' formula.

In the preceding paragraph the definition of $b_{Kk}(LM\,|l)$ was made so that $\gamma(LM\,|l)$ may take the same formula as Sharma's $b_0(0lLM)$. Thereby it is assured that necessarily

$$b_{00}(LM\,|l) = 1. \tag{2.10}$$

Owing to this definition of $b_{Kk}(LM\,|l)$, the properties of $b_{Kk}(LM\,|l)$ can be manifested readily, and furthermore some easy recurrence formulas to evaluate $b_{Kk}(LM\,|l)$ can be derived without difficulty.

The coefficient $b_{Kk}(LM\,|l)$ is equivalent to $c_\nu(NLM\,|ls)$ introduced by Löwdin[2] (also in Ref. 2, s and ν take the place of K and k, respectively). These two are connected by the relation

$$c_k(NLM\,|lK) = (-1)^L \frac{(2L-1)!!(2l-1)!!}{(L-M)!(l+M)!} b_{Kk}(LM\,|l). \tag{2.11}$$

Eventually, by making use of $b_{Kk}(LM\,|l)$ and $\gamma(LM\,|l)$, we can rewrite Eq. (2.2) in a compact form as

$$\alpha_l(fLM\,|a,r) = \gamma(LM\,|l) \sum_{K=0}^{L+l} \sum_{k=0}^{L+l-K} b_{Kk}(LM\,|l) \times \left(\frac{r}{a}\right)^{2k-l} \int_{|r-a|}^{r+a} dR \left(\frac{R}{a}\right)^{2K-L+1} f(R). \tag{2.12}$$

3. PROPERTIES OF $b_{Kk}(LM|l)$

In this section the properties of $b_{Kk}(LM\,|l)$ will be manifested, since they may help us to derive some recurrence formulas for evaluating $b_{Kk}(LM\,|l)$: first its symmetry property is derived; second it is shown that for a few specific values of K and k, $b_{Kk}(LM\,|l)$ take particular values; and third it is proved that a few special sums of $b_{Kk}(LM\,|l)$ vanish on the respective conditions.

Since $b_{Kk}(LM\,|l)$ has been defined by Eq. (2.6), its symmetry property can be found from consideration of the symmetry properties of $\beta(LM\,|l)$, $C(Ll\lambda;M\,{-M})$, and $D_{l\lambda LKk}$. Those are given by

$$\beta(L\,{-M}\,|l) = \beta(lM\,|L) = \beta(LM\,|l), \tag{3.1}$$

$$C(Ll\lambda;\,{-MM}) = C(Ll\lambda;M\,{-M}) \quad \text{for } L+l-\lambda \text{ even,}^{13} \tag{3.2}$$

$$C(lL\lambda;M\,{-M}) = C(Ll\lambda;M\,{-M}),^{13} \tag{3.3}$$

and

$$D_{L\lambda lkK} = D_{l\lambda LKk}. \tag{3.4}$$

Introduction of Eqs. (3.1)–(3.4) into Eq. (2.6) leads immediately to the symmetry property of $b_{Kk}(LM\,|l)$. It may be expressed as

$$b_{Kk}(L\,{-M}\,|l) = b_{Kk}(LM\,|l) \tag{3.5}$$

and

$$b_{kK}(lM\,|L) = b_{Kk}(LM\,|l). \tag{3.6}$$

Therefore, henceforth only $b_{Kk}(LM\,|l)$ with $M \geqslant 0$ and $l \geqslant L$ will be taken into account.

The coefficient $b_{Kk}(LM\,|l)$ may be regarded as the element on the $(K+1)$th row and the $(k+1)$th column in a square matrix $\mathbb{b}(LM\,|l)$ of dimension $L+l+1$. Since the elements in which $K+k > L+l$ never appear in the matrix as seen in Eq. (2.12), the symbol $\times$ will be placed on the sites which those elements should occupy.

Next we will consider what values the two elements, $b_{0L+l}(LM|l)$ and $b_{L+l0}(LM|l)$, on the right-top and left-bottom of the matrix $\mathbf{b}(LM|l)$ take. When the explicit expression[14] for $C(LlL+l;M-M)$:

$$C(LlL+l;M-M)=(L+l)!\times\left[\frac{(2L)!(2l)!}{(2L+2l)!(L-M)!(L+M)!(l-M)!(l+M)!}\right]^{1/2}, \tag{3.7}$$

and that for $C(LlL+l;00)$ which can be obtained by setting $M=0$ in the above formula, are introduced into Eq. (2.6), it is easily found that

$$b_{0L+l}(LM|l)=(-1)^{L-M} \tag{3.8}$$

and

$$b_{L+l0}(LM|l)=(-1)^{l-M}. \tag{3.9}$$

Therefore, in general, the matrix $\mathbf{b}(LM|l)$ may be written as

$$\mathbf{b}(LM|l)=\begin{pmatrix} 1 & \cdots & (-1)^{L-M} \\ & & \times \\ \vdots & ⋰ & \vdots \\ (-1)^{l-M} & \times\cdots & \times \end{pmatrix}. \tag{3.10}$$

Finally we will take into consideration a few special sums of $b_{Kk}(LM|l)$. The sum over all the first-row elements is expressed by

$$\sum_{k=0}^{L+l} b_{0k}(LM|l) =(-1)^{L+l}(2L-1)!!\beta(LM|l)\times 2\sum_{\lambda=l-L}^{L+l}\frac{C(Ll\lambda;M-M)C(Ll\lambda;00)}{(L+l+\lambda)!!}\times\sum_{k=0}^{(L+l+\lambda)/2}(-1)^k\binom{\frac{L+l+\lambda}{2}}{k}\times\frac{(2l-2k-1)!!}{(L+l-\lambda-2k-1)!!}. \tag{3.11}$$

This is derived by introducing the formula for $D_{l\lambda L0k}$ [see Eq. (2.4)] into Eq. (2.6) and then changing the summation order. $\binom{n}{k}$ denotes the binomial coefficient $n!/(n-k)!k!$. In Eq. (3.11), if $(L+l+\lambda)/2>(-L+l+\lambda)/2$, i.e., $L>0$, then the inner sum vanishes regardless of λ. This validity is proved generally in Appendix. As a result, it follows that

$$\sum_{k=0}^{L+l} b_{0k}(LM|l)=0 \quad \text{if } L>0. \tag{3.12}$$

Also, from the same consideration as the above, it can be seen that the sum over all the first-column elements vanishes if $l>0$, i.e.,

$$\sum_{K=0}^{L+l} b_{K0}(LM|l)=0 \quad \text{if } l>0. \tag{3.13}$$

When Eq. (2.4) is used for $D_{l\lambda LKk}$ in Eq. (2.6) and then $K+k$ is replaced by p, $b_{Kk}(LM|l)$ may be rewritten as

$$b_{Kp-K}(LM|l) = \frac{(-1)^{l-p+K}}{(2p)!!}\beta(LM|l)\binom{p}{K}\frac{(2K-2p+2l-1)!!}{(2K-2L-1)!!}\times 2\sum_{\lambda=\lambda_{\min}}^{L+l}\frac{C(Ll\lambda;M-M)C(Ll\lambda;00)}{(L+l+\lambda-2p)!!(L+l-\lambda-2p-1)!!}, \tag{3.14}$$

where

$$\lambda_{\min}=\max\{l-L,2p-(L+l)\}. \tag{3.15}$$

Here p indicates that $(p+1)$th, from the left, of $L+l+1$ arrays which consist of the elements aligning obliquely from upper-right to lower-left in the matrix $\mathbf{b}(LM|l)$. In particular, for $p=L+l$, by introducing Eq. (3.7) into Eq. (3.14), one finds that

$$b_{KL+l-K}(LM|l)=(-1)^{L-M}\cdot(-1)^K\binom{L+l}{K}. \tag{3.16}$$

This is nothing but the product of one of the binomial coefficients in $(x-y)^{L+l}$ with $b_{0L+l}(LM|l)$. Thus the products with $K=0$ to $L+l$ appear in turn on the most right oblique array. Needless to say, the sum of $b_{KL+l-K}(LM|l)$ over the possible range of K vanishes so long as $L+l\neq 0$. In general, the sum of $b_{Kp-K}(LM|l)$ over the possible range of K is expressed in a form similar to Eq. (3.11):

$$\sum_{K=0}^{p} b_{Kp-K}(LM|l) = \frac{(-1)^{l-p}}{(2p)!!}\beta(LM|l)\times\sum_{\lambda=\lambda_{\min}}^{L+l}\frac{C(Ll\lambda;M-M)C(Ll\lambda;00)}{(L+l+\lambda-2p)!!(L+l-\lambda-2p-1)!!}\times\sum_{K=0}^{p}(-1)^K\binom{p}{K}\frac{(2K-2p+2l-1)!!}{(2K-2L-1)!!}. \tag{3.17}$$

If $p>L+l-p$, i.e., $p>(L+l)/2$, then the above sum vanishes in the same manner as Eqs. (3.12) and (3.13). That is,

$$\sum_{K=0}^{p} b_{Kp-K}(LM|l)=0 \quad \text{if } p>\frac{L+l}{2}. \tag{3.18}$$

In conclusion, all the properties of $b_{Kk}(LM|l)$ derived in this section may help us to calculate $b_{Kk}(LM|l)$. At least they are useful for us to check for some particular $b_{Kk}(LM|l)$ whether their calculated values are correct or not.

4. RECURRENCE FORMULAS FOR EVALUATING $b_{Kk}(LM|l)$

In this section a few formulas necessary to numerically calculate $b_{Kk}(LM|l)$ repeatedly will be derived. Further some result obtained by the use of them will be presented.

In Eq. (3.14) the factor on the sum depends on both p and K, while the sum only p. Therefore, the direct use of Eq. (3.14) leads to the recurrence formula in K. It may be written as

$$b_{K+1p-K-1}(LM|l) = \frac{(K-p)(2K-2p+2l+1)}{(K+1)(2K-2L+1)} b_{Kp-K}(LM|l). \tag{4.1}$$

It should be noted that the factor on the right-hand side of Eq. (4.1) is independent of M. Accordingly, evaluating the factor once is enough for $b_{Kp-K}(LM|l)$ with the same set of L and l. By the iterative use of Eq.(4.1) all the elements in each oblique array except the one on the first row or column are evaluated without difficulty. Then the problem how to easily find the value of the first-row or -column elements arises. Here we will derive a recurrence formula for evaluating the first-row elements.

Going back to the definition of $b_{Kk}(LM|l)$, Eq. (2.6), and then introducing into it the recurrence formula[15] for $C(Ll\lambda; M -M)$ in M which is expressed by

$$\begin{aligned}
&C(Ll\lambda;M-M)\\
&\quad= \{[\lambda(\lambda+1) - L(L+1) - l(l+1) + 2(M-1)^2]\\
&\qquad\times C(Ll\lambda;M-1-M+1)\\
&\qquad- [(L-M+2)(L+M-1)(l-M+2)(l+M-1)]^{1/2} C(Ll\lambda;M-2-M+2)\}/\\
&\qquad\times[(L+M)(L-M+1)(l+M)(l-M+1)]^{1/2},
\end{aligned} \tag{4.2}$$

one obtains an alternative expression for $b_{0k}(LM|l)$:

$$\begin{aligned}
b_{0k}(LM|l) &= (-1)^L(2L-1)!!\beta(LM|l)/\{(2k)!!(2k-2l-1)\,\overline{!!}\,[(L+M)(L-M+1)(l+M)(l-M+1)]^{1/2}\}\\
&\quad\times 2\sum_{\lambda=\lambda_{\min}}^{L+l}\{[\lambda(\lambda+1) - L(L+1) - l(l+1) + 2(M-1)^2]C(Ll\lambda;M-1-M+1)\\
&\quad- [(L-M+2)(L+M-1)(l-M+2)(l+M-1)]^{1/2} C(Ll\lambda;M-2-M+2)\}C(Ll\lambda;00)/\\
&\quad\times[(L+l+\lambda-2k)!!(L+l-\lambda-2k-1)\,\overline{!!}\,],
\end{aligned} \tag{4.3}$$

where

$$\lambda_{\min} = \max\{l-L, 2k-(L+l)\}. \tag{4.4}$$

Furthermore, by using the equality

$$\begin{aligned}
\lambda(\lambda+1) = &-(L+l+\lambda-2k)(L+l-\lambda-2k-1)\\
&+(L+l-2k)(L+l-2k-1)
\end{aligned} \tag{4.5}$$

and again returning to Eq. (2.6), Eq. (4.3) can be rewritten as

$$\begin{aligned}
&b_{0k}(LM|l)\\
&\quad= [(L-M+1)(l-M+1)]^{-1}\{2(k+1)(2k-2l+1)\\
&\qquad\times b_{0k+1}(LM-1|l) - 2[(L-2k-1)(l-2k-1)\\
&\qquad-(k+1)(2k+1)+(M-1)^2]\\
&\qquad\times b_{0k}(LM-1|l) - (L+M-1)(l+M-1)\\
&\qquad\times b_{0k}(LM-2|l)\}.
\end{aligned} \tag{4.6}$$

When the first term in the braces in the above equation is derived, the observation has been used that $\lambda_{\min} = \max\{l-L, 2k-(L+l)\}$ is substituted by $\lambda_{\min} = \max\{l-L, 2(k+1)-(L+l)\}$ because the term with $\lambda = 2k-(L+l)$ never arises. Replacing $M-1$ by M in Eq. (4.6) and then transforming Eq. (4.6), we can obtain the recurrence formula for $b_{0k}(LM|l)$ in k:

$$\begin{aligned}
&b_{0k+1}(LM|l)\\
&\quad= \frac{1}{2(k+1)(2k-2l+1)}\{(L-M)(l-M)\\
&\qquad\times b_{0k}(LM+1)|l) + 2[(L-2k-1)(l-2k-1)\\
&\qquad-(k+1)(2k+1)+M^2]b_{0k}(LM|l)\\
&\qquad+(L+M)(l+M)b_{0k}(LM-1|l)\}.
\end{aligned} \tag{4.7}$$

This is applicable even for $M=0$ provided that the symmetry property of $b_{Kk}(LM|l)$ expressed by Eq. (3.5) is taken into consideration. Especially, Eq. (4.7) for $k=0$ is reduced to a strikingly simple formula owing to Eq. (2.10)

$$b_{01}(LM|l) = -\frac{L(l-1)+l(L-1)+2M^2}{2l-1}. \tag{4.8}$$

Thus all the first-row elements $b_{0k}(LM|l)$ with $M=0$ to L are successively evaluated in the increasing order of k by the repeated use of Eq. (4.7) with the aid of Eq. (4.8).

A recurrence formula for evaluating the first-column elements $b_{K0}(LM|l)$ can be also obtained by introducing into Eq. (4.7) the symmetry relation expressed by Eq. (3.6). The formula, however, will not be given here since it is not necessary for finding the values of those elements.

Finally, some of $b_{Kk}(LM|l)$ evaluated in the above manner are presented in the matrix form. The matrices $\mathbb{b}(LM|l)$ with the restriction, $0\leqslant M\leqslant L\leqslant l\leqslant 2$, are as follows:

$$\mathbf{b}(00|0) = (1)$$

$$\mathbf{b}(00|1) = \begin{pmatrix} 1 & 1\\ -1 & \times \end{pmatrix},$$

$$\mathbf{b}(10|1) = \begin{pmatrix} 1 & 0 & -1\\ 0 & 2 & \times\\ -1 & \times & \times \end{pmatrix}, \quad \mathbf{b}(11|1) = \begin{pmatrix} 1 & -2 & 1\\ -2 & -2 & \times\\ 1 & \times & \times \end{pmatrix},$$

$$\mathbf{b}(00|2)=\begin{pmatrix} 1 & \frac{2}{3} & 1 \\ -2 & -2 & \times \\ 1 & \times & \times \end{pmatrix},$$

$$\mathbf{b}(10|2)=\begin{pmatrix} 1 & -\frac{1}{3} & \frac{1}{3} & -1 \\ -1 & \frac{2}{3} & 3 & \times \\ -1 & -3 & \times & \times \\ 1 & \times & \times & \times \end{pmatrix},\quad \mathbf{b}(11|2)=\begin{pmatrix} 1 & -1 & -1 & 1 \\ -3 & -2 & -3 & \times \\ 3 & 3 & \times & \times \\ -1 & \times & \times & \times \end{pmatrix},$$

$$\mathbf{b}(20|2)=\begin{pmatrix} 1 & -\frac{4}{3} & \frac{2}{3} & -\frac{4}{3} & 1 \\ -\frac{4}{3} & \frac{4}{9} & \frac{4}{3} & -4 & \times \\ \frac{2}{3} & \frac{4}{3} & 6 & \times & \times \\ -\frac{4}{3} & -4 & \times & \times & \times \\ 1 & \times & \times & \times & \times \end{pmatrix},\quad \mathbf{b}(21|2)=\begin{pmatrix} 1 & -2 & 0 & 2 & -1 \\ -2 & 0 & -2 & 4 & \times \\ 0 & -2 & -6 & \times & \times \\ 2 & 4 & \times & \times & \times \\ -1 & \times & \times & \times & \times \end{pmatrix},$$

$$\mathbf{b}(22|2)=\begin{pmatrix} 1 & -4 & 6 & -4 & 1 \\ -4 & 4 & 4 & -4 & \times \\ 6 & 4 & 6 & \times & \times \\ -4 & -4 & \times & \times & \times \\ 1 & \times & \times & \times & \times \end{pmatrix}. \tag{4.9}$$

5. ASYMPTOTIC EXPRESSION FOR $\alpha_l(fLM|a,r)$ WHEN $r\approx 0$

In this section, for $r\approx 0$ we will derive a more practical expression for $\alpha_l(fLM\,|a,r)$.

When $r\approx 0$, the direct integration of Eq. (2.12) is not adequate to calculate $\alpha_l(fLM\,|a,r)$, because numerically it yields an extreme reduction of the significant figures due to appearance of a difference between two nearly equal values. Therefore, in such a case the formula derived by applying Taylor's theorem to the integral in Eq. (2.12) should be used instead of Eq. (2.12). It may be written as

$$\alpha_l(fLM\,|a,r)=2\gamma(LM\,|l\,)\sum_{n=l}^{\infty}\left(\frac{r}{a}\right)^{2n-l+1} \times\sum_{k=0}^{\min\{L+l,n\}}\sum_{K=0}^{L+l-k} b_{Kk}(LM\,|l\,) \times E^{2K}_{2n-2k}(a), \tag{5.1}$$

where

$$E^{2K}_{2n-2k}(a) \equiv \frac{a^{2n-2(K+k)+L}}{(2n-2k+1)!}\left[\left(\frac{d}{dR}\right)^{2n-2k} R^{\,2K}g(R\,)\right]_{R=a} \tag{5.2}$$

with

$$g(R\,)\equiv f(R\,)/R^{\,L-1}. \tag{5.3}$$

Here it should be noted that the terms with $n<l$ never arise in Eq. (5.1) because every inner double sum in those terms necessarily vanishes.[2]

The inner sum, denoted by $\alpha_{l,2n-l}(fLM\,|a)$, may be expressed as

$$\alpha_{l,2n-l}(fLM\,|a)=\sum_{s=s_{\min}}^{2n} h_{n,s}(LM\,|l\,)a^{s+L}g^{(s)}(a)/s!, \tag{5.4}$$

where

$$s_{\min}=\max\{0,2n-2(L+l\,)\}, \tag{5.5}$$

$$h_{n,s}(LM\,|l\,) = \sum_{k=0}^{[(2n-s)/2]}\ \sum_{K=[(2n-s+1)/2]-k}^{L+l-k} \{(2K\,)! \times[(2n-2k+1)(2n-2k-s)!(2K-2n+2k+s)!]^{1/2}\}b_{Kk}(LM\,|l\,), \tag{5.6}$$

and

$$g^{(s)}(a)=\left[\left(\frac{d}{dR}\right)^{s}g(R\,)\right]_{R=a}. \tag{5.7}$$

$[q/2]$ denotes the quotient of $q/2$. For $s=2n$, Eq. (5.6) can be reduced to

$$h_{n,2n}(LM\,|l\,)=\frac{1}{2n+1}\sum_{K=0}^{L+l} b_{K0}(LM\,|l\,). \tag{5.8}$$

From Eq. (3.13) this sum vanishes so long as $l\neq 0$.

When assumed that $g(R\,)$ is expressed by a polynomial of R, $\alpha_{l,2n-l}(fLM\,|a)$ becomes a polynomial of a with the same order as $a^Lg(a)$ irrespective of n. Therefore, $\alpha_{l,2n-l}(fLM\,|a)$ never increases rapidly in magnitude with the increase of n. Thus, it is ensured that if r/a is sufficiently near to zero, $\alpha_l(fLM\,|l\,)$ converges within a finite value of n.

6. DISCUSSION

In this final section we will give some notes on calculation of molecular integrals by using $\alpha_l(LM\,|a,r)$.

At the outset we will consider how, when the spherical harmonic in the function to be expanded, denoted by $Y^M_L(\Theta',\Phi')$, is not quantized in the direction of center A to center B, $f(R\,)Y^M_L(\Theta',\Phi')$ is expressed in terms of $Y^m_l(\vartheta,\varphi)$.

Then $Y_L^M(\Theta',\Phi')$ may be expressed in terms of $Y_L^{M'}(\Theta,\Phi)$ quantized in the direction of A to B, as below:

$$Y_L^M(\Theta',\Phi') = \mathrm{Rot}\{Y_L^M(\Theta,\Phi)\} = \sum_{M'=-|M|}^{|M|} Y_L^{M'}(\Theta,\Phi) D_{M'M}^{(L)}(\mathrm{Rot}), \quad (6.1)$$

where Rot denotes the rotation which brings $Y_L^M(\Theta,\Phi)$ into $Y_L^M(\Theta',\Phi')$, and $D_{M'M}^{(L)}(\mathrm{Rot})$ the element on the $(L+M'+1)$th row and the $(L+M+1)$th column in the $2L+1$-dimensional irreducible representation matrix for Rot. Therefore, it is essential to carry out the expansion of $f(R)Y_L^{M'}(\Theta,\Phi)$ in terms of $Y_l^{M'}(\vartheta,\varphi)$ as shown by Eq. (2.1). By using Eq. (6.1), the expansion formula for $f(R)Y_L^M(\Theta',\Phi')$ is given by

$$f(R)Y_L^M(\Theta',\Phi') = \sum_{M'=-|M|}^{|M|} D_{M'M}^{(L)}(\mathrm{Rot}) \times \sum_{l=|M'|}^{\infty} \left(\frac{1}{r}\right) \alpha_l(fLM'|a,r) Y_l^{M'}(\vartheta,\varphi). \quad (6.2)$$

Second we emphasize that, since the products of $b_{Kk}(LM|l)$ with $\gamma(LM|l)$ are irrespective of the form of $f(R)$ as seen in Eqs. (2.6), (2.7), and (2.9), those products can be utilized even for any $f(R)$ once they have been evaluated.

Finally we will give a note on the computation of molecular integrals between atomic orbitals around separate centers. In such computation, integration over r must be carried out. Then the integration region is divided into three subregions, i.e., $r \geqslant a$, $a > r \geqslant \epsilon$, and $\epsilon > r \geqslant 0$, where a definite value of ϵ must be determined from the result of examination of the convergence of Eq. (5.1). Thus for each subregion each integrand is necessary. Any explicit form of those integrands, however, cannot be obtained as far as a definite form of $f(R)$ is not determined. How those integrands are expressed when $f(R)$ takes a special form such as Slater-type orbital and further how the coefficients appearing there can be evaluated will be presented in a separate paper.

ACKNOWLEDGMENTS

The author thanks Dr. Osamu Matsuoka for directing his attention to Silvertone and Moats' expansion formula.

APPENDIX

In general, the sum S_n expressed by

$$S_n \equiv \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{(2p-2k-1)!!}{(2q-2k-1)!!} \quad (A1)$$

can be rewritten as

$$S_n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \times \left[\Gamma_0^{p-q} + (1-\delta_{p,q}) \sum_{i=1}^{p-q} \Gamma_i^{p-q} \times \prod_{\mu=1}^{i} (2k - 2\mu + 2) \right], \quad (A2)$$

where

$$\Gamma_i^{p-q} \equiv (-1)^i \binom{p-q}{i} \frac{(2p-2i-1)!!}{(2q-1)!!}. \quad (A3)$$

Here it has been assumed that $p \geqslant q$. It should be noted that the equality of Eq. (A2) holds valid even if $p-k$ and/or $q-k$ are negative. The repeated use of the relation

$$\binom{n-\mu}{k-\mu}(k-\mu) = (n-\mu)\binom{n-\mu-1}{k-\mu-1} \quad (A4)$$

leads to

$$S_n = \sum_{i=0}^{\min\{n,p-q\}} (-1)^i \frac{(2n)!!}{(2n-2i)!!} \times \Gamma_i^{p-q} \sum_{k=0}^{n-i} (-1)^k \binom{n-i}{k}. \quad (A5)$$

As easily seen from the above equation, if and only if $i = n$, the inner sum remains nonzero. Otherwise it vanishes. Thus it follows that

$$S_n = \begin{cases} 0 & \text{if } n > p-q. \\ (-1)^n (2n)!! \Gamma_n^{p-q} & \text{otherwise.} \end{cases} \quad (A6)$$

[1]M. P. Barnett and C. A. Coulson, Philos. Trans. R. Soc. London, Ser. A **243**, 221 (1951).
[2]P.-O. Löwdin, Adv. Phys. **5**, 96 (1956).
[3]Y. Tanabe, J. Phys. Soc. Jpn. **11**, 980 (1956).
[4]F. E. Harris and H. H. Michels, J. Chem. Phys. **43**, S165 (1965).
[5]H. J. Silverstone, J. Chem. Phys. **47**, 537 (1967).
[6]R. R. Sharma, J. Math. Phys. **9**, 505 (1968).
[7]E. O. Steinborn and E. Filter, Theor. Chim. Acta **38**, 247 (1975).
[8]R. R. Sharma, Phys. Rev. A **13**, 517 (1976).
[9]H. J. Silverstone and R. K. Moats, Phys. Rev. A **16**, 1731 (1977).
[10]H. W. Jones and C. A. Weatherford, Int. J. Quantum Chem. **S12**, 483 (1978).
[11]I. I. Guseinov, J. Chem. Phys. **69**, 4990 (1978).
[12]For example, M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957).
[13]See Eqs. (3.16a) and (3.16b) in Ref. 12.
[14]See Eq. (3.18) or (3.19) in Ref. 12.
[15]See Eq. (I.1) in Appendix I of Ref. 12.

Exact invariants quadratic in the momentum for a particle in a three-dimensional electromagnetic field

H. Ralph Lewis
Los Alamos National Laboratory, Los Alamos, New Mexico 87575

(Received 3 June 1982; accepted for publication 9 November 1983)

For a Hamiltonian of the form $H = \frac{1}{2}(\mathbf{p} - \mathbf{A}(\mathbf{r},t))^2 + V(\mathbf{r},t)$, the conditions on the scalar and vector potentials $V(\mathbf{r},t)$ and $\mathbf{A}(\mathbf{r},t)$ have been found such that there exists an exact invariant that is quadratic in each component of the canonical momentum $\mathbf{p}$. The invariant is written in the form $I(\mathbf{r},\mathbf{p},t) = \frac{1}{2}(\mathbf{f}_2(\mathbf{r},t)\cdot\mathbf{p} - \mathbf{f}_1(\mathbf{r},t))^2 + f_0(\mathbf{r},t)$, where $\mathbf{f}_2$ is a nonsingular symmetric real dyad, $\mathbf{f}_1$ is a real vector, and f_0 is a real scalar. For the cases in which $\mathbf{f}_2$ is proportional to the unit dyad, or the vector potential is identically zero and $\mathbf{f}_2$ commutes with the time derivative of $\mathbf{f}_2$, all of the potentials that satisfy the conditions for such invariants to exist have been found explicitly and the invariants have been found in terms of the potentials. The derivation is a generalization of a method reported by Lewis and Leach for the corresponding one-dimensional problem. For the case in which the vector potential vanishes identically, the result is a generalization of a result found in three dimensions by Chandrasekhar. The results presented are applicable in plasma physics and stellar dynamics.

PACS numbers: 41.70. + t, 02.30.Tb

I. INTRODUCTION

Lewis and Leach[1] have reported a method with which they found all one-dimensional potentials for which there exists an exact invariant that is quadratic in the momentum; for those potentials, they found the invariants. In this paper a generalization to three dimensions is given for the Hamiltonian

$$H = \tfrac{1}{2}(\mathbf{p} - \mathbf{A}(\mathbf{r},t))^2 + V(\mathbf{r},t), \tag{1}$$

where $V(\mathbf{r},t)$ and $\mathbf{A}(\mathbf{r},t)$ are scalar and vector potentials, respectively, as functions of position vector $\mathbf{r}$ and time t, and $\mathbf{p}$ is the canonical momentum vector. For this Hamiltonian, the conditions on the scalar and vector potentials are derived such that there exists an exact invariant that is quadratic in each component of the canonical momentum vector.

The invariant is written in the form

$$I(\mathbf{r},\mathbf{p},t) = \tfrac{1}{2}(\mathbf{f}_2(\mathbf{r},t)\cdot\mathbf{p} - \mathbf{f}_1(\mathbf{r},t))^2 + f_0(\mathbf{r},t), \tag{2}$$

where $\mathbf{f}_2$ is a nonsingular symmetric real dyad, $\mathbf{f}_1$ is a real vector, and f_0 is a real scalar. It is shown at the beginning of the derivation that $\mathbf{f}_2$ must depend on t only. For the cases in which $\mathbf{f}_2$ is proportional to the unit dyad, or the vector potential vanishes identically and $\mathbf{f}_2$ commutes with the time derivative of $\mathbf{f}_2$, all of the potentials that satisfy the conditions for such invariants to exist have been found explicitly and the invariants have been found in terms of the potentials. For the case in which the vector potential vanishes identically, the result is a generalization of a result found in three dimensions by Chandrasekhar,[2] who considered the problem of what scalar potentials in three dimensions admit a certain class of invariants quadratic in the momentum. Although he did not completely solve the problem of satisfying all of the compatability conditions that arose in his analysis, he discussed various cases within the context of stellar dynamics. The results presented are applicable in plasma physics and stellar dynamics. The one-dimensional result of Lewis and Leach has been used to derive exact time-dependent solutions of the Vlasov–Poisson equations.[3] The three-dimensional result presented here can be used to derive exact time-dependent solutions of the Vlasov–Maxwell equations.

In order to avoid confusion about the notation used in the derivation, we begin with some explanatory remarks. The vector $\mathbf{p}$ is the canonical momentum vector, which means that the velocity $\mathbf{v}$ of a particle whose motion is governed by the Hamiltonian (1) is

$$\mathbf{v} = \mathbf{p} - \mathbf{A}(\mathbf{r},t). \tag{3}$$

The scalar canonical momenta conjugate to the coordinates of a curvilinear coordinate system, as usually defined in Hamiltonian dynamics, are proportional to, but not generally equal to, the components of $\mathbf{p}$ in that coordinate system. The scalar canonical momenta are neither used in the derivation in this paper nor in the statement of results. Nevertheless, confusion may arise because the symbols that usually denote the scalar canonical momenta are the same as those that are normally used to denote the components of a vector $\mathbf{p}$. As an illustrative example, consider a cylindrical coordinate system with radial, angular, and axial coordinates (r,θ,z). The r- and z-components of $\mathbf{p}$ are $\dot{r} + A_r(\mathbf{r},t)$ and $\dot{z} + A_z(\mathbf{r},t)$, respectively, where $\dot{}$ denotes differentiation with respect to t. These are also the canonical momenta conjugate to the coordinates r and z. However, the θ-component of $\mathbf{p}$ is $r\dot{\theta} + A_\theta(\mathbf{r},t)$, which is $1/r$ times the canonical momentum conjugate to the coordinate θ.

The symbol ∇ is used to denote the usual gradient operator with respect to the vector $\mathbf{r}$. However, when there are two vector arguments of a function, then ∇ is used with a subscript to denote the vector with respect to which the differentiation is to occur. For example, let $f(\mathbf{r},\mathbf{p})$ be an arbitrary differentiable function of the vectors $\mathbf{r}$ and $\mathbf{p}$. The definition of the operators $\nabla_{\mathbf{r}}$ and $\nabla_{\mathbf{p}}$ is

$$\delta f = \delta \mathbf{r}\cdot\nabla_{\mathbf{r}} f + \delta \mathbf{p}\cdot\nabla_{\mathbf{p}} f, \tag{4}$$

where δf is the differential change of f associated with differential changes $\delta\mathbf{r}$ and $\delta\mathbf{p}$ of $\mathbf{r}$ and $\mathbf{p}$. $\nabla_{\mathbf{r}}$ is the gradient operator with respect to $\mathbf{r}$ holding the *vector* $\mathbf{p}$ fixed; $\nabla_{\mathbf{p}}$ is the gradient operator with respect to $\mathbf{p}$ holding the *vector* $\mathbf{r}$ fixed. Hamilton's equations can be expressed in the following equivalent vector form that is valid for an arbitrary curvilinear coordinate system:

$$\dot{\mathbf{r}} = \nabla_{\mathbf{p}} H, \tag{5a}$$

$$\dot{\mathbf{p}} = -\nabla_{\mathbf{r}} H. \tag{5b}$$

A detailed expository discussion of the canonical momentum vector, the general meaning of these gradient operators, and the vector form of Hamilton's equations can be found in Sec. II and Appendix A of Ref. 4.

The derivation of invariants of the form (2) for the Hamiltonian (1) is presented in Sec. II.

II. THE DERIVATION

The definition of an invariant $I(\mathbf{r},\mathbf{p},t)$ is that the total time derivative of I vanish; for the Hamiltonian given by (1), we have

$$\begin{aligned}\frac{dI}{dt} &= \frac{\partial I}{\partial t} + [I,H]\\ &= \frac{\partial I}{\partial t} + (\nabla_{\mathbf{p}} H)\cdot\nabla_{\mathbf{r}} I - (\nabla_{\mathbf{r}} H)\cdot\nabla_{\mathbf{p}} I\\ &= \frac{\partial I}{\partial t} + (\mathbf{p}-\mathbf{A})\cdot\nabla_{\mathbf{r}} I - (\nabla V)\cdot\nabla_{\mathbf{p}} I + (\nabla_{\mathbf{p}} I)\cdot(\nabla\mathbf{A})\cdot(\mathbf{p}-\mathbf{A})\\ &= 0.\end{aligned} \tag{6}$$

The most general real invariant quadratic in each component of the canonical momentum $\mathbf{p}$ can be written in the form given by (2),

$$I(\mathbf{r},\mathbf{p},t) = \tfrac{1}{2}(\mathbf{f}_2(\mathbf{r},t)\cdot\mathbf{p} - \mathbf{f}_1(\mathbf{r},t))^2 + f_0(\mathbf{r},t), \tag{7}$$

where $\mathbf{f}_2$ is a nonsingular symmetric real dyad, $\mathbf{f}_1$ is a real vector, and f_0 is a real scalar, which may depend on $\mathbf{r}$ and t. By using this form in (6), we obtain

$$\begin{aligned}\frac{dI}{dt} &= (\dot{\mathbf{f}}_2\cdot\mathbf{p} - \dot{\mathbf{f}}_1)\cdot(\mathbf{f}_2\cdot\mathbf{p} - \mathbf{f}_1) + \dot{f}_0\\ &\quad + (\mathbf{p}-\mathbf{A})\cdot\{\tfrac{1}{2}[\nabla\mathbf{f}_2^2]:\mathbf{pp} - [\nabla(\mathbf{f}_1\cdot\mathbf{f}_2)]\cdot\mathbf{p} + \nabla(\tfrac{1}{2}f_1^2 + f_0)\}\\ &\quad - (\nabla V)\cdot\mathbf{f}_2\cdot(\mathbf{f}_2\cdot\mathbf{p} - \mathbf{f}_1) + (\mathbf{p}\cdot\mathbf{f}_2 - \mathbf{f}_1)\cdot\mathbf{f}_2\cdot(\nabla\mathbf{A})\cdot(\mathbf{p}-\mathbf{A})\\ &= 0,\end{aligned} \tag{8}$$

where $\mathbf{f}_2^2 \equiv \mathbf{f}_2\cdot\mathbf{f}_2$, $f_1^2 \equiv \mathbf{f}_1\cdot\mathbf{f}_1$, and a dot over a symbol denotes partial differentiation with respect to t. Because dI/dt must equal zero for all values of $\mathbf{p}$, the coefficients of $\mathbf{ppp}$, $\mathbf{pp}$, and $\mathbf{p}$ must vanish and the part of dI/dt independent of $\mathbf{p}$ must vanish also. This implies the following conditions that are equivalent to (8):

$$\nabla\mathbf{f}_2^2 = 0, \tag{9a}$$

$$\frac{1}{2}\frac{\partial}{\partial t}\mathbf{f}_2^2 - \frac{1}{2}\mathbf{A}\cdot\nabla\mathbf{f}_2^2 + \mathbf{f}_2^2\cdot(\nabla\mathbf{A}) - \nabla(\mathbf{f}_2\cdot\mathbf{f}_1) = 0, \tag{9b}$$

$$\begin{aligned}&\frac{\partial}{\partial t}(\mathbf{f}_2\cdot\mathbf{f}_1) - \mathbf{A}\cdot\nabla(\mathbf{f}_2\cdot\mathbf{f}_1) - \nabla\Big(\frac{1}{2}f_1^2 + f_0\Big)\\ &\quad + \mathbf{f}_2^2\cdot\nabla\Big(V + \frac{1}{2}A^2\Big) + (\mathbf{f}_2\cdot\mathbf{f}_1)\cdot(\nabla\mathbf{A}) = 0,\end{aligned} \tag{9c}$$

$$\begin{aligned}&\frac{\partial}{\partial t}\Big(\frac{1}{2}f_1^2 + f_0\Big) - \mathbf{A}\cdot\nabla\Big(\frac{1}{2}f_1^2 + f_0\Big)\\ &\quad + (\mathbf{f}_2\cdot\mathbf{f}_1)\cdot\nabla\Big(V + \frac{1}{2}A^2\Big) = 0.\end{aligned} \tag{9d}$$

The solution of (9a) is that $\mathbf{f}_2$ be independent of $\mathbf{r}$,

$$\mathbf{f}_2 = \mathbf{f}_2(t). \tag{10}$$

Taking the curl of (9b), we find

$$\nabla\times(\mathbf{f}_2^2\cdot\nabla\mathbf{A}) = 0, \tag{11}$$

which implies that $\mathbf{f}_2^2\cdot\nabla\mathbf{A}$ is the gradient of a vector, as is readily apparent if $\mathbf{A}$ be written in Cartesian coordinates. Certain restrictions must be placed on $\mathbf{A}$ in order to satisfy this condition. We now determine those restrictions and calculate the vector $\mathbf{S}(\mathbf{r},t)$ whose gradient is $\mathbf{f}_2^2\cdot\nabla\mathbf{A}$,

$$\mathbf{f}_2^2\cdot\nabla\mathbf{A} = \nabla\mathbf{S}. \tag{12}$$

Let $\sigma(\mathbf{r},t)$ be a scalar function such that $\mathbf{f}_2^2\cdot\nabla\sigma$ is the gradient of a scalar $\tau(\mathbf{r},t)$,

$$\mathbf{f}_2^2\cdot\nabla\sigma = \nabla\tau. \tag{13}$$

In the present context, $\sigma(\mathbf{r},t)$ can be any Cartesian component of $\mathbf{A}(\mathbf{r},t)$. Later, when we consider the special case in which $\mathbf{A}(\mathbf{r},t)$ vanishes identically and $\mathbf{f}_2$ commutes with the time derivative of $\mathbf{f}_2$, we shall also take $\sigma(\mathbf{r},t)$ to be $V(\mathbf{r},t)$. Because $\mathbf{f}_2^2$ is real and symmetric, its representation in a Cartesian coordinate system can be diagonalized by a similarity transformation. Because $\mathbf{f}_2^2$ is only a function of t, such a similarity transformation can be viewed as a linear transformation, parametrized by t, between two Cartesian representations. If the new Cartesian coordinates are denoted by (η_1, η_2, η_3), then (13) is equivalent to

$$\frac{\partial}{\partial\eta_i}(\lambda_i\sigma - \tau) = 0, \quad i = 1,2,3, \tag{14}$$

where $\lambda_i(t)$ is the ith eigenvalue of $\mathbf{f}_2^2$. For (14) to be satisfied, there must exist functions $g_i(\eta_j, \eta_k)$, where (i,j,k) is a cyclic permutation of (1,2,3), such that

$$\begin{aligned}\lambda_1\sigma - \tau &= g_1(\eta_2,\eta_3),\\ \lambda_2\sigma - \tau &= g_2(\eta_3,\eta_1),\\ \lambda_3\sigma - \tau &= g_3(\eta_1,\eta_2).\end{aligned} \tag{15}$$

The function g_i cannot depend on η_i; however, it may depend on t. In order that these three equations can be satisfied, a condition among the eigenvalues λ_i and the functions g_i must be satisfied. The condition, obtained by eliminating τ and σ from (15), is

$$\begin{aligned}&(\lambda_1 - \lambda_2)g_3(\eta_1,\eta_2) + (\lambda_3 - \lambda_1)g_2(\eta_3,\eta_1)\\ &\quad + (\lambda_2 - \lambda_3)g_1(\eta_2,\eta_3) = 0.\end{aligned} \tag{16}$$

In some cases, (16) leads to a restriction on the functional form of σ. Because σ can be any Cartesian component of $\mathbf{A}(\mathbf{r},t)$, the restrictions on σ are also restrictions on the functional form of $\mathbf{A}(\mathbf{r},t)$. There are three cases.

(i) All of the eigenvalues of $\mathbf{f}_2^2$ are equal: $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$. Then (16) is satisfied identically and, from (15), all of the functions $g_i(\eta_j,\eta_k)$ are equal. Because g_i does not depend on η_i, the functions can be equal only if they do not depend on any of the variables (η_1, η_2, η_3): $g_i = c$, where c may only depend on t. Then Eqs. (15) are all solved by taking

$$\tau = \lambda\sigma - c, \tag{17}$$

where σ is arbitrary. No restriction is imposed by (13) on σ or $\mathbf{A}(\mathbf{r},t)$. The vector $\mathbf{S}(\mathbf{r},t)$ is

$$\mathbf{S}(\mathbf{r},t) = \mathbf{f}_2^2\cdot\mathbf{A} = \lambda(t)\mathbf{A}(\mathbf{r},t). \tag{18}$$

(ii) Two of the eigenvalues of $\mathbf{f}_2^2$ are equal to one another but unequal to the third: $\lambda_i = \lambda_j = \lambda \neq \lambda_k$, where (i,j,k) is a cyclic permutation of (1,2,3). Then, from (16), $g_i(\eta_j,\eta_k) = g_j(\eta_k,\eta_i)$, which is possible only if $g_i(\eta_j,\eta_k) = g_j(\eta_k,\eta_i) = g(\eta_k)$. The solution of (15) is

$$\sigma = \frac{g(\eta_k) - g_k(\eta_i,\eta_j)}{\lambda - \lambda_k}, \quad \tau = \frac{\lambda_k g(\eta_k) - \lambda g_k(\eta_i,\eta_j)}{\lambda - \lambda_k}, \tag{19}$$

where $g(\eta_k)$ and $g_k(\eta_i,\eta_j)$ are arbitrary. This restriction on σ means that $\mathbf{A}(\mathbf{r},t)$ must have the form

$$\mathbf{A}(\mathbf{r},t) = \mathbf{A}^{(1)}(\eta_k,t) + \mathbf{A}^{(2)}(\eta_i,\eta_j,t) \tag{20}$$

when $\lambda_i = \lambda_j = \lambda \neq \lambda_k$. The vector $\mathbf{S}(\mathbf{r},t)$ is

$$\mathbf{S}(\mathbf{r},t) = \lambda_k(t)\mathbf{A}^{(1)}(\eta_k,t) + \lambda(t)\mathbf{A}^{(2)}(\eta_i,\eta_j,t). \tag{21}$$

(iii) All of the eigenvalues of $\mathbf{f}_2^2$ are unequal: $\lambda_1 \neq \lambda_2 \neq \lambda_3$. By differentiating (16) with respect to η_k we obtain

$$(\lambda_k - \lambda_i)\frac{\partial g_j}{\partial \eta_k} = (\lambda_k - \lambda_j)\frac{\partial g_i}{\partial \eta_k}, \tag{22}$$

where (i,j,k) is a cyclic permutation of (1,2,3). Because g_j does not depend on η_j, and g_i does not depend on η_i, this is possible only if the expression on each side of (22) is only a function of η_k (and possibly t),

$$(\lambda_k - \lambda_i)\frac{\partial g_j}{\partial \eta_k} = \frac{\partial}{\partial \eta_k} h_k(\eta_k), \tag{23a}$$

$$(\lambda_k - \lambda_j)\frac{\partial g_i}{\partial \eta_k} = \frac{\partial}{\partial \eta_k} h_k(\eta_k), \tag{23b}$$

where $h_k(\eta_k)$ is arbitrary. The solution of (23a) must agree with the solution of (23b). The common solution is

$$(\lambda_k - \lambda_i)g_j(\eta_k,\eta_i) = h_k(\eta_k) - h_i(\eta_i). \tag{24}$$

The solution of (15) is

$$\sigma = \frac{(\lambda_2 - \lambda_3)h_1 + (\lambda_3 - \lambda_1)h_2 + (\lambda_1 - \lambda_2)h_3}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)}, \tag{25}$$

$$\tau = \frac{\lambda_1(\lambda_2 - \lambda_3)h_1 + \lambda_2(\lambda_3 - \lambda_1)h_2 + \lambda_3(\lambda_1 - \lambda_2)h_3}{(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)},$$

where the functions $h_i(\eta_i)$ are arbitrary. This restriction on σ means that $\mathbf{A}(\mathbf{r},t)$ must have the form

$$\mathbf{A}(\mathbf{r},t) = \mathbf{A}^{(1)}(\eta_1,t) + \mathbf{A}^{(2)}(\eta_2,t) + \mathbf{A}^{(3)}(\eta_3,t) \tag{26}$$

when $\lambda_1 \neq \lambda_2 \neq \lambda_3$. The vector $\mathbf{S}(\mathbf{r},t)$ is

$$\mathbf{S}(\mathbf{r},t) = \lambda_1(t)\mathbf{A}^{(1)}(\eta_1,t) + \lambda_2(t)\mathbf{A}^{(2)}(\eta_2,t) + \lambda_3(t)\mathbf{A}^{(3)}(\eta_3,t). \tag{27}$$

To continue with the solution of (9b)–(9d), it is convenient to define functions $\mathbf{f}$, y, and Z by

$$\begin{aligned} \mathbf{f} &= \mathbf{f}_2\cdot\mathbf{f}_1, \\ y &= \tfrac{1}{2}f_1^2 + f_0, \\ Z &= V + \tfrac{1}{2}A^2, \end{aligned} \tag{28}$$

and to write (9b)–(9d) in terms of them,

$$\frac{1}{2}\frac{d}{dt}\mathbf{f}_2^2 + \mathbf{f}_2^2\cdot(\nabla\mathbf{A}) - \nabla\mathbf{f} = 0, \tag{29a}$$

$$\frac{\partial\mathbf{f}}{\partial t} - \mathbf{A}\cdot\nabla\mathbf{f} - \nabla y + \mathbf{f}_2^2\cdot\nabla Z + \mathbf{f}\cdot\nabla\mathbf{A} = 0, \tag{29b}$$

$$\frac{\partial y}{\partial t} - \mathbf{A}\cdot\nabla y + \mathbf{f}\cdot\nabla Z = 0. \tag{29c}$$

We change variables from $(\mathbf{r},t)$ to $(\mathbf{x},t)$, where $\mathbf{x}$ and $\mathbf{r}$ are related by

$$\mathbf{x} = [\mathbf{r} - \boldsymbol{\alpha}(t)]\cdot\boldsymbol{\rho}^{-1}(t), \quad \mathbf{r} = \mathbf{x}\cdot\boldsymbol{\rho}(t) + \boldsymbol{\alpha}(t). \tag{30}$$

$\boldsymbol{\rho}(t)$ is a nonsingular time-dependent dyad and $\boldsymbol{\alpha}(t)$ is a time-dependent vector; they will be chosen to facilitate the solution of (29a)–(29c). The variables $(\mathbf{x},t)$ are appropriate for generalizing the one-dimensional result of Lewis and Leach.[1] With this change of variables, derivatives transform according to

$$\nabla_{\mathbf{r}} = \boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}, \tag{31a}$$

$$\begin{aligned} \left.\frac{\partial}{\partial t}\right|_{\mathbf{r}} &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} + \left.\frac{\partial\mathbf{x}}{\partial t}\right|_{\mathbf{r}}\cdot\nabla_{\mathbf{x}} \\ &= \left.\frac{\partial}{\partial t}\right|_{\mathbf{x}} - [(\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}})\cdot\boldsymbol{\rho}^{-1}]\cdot\nabla_{\mathbf{x}}. \end{aligned} \tag{31b}$$

The identity

$$\frac{d\boldsymbol{\rho}^{-1}}{dt} = -\boldsymbol{\rho}^{-1}\cdot\dot{\boldsymbol{\rho}}\cdot\boldsymbol{\rho}^{-1} \tag{32}$$

was used to obtain this formula. Henceforth, all functions will be regarded as functions of $(\mathbf{x},t)$ instead of $(\mathbf{r},t)$ and $\partial/\partial t$ will denote the partial derivative with respect to t holding $\mathbf{x}$ fixed.

Equations (29a)–(29c) written in terms of $(\mathbf{x},t)$ are

$$\frac{1}{2}\frac{d}{dt}\mathbf{f}_2^2 + \mathbf{f}_2^2\cdot\boldsymbol{\rho}^{-1}\cdot(\nabla_{\mathbf{x}}\mathbf{A}) - \boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}\mathbf{f} = 0, \tag{33a}$$

$$\begin{aligned} &\frac{\partial\mathbf{f}}{\partial t} - (\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}\mathbf{f} - \boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}y \\ &\qquad + \mathbf{f}_2^2\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}Z + \mathbf{f}\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}\mathbf{A} = 0, \end{aligned} \tag{33b}$$

$$\frac{\partial y}{\partial t} - (\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}y + \mathbf{f}\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}Z = 0. \tag{33c}$$

We can obtain a convenient equation for $\partial y/\partial t$ by eliminating $\nabla_{\mathbf{x}}y$ from (33c) by using (33b) and then eliminating $\nabla_{\mathbf{x}}\mathbf{A}$ by using (33a). The equation can be written in the form

$$\begin{aligned} &\frac{\partial y}{\partial t} - (\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\left(\frac{\partial\mathbf{f}}{\partial t} - \frac{1}{2}\mathbf{f}\cdot\mathbf{f}_2^{-2}\cdot\frac{d}{dt}\mathbf{f}_2^2\right) \\ &+ (\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\{[(\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\mathbf{f}_2^2 - \mathbf{f}]\cdot\mathbf{f}_2^{-2}\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}\mathbf{f}\} \\ &\qquad - [(\mathbf{x}\cdot\dot{\boldsymbol{\rho}} + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\mathbf{f}_2^2 - \mathbf{f}]\cdot\boldsymbol{\rho}^{-1}\cdot\nabla_{\mathbf{x}}Z = 0. \end{aligned} \tag{34}$$

By virtue of (12) and (31a), the general solution of (33a) is

$$\mathbf{f} = \mathbf{S} + \frac{1}{2}\mathbf{x}\cdot\boldsymbol{\rho}\cdot\frac{d}{dt}\mathbf{f}_2^2 + \text{const}, \tag{35}$$

where $\mathbf{S}$ is defined by (18), (21), and (27). If $\mathbf{f}_2$ is proportional to the unit dyad, so that (18) applies, or if $\mathbf{A}$ vanishes identically and the time derivative of $\mathbf{f}_2$ commutes with $\mathbf{f}_2$, then the coefficients of $\nabla_x \mathbf{f}$ and $\nabla_x Z$ in (34) can be made to vanish by suitable choice of $\boldsymbol{\rho}$; in that case, (34) can be solved easily. There may be other cases in which (34) can be solved, but, if so, they have not been found. Henceforth, we shall assume that either $\mathbf{f}_2$ is proportional to the unit dyad, or that $\mathbf{A}$ vanishes identically and the time derivative of $\mathbf{f}_2$ commutes with $\mathbf{f}_2$, or both. $\mathbf{f}_2$ and its time derivative will commute if $\mathbf{f}_2$ is diagonal. They will also commute if $\mathbf{f}_2$ can be diagonalized by a time-independent similarity transformation. We choose

$$\boldsymbol{\rho} = \mathbf{f}_2, \tag{36}$$

and choose the arbitrary constant in (35) so that the coefficients of $\nabla_x \mathbf{f}$ and $\nabla_x Z$ in (34) vanish,

$$\mathbf{f} = (\mathbf{x}\cdot\dot{\mathbf{f}}_2 + \dot{\boldsymbol{\alpha}} + \mathbf{A})\cdot\mathbf{f}_2^2. \tag{37}$$

Then (34) can be written in the form

$$\frac{\partial y}{\partial t} = \frac{1}{2}\frac{\partial}{\partial t}(\mathbf{f}\cdot\mathbf{f}_2^{-2}\cdot\mathbf{f}), \tag{38}$$

with the solution

$$y = U(\mathbf{x}) + \tfrac{1}{2}\mathbf{f}\cdot\mathbf{f}_2^{-2}\cdot\mathbf{f}, \tag{39}$$

where $U(\mathbf{x})$ is an arbitrary function of $\mathbf{x}$.

The result of substituting (37) and (39) into (33b) can be written as

$$\mathbf{f}_2^2\cdot\frac{\partial}{\partial t}(\mathbf{f}_2\cdot\mathbf{A}) = \nabla_x\left[y - \mathbf{x}\cdot\mathbf{f}_2^2\cdot\frac{\partial}{\partial t}\left[\frac{1}{4}\frac{d\mathbf{f}_2^2}{dt}\cdot\mathbf{x} + \mathbf{f}_2\cdot\dot{\boldsymbol{\alpha}}\right]\right] - \mathbf{f}_2^2\cdot\nabla_x\left(V + \frac{1}{2}A^2\right). \tag{40}$$

This is a condition on V and $\mathbf{A}$; when it is satisfied and $\mathbf{f}$ and y are given by (37) and (39), then (33a)–(33c) are satisfied. We consider our two cases separately.

(i) $\mathbf{f}_2$ is proportional to the unit dyad: $\mathbf{f}_2 = \rho(t)\mathbf{1}$, where $\rho(t)$ is an arbitrary scalar function. Then (40) has the form

$$\rho^2\frac{\partial}{\partial t}(\rho\mathbf{A}) = \nabla_x\left\{U(\mathbf{x}) + \frac{1}{2}\rho^2|\dot\rho\mathbf{x} + \dot{\boldsymbol{\alpha}} + \mathbf{A}|^2 - \rho^2\mathbf{x}\cdot\frac{\partial}{\partial t}\left[\frac{1}{4}\frac{d\rho^2}{dt}\mathbf{x} + \rho\dot{\boldsymbol{\alpha}}\right] - \rho^2\left(V + \frac{1}{2}A^2\right)\right\}. \tag{41}$$

The solution is

$$\rho\mathbf{A} = \nabla_x\psi(\mathbf{x},t) + \boldsymbol{\beta}(\mathbf{x}), \tag{42}$$

where $\psi(\mathbf{x},t)$ and $\boldsymbol{\beta}(\mathbf{x})$ are arbitrary, and

$$\rho^2\left(V + \frac{1}{2}A^2\right) = U(\mathbf{x}) + \frac{1}{2}\rho^2|\dot\rho\mathbf{x} + \dot{\boldsymbol{\alpha}} + \mathbf{A}|^2 - \rho^2\mathbf{x}\cdot\frac{\partial}{\partial t}\left[\frac{1}{4}\frac{d\rho^2}{dt}\mathbf{x} + \rho\dot{\boldsymbol{\alpha}}\right] - \rho^2\frac{\partial\psi}{\partial t} + G(t), \tag{43}$$

where $U(\mathbf{x})$, $\rho(t)$, $\boldsymbol{\alpha}(t)$, and $G(t)$ are arbitrary. When $\mathbf{A}$ and V are given by (42) and (43) and $\mathbf{f}_2 = \rho(t)\mathbf{1}$, then $I(\mathbf{r},\mathbf{p},t)$ given by (2) is an invariant. According to (28), (30), (37), and (39), $\mathbf{f}_1$ and f_0 are given by

$$\mathbf{f}_1 = \rho(\dot\rho\mathbf{x} + \dot{\boldsymbol{\alpha}} + \mathbf{A}), \tag{44}$$

$$f_0 = U(\mathbf{x}), \tag{45}$$

where

$$\mathbf{x} = \frac{\mathbf{r} - \boldsymbol{\alpha}(t)}{\rho(t)}. \tag{46}$$

(ii) $\mathbf{A}$ vanishes identically and $\mathbf{f}_2$ commutes with the time derivative of $\mathbf{f}_2$. Then (40) has the form

$$0 = \nabla_x\left\{U(\mathbf{x}) + \frac{1}{2}(\mathbf{x}\cdot\dot{\mathbf{f}}_2 + \dot{\boldsymbol{\alpha}})\cdot\mathbf{f}_2^2\cdot(\dot{\mathbf{f}}_2\cdot\mathbf{x} + \dot{\boldsymbol{\alpha}}) - \mathbf{x}\cdot\mathbf{f}_2^2\cdot\frac{\partial}{\partial t}\left[\frac{1}{4}\frac{d\mathbf{f}_2^2}{dt}\cdot\mathbf{x} + \mathbf{f}_2\cdot\dot{\boldsymbol{\alpha}}\right]\right\} - \mathbf{f}_2^2\cdot\nabla_x V. \tag{47}$$

The last term must be the gradient with respect to $\mathbf{x}$ of some function,

$$\mathbf{f}_2^2\cdot\nabla_x V = \nabla_x W. \tag{48}$$

Because of (31a) and (36), this is equivalent to

$$\mathbf{f}_2^2\cdot\nabla_r V = \nabla_r W, \tag{49}$$

which is of the same form as (13). Thus, the discussion following (13) is applicable here. The restrictions on V are the following.

(a) If all of the eigenvalues of $\mathbf{f}_2^2$ are equal, $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, then there is no restriction on V and

$$W(\mathbf{x},t) = \lambda(t)V(\mathbf{r},t). \tag{50}$$

(b) If two of the eigenvalues of $\mathbf{f}_2^2$ are equal to one another but unequal to the third, $\lambda_i = \lambda_j = \lambda \neq \lambda_k$, where (i,j,k) is a cyclic permutation of (1,2,3), then V and W must be of the form

$$V(\mathbf{r},t) = V^{(1)}(\eta_k,t) + V^{(2)}(\eta_i,\eta_j,t), \tag{51}$$

$$W(\mathbf{x},t) = \lambda_k(t)V^{(1)}(\eta_k,t) + \lambda(t)V^{(2)}(\eta_i,\eta_j,t). \tag{52}$$

(c) If all of the eigenvalues of $\mathbf{f}_2^2$ are unequal, $\lambda_1 \neq \lambda_2 \neq \lambda_3$, then V and W must be of the form

$$V(\mathbf{r},t) = V^{(1)}(\eta_1,t) + V^{(2)}(\eta_2,t) + V^{(3)}(\eta_3,t), \tag{53}$$

$$W(\mathbf{x},t) = \lambda_1(t)V^{(1)}(\eta_1,t) + \lambda_2(t)V^{(2)}(\eta_2,t) + \lambda_3(t)V^{(3)}(\eta_3,t). \tag{54}$$

The solution of (47) is

$$W(\mathbf{x},t) = U(\mathbf{x}) + \frac{1}{2}(\mathbf{x}\cdot\dot{\mathbf{f}}_2 + \dot{\boldsymbol{\alpha}})\cdot\mathbf{f}_2^2\cdot(\dot{\mathbf{f}}_2\cdot\mathbf{x} + \dot{\boldsymbol{\alpha}}) - \mathbf{x}\cdot\mathbf{f}_2^2\cdot\frac{\partial}{\partial t}\left[\frac{1}{4}\frac{d\mathbf{f}_2^2}{dt}\cdot\mathbf{x} + \mathbf{f}_2\cdot\dot{\boldsymbol{\alpha}}\right] + G(t), \tag{55}$$

where $\mathbf{f}_2(t)$ is an arbitrary dyad that commutes with its time derivative, where $\boldsymbol{\alpha}(t)$ and $G(t)$ are arbitrary and where $U(\mathbf{x})$ must be consistent with the form of $W(\mathbf{x},t)$. If we denote by (ξ_1,ξ_2,ξ_3) the Cartesian coordinates of $\mathbf{x}$ that are associated with (η_1,η_2,η_3), which are the coordinates of $\mathbf{r}$ in the Cartesian representation in which $\mathbf{f}_2^2$ is diagonal, then the restrictions on $U(\mathbf{x})$ are the following.

(a) If $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, then there is no restriction on $U(\mathbf{x})$.

(b) If $\lambda_i = \lambda_j = \lambda \neq \lambda_k$, then $U(\mathbf{x})$ must be of the form

$$U(\mathbf{x}) = U^{(1)}(\xi_k) + U^{(2)}(\xi_i,\xi_j). \tag{56}$$

(c) If $\lambda_1 \neq \lambda_2 \neq \lambda_3$, then $U(\mathbf{x})$ must be of the form

$$U(\mathbf{x}) = U^{(1)}(\xi_1) + U^{(2)}(\xi_2) + U^{(3)}(\xi_3). \tag{57}$$

When $\mathbf{A}$ vanishes identically, $\mathbf{f}_2$ commutes with the time derivative of $\mathbf{f}_2$, and V and W are given by (50)–(57), then $I(\mathbf{r},\mathbf{p},t)$ given by (2) is an invariant. According to (28), (30), (37), and (39), $\mathbf{f}_1$ and f_0 are given by

$$\mathbf{f}_1 = (\mathbf{x}\cdot\dot{\mathbf{f}}_2 + \dot{\boldsymbol{\alpha}})\cdot\mathbf{f}_2, \tag{58}$$

$$f_0 = U(\mathbf{x}), \tag{59}$$

where

$$\mathbf{x} = \mathbf{f}_2^{-1}(t)\cdot[\mathbf{r} - \boldsymbol{\alpha}(t)]. \tag{60}$$

The derivation is now complete.

[1]H. R. Lewis and P. G. L. Leach, J. Math. Phys. **23**, 2371 (1982).

[2]S. Chandrasekhar, *Principles of Stellar Dynamics* (Dover, New York, 1960), Chaps. 3 and 4.

[3]H. R. Lewis and K. R. Symon, Phys. Fluids. **27**, 192 (1984).

[4]H. R. Lewis and K. R. Symon, J. Math. Phys. **20**, 413 (1979); also, Erratum in J. Math. Phys. **20**, 2372 (1979).

Green's functions at zero viscosity[a)]

H. M. Fried and J. Tessendorf
Department of Physics, Brown University, Providence, Rhode Island 02912

(Received 18 January 1983; accepted for publication 17 June 1983)

Fradkin-type propagator representations are written for solutions to Navier–Stokes and related equations, for arbitrary dimension D and arbitrary source geometry. In the limit of very small viscosity, velocity/vorticity solutions are given in terms of Cauchy position coordinates q of a particle advected by the velocity flow v, using a set of coupled equations for q and v. For localized point vortices in two dimensions, the vectors q become the time-dependent position coordinates of interacting vortices, and our equations reduce to those of the familiar, coupled vortex problem. The formalism is, however, able to discuss three-dimensional vortex motion, discrete or continuous, including the effects of vortex stretching. The mathematical structure of vortex stretching in a D-dimensional fluid without boundaries is conveniently described in terms of an SU(D) representation of these equations. Several simple examples are given in two dimensions, to anchor the method in the context of previously known, exact solutions. In three dimensions, vortex stretching effects are approximated using a previous "strong coupling" technique of particle physics, enabling one to build a crude model of the intermittent growth of enstrophy, which may signal the onset of turbulence. For isotropic turbulence, the possibility of a singularity in the inviscid enstrophy at a finite time is related to the behavior of a single function characterizing the intermittency.

PACS numbers: 47.30. + s, 47.10. + g, 03.40.Gc

I. INTRODUCTION

Motivated by recent work suggesting chaotic behavior of a particle advected by the velocity flow due to three or more interacting vortices,[1] we would like to present a formalism based directly on the coupled velocity/vorticity equations of Navier–Stokes (NS) theory, which should be capable of describing situations involving specified, externally produced vortices as well as the possibility of spontaneous vortex generation at low viscosity. Based on this formalism, in this paper we offer an alternate way of characterizing known, two-dimensional flows, and a somewhat crude picture of the onset of turbulence in an infinite, three-dimensional, inviscid fluid without rigid boundaries.

To the best of our knowledge, the representations derived here are new; they are based upon a Green's function method invented by Fradkin[2] in the context of scattering problems in potential theory and quantum field theory. Fradkin's original functional differential forms may readily be converted to an equivalent functional integral formalism, and when applied to the NS problem it turns out that a great simplification can be made in the limit of small viscosity. There, the necessary functional integral can be well represented by an extremum method, analogous to the small $\hbar$ or large N expansions of quantum field theory, with corrections given in ascending powers of viscosity. We expect the spontaneous generation of vortices to be associated with such corrections; however, the general structure for arbitrary spatial dimension D and arbitrary source geometry and time dependence is most simply discussed in the limit of zero viscosity.

In this formalism, solutions to the inviscid NS problem are given in terms of a time-dependent position vector of a fictitious particle, or of a passive marker, whose position and velocity are codetermined by the exact velocity flow, in a construction which emerges from the extremum calculation appropriate to the small viscosity limit. For a system of distinct, point vortices, these marker coordinates represent the time- and position-dependent coordinates of the vortices themselves. An interesting feature of this method, which explicitly couples velocity and vorticity flow of the NS system in D dimensions, is the natural appearance of what may be an underlying SU(D) symmetry, associated in a nontrivial way with the presence of vortex stretching.

In this paper we present an alternative way of characterizing known, two-dimensional velocity/vortex flows, along with some simple generalizations, and a somewhat crude picture of what may be the onset of turbulence in an infinite, three-dimensional, inviscid fluid without rigid boundaries. This is easiest to see for the case of isotropic turbulence, where there is but one length scale characterizing the spatial scale of vortex motion. In the nonisotropic case where (at least) two length scales enter, there is another function which must be determined, and which acts to damp the intermittent growth, and the possibility of a finite-time singularity. These results, special to three dimensions, follow from a representation of vortex stretching effects in terms of the growth of a nonunitary SU(3) matrix, whose components are related to time integrals of velocity gradients. With the aid of a "strong-coupling" technique previously used in particle physics, one can obtain approximate forms valid for large velocity gradients. The nonlinearities of the problem are still formidable; but, in a crude, dimensional way, it is easy to watch the growth of vorticity, and, in an intermittent way, as a function which essentially controls the

[a)] Supported in part by the U.S. Department of Energy under Contract DE-AC02-76ER03130.A009—Task A.

increase of enstrophy becomes large, not smoothly but in spurts.

The output of our method is closely related to solutions for vorticity suggested more than a century ago by Cauchy,[3] $\boldsymbol{\omega}(t) = \omega_i(t_0)\partial/\partial a_i \mathbf{X}(\mathbf{a},t)$, $\mathbf{X}(\mathbf{a},t_0) = \mathbf{a}$, in terms of the position coordinates $\mathbf{X}(\mathbf{a},t)$ of the moving, material fluid; our "Marker" coordinates q are in essence Cauchy's $\mathbf{X}$, generalized to arbitrary vorticity source distributions.

II. THE FORMALISM

We begin with the standard NS equation

$$\partial_t \mathbf{v} + (\mathbf{v}\cdot\nabla)\mathbf{v} - \nu\nabla^2\mathbf{v} = -\nabla p + \mathbf{f}(\mathbf{r},t), \tag{1}$$

where p denotes the fluid pressure, ν is the kinematic viscosity, and $\mathbf{f}$ represents a divergenceless velocity source. Taking the curl of (1) generates

$$\partial_t \boldsymbol{\omega} + (\mathbf{v}\cdot\nabla)\boldsymbol{\omega} - \nu\nabla^2\boldsymbol{\omega} - (\boldsymbol{\omega}\cdot\nabla)\mathbf{v} = \mathbf{g}(\mathbf{r},t), \tag{2}$$

where $\boldsymbol{\omega} = \nabla\times\mathbf{v}$ is the vorticity and $\mathbf{g}$ denotes a corresponding divergenceless vorticity source, $\mathbf{g} = \nabla\times\mathbf{f}$. For a specified velocity field $\mathbf{v}(\mathbf{r},t)$, (2) is linear in $\boldsymbol{\omega}$, and can be solved in terms of an appropriate, if formal, functional representation, $\boldsymbol{\omega} = \boldsymbol{\omega}\{v\}$. That solution must then be combined with the relation $\boldsymbol{\omega} = \nabla\times\mathbf{v}$, or its inverse $\mathbf{v} = \mathbf{v}_0 - \nabla^{-2}\nabla\times\boldsymbol{\omega}$, where $\mathbf{v}_0$ is some specified, initial velocity field in the absence of $\mathbf{g}$ and $\boldsymbol{\omega}$, satisfying $\nabla\times\mathbf{v}_0 = 0$. Other methods of resolution of (2) are possible, but this is the simplest for our purposes; in fact, we shall further simplify matters by supposing that $\mathbf{v}_0$ denotes a constant, not-too-large velocity flow. A vorticity source of arbitrary strength $\mathbf{g}(\mathbf{r},t)$ is then turned on at $t=0$, and we ask for the subsequent velocity/vorticity flow.

The corresponding solution to (2), for specified $\mathbf{v}(\mathbf{r},t)$ and under the assumption that $\boldsymbol{\omega}$ vanishes for all $t<0$, is given by

$$\omega_a(\mathbf{r},t) = \int d^D y \int dy_0\, G_{ab}(\mathbf{r},\mathbf{y};t,y_0|v) g_b(\mathbf{y},y_0), \tag{3}$$

where $G[v]$ denotes that casual Green's function which satisfies[4]

$$\sum_{b=1}^{D} \{[\partial_t + (\mathbf{v}\cdot\nabla) - \nu\nabla^2]\delta_{ab} - \partial_b v_a\} G_{bc}(\mathbf{r},\mathbf{y};t,y_0|v) = \delta^D(\mathbf{r}-\mathbf{y})\delta(t-y_0)\delta_{ac} \tag{4}$$

in D spatial dimensions. Using an obvious shorthand, one can write the formal equivalent of (4)

$$([\partial_t + (\mathbf{v}\cdot\nabla) - \nu\nabla^2] - (\partial v))G[v] = 1, \tag{5}$$

and a corresponding formal representation

$$G[v] = \int_0^\infty ds \exp\{-s([\partial_t + (\mathbf{v}\cdot\nabla) - \nu\nabla^2] - (\partial v))\}. \tag{6}$$

It will become clear, subsequently, that (6) does indeed define a retarded Green's function. The symbol $\mathbf{v}$ in (5) and (6) now denotes an operator with matrix elements diagonal in configuration space and time, $\langle \mathbf{x},x_0|\mathbf{v}|\mathbf{y},y_0\rangle = \mathbf{v}(\mathbf{x},x_0)\delta^D(\mathbf{x}-\mathbf{y})\delta(x_0-y_0)$, while ∂_+ and ∇ retain their customary operator meanings.

It is convenient to rewrite the square bracket of (6) as

$$[\partial_t - \nu(\nabla - \mathbf{v}/2\nu)^2 + \mathbf{v}^2/4\nu],$$

since it is assumed that $\mathbf{v}$ is divergenceless, $\Sigma_a[\partial_a, v_a] = 0$; all the complications of the representation that follow arise because a general component of $\mathbf{v}(\mathbf{x},x_0)$ does not commute with ∇, or with ∇^2. With this step, one has

$$G[v] = \int_0^\infty ds \times\exp\left\{-s\left(\left[\partial_t - \nu\left(\nabla - \frac{\mathbf{v}}{2\nu}\right)^2 + \frac{\mathbf{v}^2}{4\nu}\right] - (\partial v)\right)\right\}. \tag{7}$$

The essence of the Fradkin method is to introduce an auxiliary, s-dependent field whose fluctuations reproduce the effects of noncommutivity of (∂_t,∇) and $\mathbf{v}$. Thus, the integrand of (7) is replaced by

$$U(s) = \left(\exp\left\{-\int_0^s ds' \left(\left[\partial_t - \nu\left(\nabla - \frac{\mathbf{v}}{2\nu}\right)^2 + \frac{\mathbf{v}^2}{4\nu} - \nu\mathbf{u}(s')\cdot\left(\nabla - \frac{\mathbf{v}}{2\nu}\right)\right] - (\partial v)\right)\right\}\right)_+, \tag{8}$$

where $\mathbf{u}(s')$ denotes a D-dimensional vector whose parametric dependence gives meaning to the s'-ordered exponential of (8); that is, in the expansion of $(\exp[-\int_0^s ds' A(s')])_+$, the noncommuting terms $A(s_1)\cdots A(s_n)$ are to be arranged in an ordered fashion, with those bearing the larger values of s_i standing to the left.[5] In terms of (8), (7) may be rewritten as

$$G[v] = \int_0^\infty ds\, U(s)|_{\mathbf{u}=0}. \tag{9}$$

The advantage of this procedure is that differential equations can now be written for $U(s)$, as a function of s and a functional of $\mathbf{u}$:

$$-\frac{\partial U}{\partial s} = \left(\left[\partial_t - \nu\left(\nabla - \frac{\mathbf{v}}{2\nu}\right)^2 + \frac{\mathbf{v}^2}{4\nu} - \nu\mathbf{u}(s)\cdot\left(\nabla - \frac{\mathbf{v}}{2\nu}\right)\right] - (\partial v)\right)U(s),$$

and

$$\frac{\delta U}{\delta \mathbf{u}(s)} = \nu\left(\nabla - \frac{\mathbf{v}}{2\nu}\right)U,$$

under the boundary condition $U(s=0)=1$. The solution to this pair of equations can be written in the form

$$U(s) = \exp\left[\frac{1}{\nu}\int_0^{s-\epsilon} ds' \frac{\delta^2}{\delta\mathbf{u}^2(s')}\right]\cdot V(s), \tag{10}$$

where

$$V(s) = \left(\exp\left\{-\int_0^s ds' \times([\partial_t + \phi(\mathbf{x},x_0;s') - \nu\mathbf{u}(s')\cdot\nabla] - (\partial v))\right\}\right)_+,$$

with ϵ a small, positive parameter subsequently set equal to zero, and $\phi(\mathbf{x},x_0;s') = \mathbf{v}^2(\mathbf{x},x_0)/4\nu + \frac{1}{2}\mathbf{u}(s')\cdot\mathbf{v}(\mathbf{x},x_0)$. Strictly speaking, one cannot display $(\mathbf{x},x_0)$ dependence before taking matrix elements of $G[v]$, and at this stage it should be understood that we are calculating $\langle\mathbf{x},x_0|G[v]$, with the $V(s)$ of (10) replaced by $\langle\mathbf{x},x_0|V(s)$.

To find a representation for $V(s)$, one may set

$$V(s) = \exp\left\{-\int_0^s ds' [\partial_t - \nu\mathbf{u}(s')\cdot\nabla]\right\}\cdot W(s), \tag{11}$$

so that $W(s)$ satisfies the equation

$$-\frac{\partial W}{\partial s}=\exp\left\{-\int_0^s ds'\,[\partial_t-\nu\mathbf{u}(s')\cdot\nabla]\right\}$$
$$\times\{\phi(\mathbf{x},x_0;s)-[\partial v(\mathbf{x},x_0)]\}$$
$$\times\exp\left\{+\int_0^s ds'\,[\partial_t-\nu\mathbf{u}(s')\cdot\nabla]\right\}\cdot W,$$

or

$$-\frac{\partial W}{\partial s}=\left[\phi\left(\mathbf{x}-\nu\int_0^s ds'\,\mathbf{u}(s'),x_0+s;s\right)\right.$$
$$\left.-\left(\partial v(\mathbf{x}-\nu\int_0^s ds'\,\mathbf{u}(s'),x_0+s)\right)\right]\cdot W. \tag{12}$$

Because of the particular sequence of translational operators written in (12), the $(\mathbf{x},x_0)$ dependence inside $W(s)$ is the same on both sides of (12); and hence (12) represents a differential equation that can be solved in terms of actual functions rather than formal operators.

Consider now the tensor quantity $\partial_b v_a(\mathbf{x},x_0)$, written in matrix notation as $Q_{ab}(\mathbf{x},x_0)$. Because $\mathbf{v}$ is assumed divergenceless, $\Sigma_a Q_{aa}=0$, and this $D\times D$ matrix is traceless. But any such matrix can be written in terms of the fundamental, or defining representation of SU(D),

$$Q_{ab}=\partial_b v_a=\sum_{i=1}^{D^2-1}(\lambda_i)_{ab}\psi_i(\mathbf{x},x_0), \tag{13}$$

where the ψ_i are a set of appropriate, complex, coefficient functions, forming a "vector" in the space of D^2-1 dimensions. For example, for $D=3$, the λ_i may be represented by the eight, traceless, Gell-Mann matrices.[6] Writing out all matrix indices, (12) takes the form

$$\frac{\partial W_{ac}}{\partial s}=-\left[\delta_{ab}\phi\left(\mathbf{x}-\nu\int_0^s ds'\,\mathbf{u}(s'),x_0+s;s\right)\right.$$
$$\left.-\sum_i(\lambda_i)_{ab}\psi_i\left(\mathbf{x}-\nu\int_0^s ds'\,\mathbf{u}(s'),x_0+s\right)\right]W_{bc},$$

and, with the boundary condition $W_{ab}[s=0]=\delta_{ab}$, has the solution

$$W(s)=\exp\left(-\int_0^s ds'\,\phi\left(\mathbf{x}-\nu\int_0^{s'}ds''\mathbf{u}(s''),x_0+s';s'\right)\right)$$
$$\cdot\left(\exp\left[\sum_i\lambda_i\int_0^s ds'\,\psi_i\left(\mathbf{x}-\nu\int_0^{s'}ds''\,\mathbf{u}(s''),x_0+s'\right)\right]\right)_+. \tag{14}$$

If the vortex stretching term was missing from the original Eq. (2), $(\boldsymbol{\omega}\cdot\nabla)\mathbf{v}\to 0$, this would be equivalent to $\psi_i=0$, and $G[v]$ would be diagonal in SU(D) space. This is precisely the case for $D=2$, where the nonsinglet terms can always be "gauged away," and possibly for those situations in three dimensions where the vorticity source geometry may be sufficiently symmetric to enforce this "singlet" property. In general, however, the specification of vorticity in terms of a given velocity field is a nonabelian problem of the same order of difficulty as that of calculating a particle propagator in the presence of a specified field containing isotopic or color degrees of freedom. There is no known way of finding an exact solution to either problem.

For the moment, we designate the s' ordered bracket of (14) by the matrix symbol $\overline{U}[\nu\mathbf{u}]$, and postpone a discussion of its properties until the three-dimensional analysis of Sec. IV. All the effects of vortex stretching are contained in $\overline{U}$. For the two-dimensional examples of Sec. III, one may set $\overline{U}=1$, and study the properties of a singlet $G[v]$,

$$[\partial_t+(\mathbf{v}\cdot\nabla)-\nu\nabla^2]G[v]=1. \tag{5'}$$

Grouping together the results of Eqs. (9)–(11) and (14), we have the representation

$$G[v]=\int_0^\infty ds\cdot\exp\left[\frac{1}{\nu}\int_0^s ds'\,\frac{\delta^2}{\delta\mathbf{u}^2(s')}\right]$$
$$\cdot\exp\left[-\int_0^s ds'\,\phi\left(\mathbf{x}+\nu\int_{s'}^s ds''\,\mathbf{u}(s''),x_0-s+s';s'\right)\right]$$
$$\cdot\overline{U}[\nu\mathbf{u}]\cdot\exp\left\{-\int_0^s ds'\,[\partial_t-\nu\mathbf{u}(s')\cdot\nabla]\right\}\Big|_{\mathbf{u}=0},$$

or

$$G(\mathbf{x},\mathbf{y};x_0,y_0|v)$$
$$=\langle\mathbf{x},x_0|G[v]|\mathbf{y},y_0\rangle$$
$$=\int_0^\infty ds\cdot\exp\frac{1}{\nu}\int_0^s ds'\,\frac{\delta^2}{\delta\mathbf{u}^2(s')}$$
$$\cdot\exp\left[-\int_0^s ds'\,\phi\left(\mathbf{x}+\nu\int_{s'}^s ds''\,\mathbf{u}(s''),x_0-s+s';s'\right)\right]$$
$$\cdot\overline{U}[\nu\mathbf{u}]\cdot\delta^D\left(\mathbf{x}-\mathbf{y}+\nu\int_0^s ds'\,\mathbf{u}(s')\right)$$
$$\cdot\delta(x_0-y_0-s)|_{\mathbf{u}=0}, \tag{15}$$

where the $(\mathbf{x},x_0)$ arguments of $\overline{U}$ may be replaced by $(\mathbf{y},y_0)$. From the temporal δ function of (15), and the positive range of integration of the variable s, it is clear that this is a retarded Green's function, nonzero only for $x_0\geqslant y_0$.

Equation (15) is an example of the Fradkin representation, given in terms of the action of a functional differential operator,

$$\exp\left[\frac{1}{\nu}\int_0^s ds'\,\frac{\delta^2}{\delta^2\mathbf{u}(s')}\right],$$

à la Schwinger. For some purposes, however, it is more convenient to recast (15) into the form of functional integration, and effectively into a path integral, à la Feynman. Imagine the continuum range of integration broken up into a summation over a set of discrete points s_i, and for each s_i replace

$$\exp\left[\frac{1}{\nu}\frac{\delta^2}{\delta\mathbf{u}^2(s_i)}\right]$$

by its Gaussian equivalent. Then take the limit of arbitrarily dense s_i, to define the functional integral replacement

$$\exp\left[\frac{1}{\nu}\int_0^s ds'\,\frac{\delta^2}{\delta\mathbf{u}^2(s')}\right]=N(s)\cdot\int d[\chi(s)]\cdot\exp\left[-\frac{\nu}{4}\int_0^s ds'\,\chi(s')^2\right.$$
$$\left.+\int_0^s ds'\,\chi(s')\cdot\frac{\delta}{\delta\mathbf{u}(s')}\right],$$

where

$$N^{-1}(s)=\int d[\chi(s)]\,\exp\left[-\frac{\nu}{4}\int_0^s ds'\,\chi^2(s')\right]. \tag{16}$$

With (16), any operation of form

$$\exp\left[\frac{1}{\nu}\int_0^s ds'\,\frac{\delta^2}{\delta\mathbf{u}^2(s')}\right]\cdot F\{\mathbf{u}\}|_{\mathbf{u}=0}$$

becomes $N(s)\cdot\int d[\chi]\,\exp[-(\nu/4)\int_0^s ds'\,\chi^2(s')]\cdot F\{\chi\}$, and

(15) may be rewritten as

$$G(\mathbf{x},\mathbf{y};x_0,y_0|\nu)=\theta(x_0-y_0)\cdot N(s)\cdot\int d[\chi]\,\delta^D\Big(\mathbf{x}-\mathbf{y}+\nu\int_0^s ds'\mathbf{u}(s')\Big)\cdot\overline{U}[\nu\chi]$$
$$\cdot\exp\Big[-\frac{\nu}{4}\int_0^s ds'\,\chi^2(s')-\int_0^s ds'\,\phi\Big(\mathbf{x}+\nu\int_{s'}^s ds''\,\chi(s''),x_0-s+s';s'\Big)\Big]\Big|_{s=x_0-y_0},$$

where ϕ is the same function as that of (10), with $\chi(s')$ replacing $\mathbf{u}(s')$. A final rescaling, $\chi(s')=\xi(s')/\nu$, puts this into the more useful, and indeed more compelling, form

$$G(\mathbf{x},\mathbf{y};x_0,y_0|\nu)=\theta(x_0-y_0)\cdot N'\int d[\xi]\cdot\delta^D\Big(\mathbf{x}-\mathbf{y}+\int_0^s ds'\xi(s')\Big)\cdot\overline{U}[\xi]$$
$$\cdot\exp\Big[-\frac{1}{4\nu}\int_0^s ds'\Big(\xi(s')+\mathbf{v}\Big(\mathbf{y}-\int_0^{s'} ds''\xi(s''),y_0+s'\Big)\Big)^2\Big]\Big|_{s=x_0-y_0},\tag{17}$$

where

$$[N']^{-1}=\int d[\xi]\exp\Big[-\frac{1}{4\nu}\int_0^s ds'\,\xi(s')^2\Big],$$

and $\theta(x)$ denotes the unit positive step function, $\theta(x)=+1, x>0, \theta(x)=0, x<0$. The $\overline{U}$ of (17) is now given by its form in (14), with the $(\mathbf{x},x_0)$ arguments of ψ_i replaced by $(\mathbf{y},y_0)$. So far, this is an exact resolution of Eq. (5).

Even a cursory glance at (17) leaves no doubt about the next, appropriate step to be taken in the limit of small viscosity. As $\nu\to 0$, the only appreciable contribution to the functional integral will come from vectors $\xi(s')$ chosen to minimize the effective action

$$S[\xi]=\frac{1}{4\nu}\int_0^s ds'\Big[\xi(s')+\mathbf{v}\Big(\mathbf{y}-\int_0^{s'} ds''\,\xi(s''),y_0+s'\Big)\Big]^2,$$

which vectors must then satisfy the Euler condition $\delta S/\delta\xi=0$ at some $\xi^{(0)}(s_1)$, $s\geqslant s_1\geqslant 0$, given by

$$0=\int_0^s ds'\Big[\delta_{ab}\delta(s'-s_1)-\partial_a v_b\Big(\mathbf{y}-\int_0^{s'} ds''\,\xi(s''),y_0+s'\Big)\cdot\theta(s'-s_1)\Big]\cdot\Big[\xi_b(s')+v_b\Big(\mathbf{y}-\int_0^{s'} ds''\xi(s''),y_0+s_1\Big)\Big].$$

The appropriate solution to this equation is, clearly, that given by the vector $\xi^{(0)}(s')$ satisfying

$$\xi_b^{(0)}(s')+v_b\Big(\mathbf{y}-\int_0^{s'} ds''\,\xi^{(0)}(s''),y_0+s'\Big)=0,\tag{18}$$

since it is only the solution to (18) which can provide a nonvanishing contribution to the functional integral of (17), as $\nu\to 0$. Because of the $(\mathbf{y},y_0)$ dependence of the $\mathbf{v}$ of (18), $\xi^{(0)}$ is then an implicit function of these variables.

In the limit of small ν, we approximately evaluate (17) by expanding $\xi(s')$ about $\xi^{(0)}(s';\mathbf{y},y_0)$, and retaining only quadratic $(\xi-\xi^{(0)})$ dependence in S. The resulting functional integral is then Gaussian, and can be evaluated without difficulty; taking into account the normalization factor N', and dropping the superscript of $\xi^{(0)}$, one obtains

$$G(\mathbf{x},\mathbf{y};x_0,y_0|\nu)|_{\nu\to 0}$$
$$=\theta(x_0-y_0)\cdot\exp(-(1/2)\,\mathrm{Tr}\ln[1+Q])\cdot\overline{U}[\xi]$$
$$\cdot\delta^D\Big(\mathbf{x}-\mathbf{y}+\int_0^s ds'\,\xi(s';\mathbf{y},y_0)\Big)\Big|_{s=x_0-y_0},\tag{19}$$

where the determinantal factor is defined by

$$\mathrm{Tr}\ln[1+Q]=\int_0^1 d\lambda\int_0^s ds_1\sum_a\langle s_1,a|Q$$
$$\cdot[1+\lambda Q]^{-1}|s_1,a\rangle$$
$$=\int_0^1 d\lambda\int_0^s ds_1\int_0^s ds_2\sum_{a,b}\langle s_1,a|Q|s_2,b\rangle$$
$$\cdot\langle s_2,b|[1+\lambda Q]^{-1}|s_1,a\rangle,\tag{20}$$

with

$$\langle s_1,a|Q|s_2,b\rangle$$
$$=-\theta(s_1-s_2)\,\partial_b v_a\Big(\mathbf{y}-\int_0^{s_1} ds'\,\xi(s';\mathbf{y},y_0),y_0+s_1\Big)$$
$$-\theta(s_2-s_1)\,\partial_a v_b\Big(\mathbf{y}-\int_0^{s_2} ds'\,\xi(s';\mathbf{y},y_0),y_0+s_2\Big)+\int_0^s ds'$$
$$\cdot\theta(s'-s_1)\theta(s'-s_2)\quad\partial_a v_c\Big(\mathbf{y}-\int_0^{s'} ds''\,\xi(s'';\mathbf{y},y_0),y_0+s'\Big)$$
$$\cdot\partial_b v_c\Big(\mathbf{y}-\int_0^{s'} ds''\,\xi(s'';\mathbf{y},y_0),y_0+s'\Big).$$

Corrections to (19) and (20), expressed in ascending powers of ν, can be generated in the standard way.

One very great simplification of these equations, the replacement of the determinantal factor of (19) by unity, can be seen from the following argument. Defining an operator R by its matrix elements

$$\langle a,s_1|R|b,s_2\rangle$$
$$=-\theta(s_1-s_2)\,\partial_b v_a\Big(\mathbf{y}-\int_0^{s_1} ds'\,\xi(s';\mathbf{y},y_0),y_0+s_1\Big),$$

it is then possible to replace the operator $1+Q$ by $(1+R^T)(1+R)=(1+R)^T(1+R)$, where the superscript T denotes "transposed" in both spatial ($\mathbf{y}$) and s variables. It follows that $\mathrm{Tr}\ln(1+Q)=\mathrm{Tr}\ln(1+R)^T+\mathrm{Tr}\ln(1+R)=2\,\mathrm{Tr}\ln(1+R)$, and one may now examine the latter, simpler quantity,

$$2\sum_{l=1}^{\infty}\frac{(-1)^{l+1}}{l}\,\mathrm{Tr}[R^l].$$

But because R is a "retarded" operator [all matrix elements $\langle as_\alpha | R | bs_\beta \rangle$ proportional to factors $\theta(s_\alpha - s_\beta)$], the trace operation vanishes for $l \geqslant 2$; and because $\nabla\cdot\mathbf{v} = 0$, the trace operation also vanishes for $l = 1$. Hence this term vanishes, and $\exp[-\frac{1}{2}\operatorname{Tr}\ln(1+Q)]$ may be replaced by unity. However, the properties of the matrix elements of Q, or of R, are of considerable importance when one asks for small-ν corrections to (19).

The special properties of the extremum vector $\boldsymbol{\xi}(s_1;\mathbf{y},y_0)$ are central to the solutions that follow. If, in terms of a specified $\mathbf{v}(\mathbf{x},x_0)$, the velocity vector $\boldsymbol{\xi}$ is to satisfy (18), it is easy to see that

$$\frac{d\boldsymbol{\xi}}{ds'} = -\left[\frac{\partial \mathbf{v}}{\partial y_0} + (\mathbf{v}\cdot\nabla_y)\mathbf{v}\right], \tag{21}$$

$\mathbf{v} = \mathbf{v}(\mathbf{y} - \int_0^s ds'\boldsymbol{\xi}, y_0 + s')$, so that variation of the parameter s' corresponds to the full, nonlinear, "hydrodynamic" variation of $\mathbf{v}$ with respect to space and time. The rhs of (21) is not zero, as it would be for the motion of a simple shock,[7] but is given by whatever is forced upon $\boldsymbol{\xi}$ by the form of the specified velocity field $\mathbf{v}(x,t)$, *via* (18). We interpret $\boldsymbol{\xi}(s';\mathbf{y},y_0)$ as the velocity vector of a fictitious particle, or passive marker, advected by $\mathbf{v}$ according to (18). If $\mathbf{v}$ is sufficiently smooth, $\boldsymbol{\xi}$ can always be developed in a power series in s',

$$\boldsymbol{\xi}(s';\mathbf{y},y_0) \simeq -\mathbf{v}(\mathbf{y},y_0) + s'\cdot\left[\frac{\partial \mathbf{v}}{\partial y_0} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right]_{s'=0} + \cdots;$$

but, because $s = x_0 - y_0$ can be arbitrarily large, this will not in general be a practical way to construct $\boldsymbol{\xi}$.

It will turn out to be useful to integrate (18) once and consider the position vector of this passive marker. Setting $\xi_\alpha(s';\mathbf{y},y_0) \equiv (dx_\alpha/ds')(s';\mathbf{y},y_0)$ and $\int_0^{s_1} ds'\, \boldsymbol{\xi}(s';\mathbf{y},y_0)$ $= \mathbf{x}(s_1;\mathbf{y},y_0) - \mathbf{x}(0;\mathbf{y},y_0) \equiv \Delta\mathbf{x}(s_1;\mathbf{y},y_0)$, we define the quantity $\mathbf{q}(s_1;\mathbf{y},y_0) \equiv \mathbf{y} - \Delta\mathbf{x}(s_1;\mathbf{y},y_0)$ as the marker's position vector, thereby replacing (18) by

$$\mathbf{q}(s_1;\mathbf{y},y_0) = \mathbf{y} + \int_0^{s_1} ds'\, \mathbf{v}(\mathbf{q}(s';\mathbf{y},y_0),y_0 + s'). \tag{22}$$

For a problem dealing with discrete, pointlike vortices, the $\mathbf{q}$ vectors turn out to represent the time-dependent position vectors of the vortices themselves.

With (3), (19), and (22), we are now in a position to calculate vorticity and its corresponding velocity in the $\nu \to 0$ limit:

$$\omega_\alpha(\mathbf{x},x_0) = \int d^D y \int dy_0\, G_{\gamma\delta}(\mathbf{x},\mathbf{y};x_0,y_0|\nu) g_\delta(\mathbf{y},y_0) \tag{23a}$$

and

$$\mathbf{v}(\mathbf{r},t) = \mathbf{v}_0 + \int d^D x\, G_0(\mathbf{r} - \mathbf{x})\nabla\times\boldsymbol{\omega}(\mathbf{x},t), \tag{23b}$$

where $G_0(Q)$ denotes the appropriate, D-dimensional Green's function, $G_0 = -\nabla^{-2}$. For simplicity and clarity, we henceforth use the three-dimensional form, $G_0(Q)$ $= [4\pi|\mathbf{Q}|]^{-1}$, except when reference is made to $D = 2$. We shall assume the vorticity is of finite spatial extent, and freely drop surface terms, replacing (23b), for example, by

$$\mathbf{v}(\mathbf{r},t) = \mathbf{v}_0 + \frac{1}{4\pi}\int d^3x \left[\nabla_r\cdot \frac{1}{|\mathbf{r} - \mathbf{x}|}\right]\times\boldsymbol{\omega}(\mathbf{x},t). \tag{23c}$$

Note that there is a zero contribution to the integrand of (23c) when x sweeps over the point $\mathbf{r}$, for, while apparently singular, that contribution vanishes by symmetry, as in the relation $\int d^3p\, \mathbf{p}(\mathbf{p}^2)^{-2} \exp[i\mathbf{p}\cdot(\mathbf{r}-\mathbf{x})]\,|_{|\mathbf{x}-\mathbf{r}|\to 0} \Rightarrow 0$.

Combining (23) and (18), we obtain

$$v_\alpha(\mathbf{r},t) = v_\alpha^{(0)} + \frac{\epsilon_{\alpha\beta\gamma}}{4\pi}\int d^3y \int_{-\infty}^{t} dy_0\, g_\delta(\mathbf{y},y_0) \times \frac{[\mathbf{r} - \mathbf{q}(s;\mathbf{y},y_0)]_\beta}{|\mathbf{r} - \mathbf{q}(s;\mathbf{y},y_0)|^3}\cdot \overline{U}_{\gamma\delta}[\mathbf{q}(s_1;\mathbf{y},y_0)], \tag{24}$$

where $s = t - y_0$ and $\overline{U}_{\gamma\delta}[\mathbf{q}(s_1;\mathbf{y},y_0)]$ denotes $(\exp[\Sigma_i \lambda_i \int_0^s ds'\, \psi_i(\mathbf{q}(s';\mathbf{y},y_0),y_0 + s')])_+$. Equations (22) and (24) represent our resolution of the inviscid NS problem.

With the aid of (24), one can rewrite (22) in the form

$$q_\alpha(s_1;\mathbf{y},y_0) = y_\alpha + v_\alpha^{(0)} s_1 + \frac{\epsilon_{\alpha\beta\gamma}}{4\pi}\int d^3y' \int_0^{s_1} ds' \int_{-\infty}^{y_0 + s'} dy_0' \cdot \overline{U}_{\gamma\delta}\cdot g_\delta(\mathbf{y}',y_0') \cdot\frac{[q_\beta(s';\mathbf{y},y_0) - q_\beta(s';\mathbf{y}',y_0')]}{|\mathbf{q}(s';\mathbf{y},y_0) - \mathbf{q}(s';\mathbf{y}',y_0')|^3}, \tag{25}$$

where the $\overline{U}$ of (25) denotes

$$(\exp[\Sigma_i \lambda_i \textstyle\int_0^{s'} ds''\, \psi_i(\mathbf{q}(s'';\mathbf{y}',y_0'),y_0' + s'')])_+.$$

It is perhaps not an easy equation to solve exactly, but there are, perhaps, reasonably simple vorticity source distributions where exact or reasonably simple approximations to (25) may be devised. Once one has solved, or approximated (25), knowledge of $\mathbf{q}(s;\mathbf{y},y_0)$ can then be used to construct the desired $\mathbf{v}(\mathbf{r},t)$. Equation (25) provides, in essence, a determining equation for the Cauchy Lagrangian-position coordinates, generalized to arbitrary source distributions $\mathbf{g}(\mathbf{y},y_0)$.

III. EXAMPLES

For simplicity we henceforth consider a "one-shot" vorticity source, turning on and off rapidly at $t \simeq 0$, and modeled by $\mathbf{g}(\mathbf{y},y_0) = \mathbf{g}(y)\cdot\delta(y_0)$; that is, at $t = 0$ a specified vorticity distribution is inserted into the fluid, and we watch the resulting fluid flow develop in time. Generalizations correspond to a continuous input of vorticity are easily written down for every example, but will not be considered in this paper. The examples of this section will deal with two-dimensional flows, for which we set $\overline{U} = 1$ and use the $D = 2$ form of $G_0 = -\nabla^{-2}$. In this way (24) and (25) become

$$\mathbf{v}(\mathbf{r},t) = \mathbf{v}_0 + \frac{\theta(t)}{2\pi}\int d^2y\, \mathbf{g}(\mathbf{y})\times\frac{[\mathbf{r} - \mathbf{q}(t;\mathbf{y})]}{|\mathbf{r} - \mathbf{q}(t;\mathbf{y})|^2} \tag{26}$$

and

$$\mathbf{q}(s_1;\mathbf{y}) = \mathbf{y} + \mathbf{v}_0 s_1 + \frac{1}{2\pi}\int_0^{s_1} ds' \int d^2y'\, \mathbf{g}(\mathbf{y}') \times\frac{[\mathbf{q}(s';\mathbf{y}) - \mathbf{q}(s';\mathbf{y}')]}{|\mathbf{q}(s';\mathbf{y}) - \mathbf{q}(s';\mathbf{y}')|^2}, \tag{27}$$

where the $y_0 = 0$ coordinate of q has been suppressed. The radius vector is here given by $\mathbf{r} = \hat{\imath}x_1 + \hat{\jmath}x_2$, with the source $\mathbf{g}(y)$ pointing in the $\hat{k}$ direction; the vector $\mathbf{y}$ is understood to lie in the (x_1,x_2) plane. We discuss the solution of this pair of equations for the following three situations.

(A) One very thin vortex tube, of radius much less than any other length dimension, modeled by

$$\mathbf{g}(\mathbf{y}) = \mathbf{g}\delta(\mathbf{y}) = \mathbf{g}\delta(y)/2\pi y.$$

Here, both the y and y' variables of (26) and (27) are to be set equal to zero. Remembering the comment following (23c), which is also true in two dimensions, one has, directly, $\mathbf{q}(s_1;\mathbf{0}) = \mathbf{v}_0 s_1$ as the solution of (27). There then follows

$$\mathbf{v}(\mathbf{r},t) = \mathbf{v}_0 + \frac{\theta(t)}{2\pi}\mathbf{g}\times\frac{[\mathbf{r}-\mathbf{v}_0 t]}{|\mathbf{r}-\mathbf{v}_0 t|^2}, \tag{28a}$$

$$\boldsymbol{\omega}(\mathbf{r},t) = \theta(t)\mathbf{g}\delta(\mathbf{r}-\mathbf{v}_0 t). \tag{28b}$$

Equations (28) should be exact; that is, they should simultaneously satisfy the Euler equation $(\partial_t + (\mathbf{v}\cdot\nabla))\boldsymbol{\omega} = \mathbf{g}\delta(\mathbf{r})\delta(t)$, together with $\boldsymbol{\omega} = \nabla\times\mathbf{v}$. This is easily verified.

(B) A circular vortex sheet, modeled by a dense collection of very thin vortex tubes, all of equal strength, all pointing in the $\hat{k}$ direction, and arranged in a ring of radius r_0 in the (x_1, x_2) plane. In the limit of a continuous number of such thin tubes, we write $\mathbf{g}(\mathbf{y}) = (\mathbf{G}/r_0)\delta(y - r_0)$, and for simplicity set $\mathbf{v}_0 = 0$. In the (x_1, x_2) plane, we set $|\mathbf{y}| = |\mathbf{y}'| = r_0$, and $\mathbf{q}(s_1;\mathbf{y}) \rightarrow \mathbf{q}(s_1;\theta) = \hat{i}q_1(s_1;\theta) + \hat{j}q_2(s_1;\theta)$. Substituting these into the two-dimensional form of (27) yields the pair of equations

$$q_1(s;\theta) = r_0\cos\theta - G\int_0^s ds'\int_0^{2\pi}\frac{d\theta'}{2\pi} \times\frac{[q_2(s';\theta) - q_2(s';\theta')]}{|\mathbf{q}(s';\theta) - \mathbf{q}(s';\theta')|^2}, \tag{29a}$$

$$q_2(s;\theta) = r_0\sin\theta + G\int_0^s ds'\int_0^{2\pi}\frac{d\theta'}{2\pi} \times\frac{[q_1(s';\theta) - q_1(s';\theta')]}{|\mathbf{q}(s';\theta) - \mathbf{q}(s';\theta')|^2}, \tag{29b}$$

where $\mathbf{y} = \hat{i}r_0\cos\theta + \hat{j}r_0\sin\theta$.

Equations (29) can be solved with the aid of the ansatz

$$q_1 = \phi_a(s_1)\cos\theta - \phi_b(s)\sin\theta,$$
$$q_2 = \phi_a(s)\sin\theta + \phi_b(s)\cos\theta, \tag{30}$$

which, when substituted into (29), leads to the pair of equations

$$\phi_a(s) = r_0 - \frac{G}{2}\int_0^s ds'\frac{\phi_b(s')}{[\phi_a^2(s') + \phi_b^2(s')]}, \tag{31a}$$

$$\phi_b(s) = +\frac{G}{2}\int_0^s ds'\frac{\phi_a(s')}{[\phi_a^2(s') + \phi_b^2(s')]}. \tag{31b}$$

Equations (31) may be solved most simply by rewriting them as differential equations parametrized in the form

$$\phi_a(s) = \rho(s)\cos\psi(s),\ \ \phi_b(s) = \rho(s)\sin\psi(s).$$

One immediately finds that $\rho(s_1) = \text{const} = r_0$, while $\psi(s) = (G/2r_0^2)s$. Regrouping and substituting into (30), one has the solutions

$$q_1 = r_0\cos[\theta + (G/2r_0^2)s],\ \ q_2 = r_0\sin[\theta + (G/2r_0^2)s]. \tag{32}$$

Inserting (32) into (26), one finds

$$\mathbf{v}(\mathbf{r},t) = \frac{\theta(t)}{2\pi}\int_0^{2\pi} d\theta\,\mathbf{G}\times\frac{[\mathbf{r}-\mathbf{q}(t;\theta)]}{|\mathbf{r}-\mathbf{q}(t;\theta)|^2}. \tag{33}$$

The quadrature of (33) can be performed, and yields

$$\mathbf{v}(\mathbf{r},t) = \mathbf{G}\times\hat{r}\phi(r),\quad \boldsymbol{\omega}(\mathbf{r},t) = \theta(t)(\mathbf{G}/r_0)\delta(r-r_0), \tag{34}$$

where $\phi(r) = r^{-1}, r > r_0$, $\phi(r) = (2r_0)^{-1}, r = r_0$, and $\phi(r) = 0$, $r < r_0$. Again, (33) represents an exact solution to the inviscid Euler equation. It will be noted that the special angular symmetry of the vorticity source distribution effectively removes all the time dependence of the q_i of (32), resulting in the relatively simple form of (34), which indeed could have more easily been guessed from an inspection of the original Euler equation. However, in other situations where such symmetry is lacking, our method may turn out to be useful.

(C) Two thin vortices of arbitrary strength generate a soluble problem without the overwhelming symmetry of the previous example. We take $\mathbf{g}(\mathbf{y}) = \mathbf{g}_1\delta(\mathbf{y}-\mathbf{y}_1) + \mathbf{g}_2\delta(\mathbf{y}-\mathbf{y}_2)$, and write the pair of equations corresponding to (27),

$$\mathbf{q}_1(s) = \mathbf{y}_1 + \mathbf{v}_0 s + \frac{1}{2\pi}\int_0^s ds'\,\mathbf{g}_2\times\frac{[\mathbf{q}_1(s') - \mathbf{q}_2(s')]}{|\mathbf{q}_1(s') - \mathbf{q}_2(s')|^2}, \tag{35a}$$

$$\mathbf{q}_2(s) = \mathbf{y}_2 + \mathbf{v}_0 s + \frac{1}{2\pi}\int_0^s ds'\,\mathbf{g}_1\times\frac{[\mathbf{q}_2(s') - \mathbf{q}_1(s')]}{|\mathbf{q}_2(s') - \mathbf{q}_1(s')|^2}. \tag{35b}$$

Writing $\mathbf{q}(s) = \mathbf{q}_1(s) - \mathbf{q}_2(s)$, $\mathbf{y} = \mathbf{y}_1 - \mathbf{y}_2$, $\mathbf{G} = \mathbf{g}_1 + \mathbf{g}_2$, the difference of Eqs. (35) becomes

$$\mathbf{q}(s) = \mathbf{y} + \mathbf{G}\times\int_0^s ds'\frac{\mathbf{q}(s')}{|\mathbf{q}(s')|^2}, \tag{36}$$

which has a strong resemblance to the equations of example (B). Since $\mathbf{y}$ is a vector in the (x_1, x_2) plane, perpendicular to $\mathbf{G}$, we can use $\hat{y}$, $\hat{G}$ and $\hat{y}\times\hat{G}$ as three orthogonal directions, writing $\mathbf{q}(s) = \hat{y}F_1(s) + (\hat{y}\times\hat{G})F_2(s)$. Substitution into (36) then generates the pair of equations

$$F_1(s) = y - G\int_0^s ds'\frac{F_2(s')}{[F_1^2(s') + F_2^2(s')]}, \tag{37a}$$

$$F_2(s) = +G\int_0^s ds'\frac{F_1(s')}{[F_1^2(s') + F_2^2(s')]}, \tag{37b}$$

with solution [by comparison with Eqs. (31)]

$$F_1(s) = y\cos(Gs/y^2),\ \ F_2(s) = y\sin(Gs/y_2).$$

Now that $\mathbf{q}(s)$ is known, one returns to the original Eqs. (35) to construct, by simple quadrature, the individual $\mathbf{q}_1(s)$, $\mathbf{q}_2(s)$; for example,

$$\mathbf{q}_1(s) = \mathbf{y}_1 + \mathbf{v}_0 s + \frac{y}{G}\mathbf{g}_2 \times\left\{\hat{y}\sin\left(\frac{Gs}{y^2}\right) + (\hat{y}x\hat{G})\left[1 - \cos\left(\frac{Gs}{y^2}\right)\right]\right\},$$

and similarly for $\mathbf{q}_2(s)$. These $\mathbf{q}_i(t)$ are then inserted into the expression corresponding to (26) to obtain

$$\mathbf{v}(\mathbf{r},t) = \mathbf{v}_0 + \frac{\theta(t)}{2\pi}\sum_{l=1}^{2}\mathbf{g}_l\times\frac{[\mathbf{r}-\mathbf{q}_l(t)]}{|\mathbf{r}-\mathbf{q}_l(t)|^2}. \tag{38}$$

From (38) one calculates

$$\boldsymbol{\omega}(\mathbf{r},t) = \theta(t)\sum_{l=1}^{2}\mathbf{g}_l\delta(\mathbf{r}-\mathbf{q}_l(t)), \tag{39}$$

showing that these vortices, inserted into the fluid at positions $\mathbf{y}_l$ at $t = 0$, subsequently move relative to the fixed coordinate system, or to each other, as specified by the time dependence of the $\mathbf{q}_l(t)$. Again, in this problem of very thin vortices, the "marker" coordinates have become those of the vortices; and, again, it is easy to show that (38) and (39) form an exact solution to the inviscid Euler equation. Of course, these solutions are well known; their usual differential equation, given in terms of complex position coordinates,[8] can be read off from the time derivative of Eqs. (35).

Generalizations to the case of N point vortices may be handled in the same manner. The complexity of the analysis, of course, increases rapidly with N, while for $N = 4$ there is evidence for a chaotic behavior of the solutions.[1]

IV. VORTEX STRETCHING

In this section we discuss some properties of the nonunitary matrix $\overline{U}$ relevant in three dimensions. For simplicity, the analysis continues to assume that $g(y,y_0) = g(y)\cdot\delta(y_0)$, using the source to insert an arbitrary vortex distribution into the fluid at $t = 0$. We suppress the $y_0 = 0$ coordinate of $\mathbf{q}(s;\mathbf{y},y_0) = \mathbf{q}(s;\mathbf{y})$, and first inspect the $\mathbf{y}$ dependence of $\mathbf{q}$, as expressed by (22).

Under the variation $\mathbf{y}\to\mathbf{y} + \delta\mathbf{y}$,

$$\mathbf{q}(s;\mathbf{y} + \delta\mathbf{y}) = \mathbf{y} + \delta\mathbf{y} + \int_0^s ds'\ \mathbf{v}(\mathbf{q}(s';\mathbf{y} + \delta\mathbf{y}),s'),$$

or to first order in $\delta\mathbf{y}$,

$$(\delta\mathbf{y}\cdot\nabla)q_c(s;\mathbf{y}) = \delta y_c + \int_0^s ds'\ [(\delta\mathbf{y}\cdot\nabla)q_a(s';\mathbf{y})]\cdot[\partial_a v_c(\mathbf{q}(s';\mathbf{y}),s')].$$

If $\delta\mathbf{y}$ has a nonzero component only in the b direction, this becomes

$$\partial_b q_c(s;\mathbf{y}) = \delta_{cb} + \int_0^s ds'\ \partial_b q_a(s';\mathbf{y})\cdot\partial_a v_c(\mathbf{q}(s';\mathbf{y}),s'). \quad (40)$$

But (40) is just an expression of the integral equation whose solution is the $\overline{U}$ of (25), since the quantity $Q_{ca}(s)\equiv\partial_a v_c(\mathbf{q}(s;\mathbf{y}),s) = \Sigma_i(\lambda_i)_{ca}\psi_i(\mathbf{q}(s;\mathbf{y}),s)$ is precisely the interaction term of the differential equation built from (40),

$$\frac{\partial\overline{U}_{cb}}{\partial s} = Q_{ca}(s)\cdot\overline{U}_{ab}(s), \quad (41)$$

where we have written $\overline{U}_{cb}(s)\equiv\partial_b q_c(s;\mathbf{y})$, in anticipation of this result. Thus, the exact solution to (41) is

$$\overline{U}(s) = \left(\exp\left[\int_0^s ds'\ Q(s')\right]\right)_+, \quad (42)$$

which is precisely the $(y_0 = 0)$ quantity $\overline{U}$ of (25).

Knowing that the vortex stretching term $\overline{U}_{cb}$ is nothing other than the spatial gradient of the vortex source, or marker, coordinate $\partial_b q_c(s;\mathbf{y})$ is interesting[3] and may even turn out to be of some practical use in finding an approximate solution to (25). For the qualitative purpose of this paper, however, we will examine the behavior of $\overline{U}$ in terms of its SU(3) coordinates.[9] The first task is to define the ψ_i, which follow from the definition of Q and the properties of the Hermitian λ_i.

For convenience, these properties—taken from Ref. 4—are grouped together in the Appendix, and will be used as needed.

From the definition, $\partial_b v_a = (\lambda_i)_{ab}\psi_i$, application of the trace property $\mathrm{Tr}[\lambda_i\lambda_j] = 2\delta_{ij}$ leads to $\psi_i = \frac{1}{2}\Sigma_{a,b}(\lambda_i)_{ba}\partial_b v_a$. From the explicit forms of the Appendix, one then constructs the components

$$\begin{aligned}
&\psi_1 = \tfrac{1}{2}[\partial_1 v_2 + \partial_2 v_1], \quad \psi_2 = \tfrac{1}{2}i[-\partial_1 v_2 + \partial_2 v_1],\\
&\psi_3 = \tfrac{1}{2}[\partial_1 v_1 + \partial_2 v_2], \quad \psi_4 = \tfrac{1}{2}[\partial_1 v_3 + \partial_3 v_1],\\
&\psi_5 = \tfrac{1}{2}i[-\partial_1 v_3 + \partial_3 v_1], \quad \psi_6 = \tfrac{1}{2}[\partial_2 v_3 + \partial_3 v_2], \quad (43)\\
&\psi_7 = \tfrac{1}{2}i[-\partial_2 v_3 + \partial_3 v_2],\\
&\psi_8 = (1/\sqrt{3})[\partial_1 v_1 + \partial_2 v_2 - 2\,\partial_3 v_3].
\end{aligned}$$

Since each Q_{ab} must be real, the three components ψ_2, ψ_5, ψ_7 are imaginary because the corresponding λ_2, λ_5, λ_7 have purely imaginary components. It is interesting to note that these three ψ_i are just proportional to the three components of vorticity, ω_a; while the remaining ψ_i are given by real symmetric velocity gradients. It will be useful to divide this collection of real and imaginary terms into two sets: $\lambda_i\psi_i\equiv i\lambda_a\omega_a + \lambda_\alpha\phi_\alpha$, where the λ_a run over the imaginary matrices, $a = (2, 5, 7)$, and the λ_α run over the real matrices, $\alpha = (1, 3, 4, 6, 8)$, with the ϕ_α corresponding to the real, symmetric, velocity gradient components of ψ. It should be emphasized that all of these λ_a,λ_α are Hermitian.

Equation (41) can then be rewritten in the form

$$\frac{\partial\overline{U}}{\partial s} = (i\lambda\cdot\omega + \lambda\cdot\phi)\overline{U},\ \ \overline{U}(s = 0) = 1, \quad (44)$$

and the nonunitary nature of the vortex stretching becomes clear upon writing the Hermitian conjugate of (44), and adding the two equations to obtain

$$\frac{\partial}{\partial s}(\overline{U}^\dagger\overline{U}) = 2\overline{U}^\dagger\lambda\cdot\phi\overline{U}, \quad (45)$$

showing that it is the symmetric components ϕ_α which govern the increase of $\overline{U}^\dagger\overline{U}$. The s dependence of ϕ, as of ω, is contained in the space-time $(\mathbf{q}(s;\mathbf{y}),s)$ arguments of these functions.

In order to construct solutions to (44), it is useful to extract the manifestly Hermitian part of $\overline{U}$ by setting $\overline{U}(s) = \overline{V}(s)\cdot\overline{W}(s)$, with $\overline{V}(s) = (\exp[i\int_0^s ds'\ \lambda\cdot\omega(s')])_+$, where the ordering symbol refers to the s' variable, as in (8). The matrix $\overline{V}(s)$ is then unitary, $\overline{V}^\dagger = \overline{V}^{-1}$, and all the nonunitary behavior of $\overline{U}$ can be transferred to $\overline{W}$, which satisfies

$$\frac{\partial\overline{W}}{\partial s} = \lambda\cdot\Phi(s)\overline{W}(s),\ \ \overline{W}(0) = 1, \quad (46)$$

where $\lambda\cdot\Phi = \overline{V}^\dagger(s)\lambda\cdot\phi(s)\overline{V}(s)$. The real components, $\Phi_i(s)$ are given by

$$\Phi_i = \tfrac{1}{2}\,\mathrm{Tr}[\lambda_i\overline{V}^\dagger\lambda\cdot\phi\overline{V}], \quad (47)$$

and it is also true that the magnitudes of Φ_i and ϕ_α are equal,

$$\Phi^2\equiv\sum_i\Phi_i^2 = \sum_\alpha\phi_\alpha^2\equiv\phi^2. \quad (48)$$

From (46), it follows that

$$\frac{\partial}{\partial s}(\overline{W}^\dagger\overline{W}) = 2\overline{W}^\dagger(\overline{V}^\dagger\lambda\cdot\phi\overline{V})\overline{W},$$

which is consistent with (45) and the unitarity of $\overline{V}$, $\overline{W}^\dagger\overline{W} = \overline{U}^\dagger\overline{U}$.

The way in which these properties shall be used is as follows. We begin with the "vectors" ϕ_α and ω_a, and assume that all velocity gradients are increasing in magnitude with time ($s \equiv t$ when $y_0 = 0$), while their "angular" variations are relatively slowly varying. That is, we take $\boldsymbol{\omega} = \hat{\omega}\omega$, $\hat{\omega}^2 \equiv 1$, and assume that

$$\left| \frac{d\hat{\omega}}{dt} \right| \ll \omega, \quad \int_0^t dt'\, \omega(t') \gg 1. \tag{49}$$

A corresponding relation for the angular behavior of the symmetric components $\boldsymbol{\phi} = \hat{\phi}\phi$ will not be necessary, although we continue to assume that the time integral of the magnitude of ϕ increases in a reasonably steady way, $\int^t dt'\, \phi(t') \gg 1$, while $\hat{\phi}$ does not change too rapidly in time. These assumptions will be justified, in a crude way, *a posteriori*.

Under these conditions we are interested in finding an approximate solution[9] for $\overline{W}(t)$, and begin by considering an explicit representation for $\overline{V}(t)$, which satisfies the differential equation

$$\frac{\partial \overline{V}}{\partial t} = i(\lambda\cdot\omega(t))\overline{V}(t), \quad \overline{V}(0) = 1. \tag{50}$$

Because $\overline{V}(t)$ is unitary, it may be written in the form $\overline{V}(t) = \exp[iF_0(t) + i\boldsymbol{\lambda}\cdot\mathbf{F}(t)]$, where $\mathbf{F}$ is, in general, a vector with eight real components. The initial condition $\overline{V}(0) = 1$ then becomes $F_0(0) = F_i(0) = 0$. Substituting this form of $\overline{V}(t)$ into (50), with the aid of the general formula

$$\frac{d}{dt} e^{\Gamma(t)} = \int_0^1 d\mu\, e^{\mu\Gamma(t)} \frac{d\Gamma}{dt} e^{(1-\mu)\Gamma(t)}, \tag{51}$$

useful when $[\Gamma, d\Gamma/dt] \neq 0$, one immediately learns that $F_0(t) \equiv 0$, as one builds a differential equation for $F(t)$,

$$\boldsymbol{\lambda}\cdot\boldsymbol{\omega}(t) = \int_0^1 d\mu\, e^{i\mu\boldsymbol{\lambda}\cdot\mathbf{F}(t)} \cdot \left(\boldsymbol{\lambda}\cdot \frac{d\mathbf{F}}{dt}\right) \cdot e^{-i\mu\boldsymbol{\lambda}\cdot\mathbf{F}(t)},$$

or

$$\omega_i(t) = \frac{1}{2} \int_0^1 d\mu\, \mathrm{Tr}\left[\lambda_i e^{i\mu\boldsymbol{\lambda}\cdot\mathbf{F}} \left(\boldsymbol{\lambda}\cdot \frac{d\mathbf{F}}{dt}\right) e^{-i\mu\boldsymbol{\lambda}\cdot\mathbf{F}}\right]. \tag{52}$$

It follows from (52), by multiplication by $\Sigma_i \hat{F}_i$, and the trace properties of the λ_j, that

$$\hat{F}\cdot\boldsymbol{\omega} = \hat{F}\cdot \frac{d\mathbf{F}}{dt} = \frac{dF}{dt},$$

where $F = \sqrt{\mathbf{F}^2} = (\Sigma_i F_i^2)^{1/2}$. Further, if $\mathbf{F}$ is written as $\hat{F}F$, (52) can be rewritten as

$$\omega_i(t) = \hat{F}_i \frac{dF}{dt} + \frac{F}{2} \int_0^1 \mu\, \mathrm{Tr}\left[\lambda_i e^{i\mu\boldsymbol{\lambda}\cdot\mathbf{F}} \left(\boldsymbol{\lambda}\cdot \frac{d\hat{F}}{dt}\right) e^{-i\mu\boldsymbol{\lambda}\cdot\mathbf{F}}\right]. \tag{53}$$

An alternate representation for (53) is obtained by expanding the exponentials of the second term, using the basic commutation relation of this Lie algebra, $[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k$, where the f_{ijk} structure constants are real, completely antisymmetric numbers as written in the Appendix. Introducing the adjoint representation, Hermitian matrices, $(A^i)_{jk} \equiv if_{ijk}$, (59) may be rewritten as

$$\omega_i(t) = \hat{F}_i \frac{dF}{dt} + F \int_0^1 d\mu\, \frac{d\hat{F}_1}{dt} (e^{2i\mu\mathbf{A}\cdot\mathbf{F}})_{ji},$$

or

$$\omega_i(t) = \hat{F}_i \frac{dF}{dt} + \frac{i}{2}\left(\frac{d\hat{F}_j}{dt}\right)[(\mathbf{A}\cdot\hat{F})^{-1}(1 - e^{2i\mathbf{A}\cdot\hat{F}F})]_{ji}. \tag{54}$$

Assuming that $(A\cdot F)^{-1}$ exists, that is, $\det(A\cdot F) \neq 0$, Eq. (54) is exact. It may be used to infer the form of $\mathbf{F}$ in the special circumstance that $F \gg 1$ and $|d\hat{F}/dt| \ll |dF/dt|$ or ω, for the second rhs term of (54) may be discarded in comparison to the first; because all the components of $\mathbf{F}$, or $\hat{F}$, are real numbers, and the $\mathbf{A}$ are Hermitian, the exponential of (54) simply oscillates rapidly as F increases. In this limit, (54) reduces to

$$\omega_i \simeq \hat{F}_i \frac{dF}{dt}, \tag{55}$$

which in comparison with (52), yields the approximate "strong-coupling" solution,

$$\hat{F}_i(t) = \hat{\omega}_i, \quad F(t) = \int_0^t dt'\, \omega(t'). \tag{56}$$

The conditions for the validity of this approximate solution,

$$\overline{V}(t) \simeq \exp\left[i\lambda\cdot\hat{\omega}(t) \int_0^t dt'\, \omega(t')\right], \tag{57}$$

are those of (49).

One can now see in a clear way, from (47) and (48), the effect of such a unitary $\overline{V}$ under the supposed conditions (49). Writing the components Φ_i as $\Phi\hat{\Phi}_i$, where $\Phi = (\Sigma_i \Phi_i^2)^{1/2}$, it follows that the magnitude $\Phi(t) = \phi(t)$ is unchanged by the $\overline{V}$ unitary transformation, while the unit vector $\hat{\Phi}(t)$ is given a rapidly oscillating time dependence, with a frequency proportional to $(1/t)\oint_0^t dt'\, \omega(t')$. That is, if one writes an alternate expression for (57), in terms of the vector $\mathbf{Q} = \hat{\omega}\int_0^t dt'\, \omega(t')$,

$$\overline{V}(t) = \exp[i\mathbf{Q}\cdot\boldsymbol{\lambda}] \equiv f_0 + i\sum_a \lambda_a f_a + \sum_\alpha \lambda_\alpha g_\alpha,$$

where $f_0 = \frac{1}{3}\mathrm{Tr}\,[\exp(i\mathbf{Q}\cdot\boldsymbol{\lambda})]$, $f_a = -\frac{1}{2}i\,\mathrm{Tr}[\lambda_a \exp(i\mathbf{Q}\cdot\boldsymbol{\lambda})]$, $g_\alpha = \frac{1}{2}\mathrm{Tr}[\lambda_\alpha \exp(i\mathbf{Q}\cdot\boldsymbol{\lambda})]$. The quantity f_0 can be calculated directly from the three eigenvalues of the matrix $\mathbf{Q}\cdot\boldsymbol{\lambda}$, which satisfy the relations $\Sigma_{n=1}^3 \xi_n = 0$, $\Sigma_{n=1}^3 \xi_n^2 = 2\mathbf{Q}^2$, and $\Sigma_{n=1}^3 \xi_n^3 = -3D \equiv 2\Sigma_{ijk} d_{ijk} Q_i Q_j Q_k$. For the above choice of the Q_a, it can be seen that $D = 0$, following from the explicit d_{abc} of the Appendix, so that $\xi_1 = -\xi_3 = +Q \equiv j\sqrt{\mathbf{Q}^2}$, $\xi_2 = 0$, generate f_0 and $f_a = -\frac{3}{2}(\partial/\partial Q_a)f_0$. But the functions g_α are not directly calculable in this way, although they may be inferred in an indirect way from the requirement of unitarity.

A more elegant way of calculating both f_a and g_α at the same time is to write $\exp[i\mathbf{Q}\cdot\boldsymbol{\lambda}] = f_0 + i\Sigma_{i=1}^8 \lambda_i f_i$ and imagine that $\mathbf{Q}$ has projections in all eight directions, although its extension in the α directions is very small. Then $|D| \ll Q^2$, and the cubic equation for the ξ_n is easily approximated to yield

$$\xi_1 \simeq Q - D/2Q^2, \quad \xi_2 \simeq +D/Q^2, \quad \xi_3 \simeq -Q - D^2/2Q^2,$$

so that $f_0(Q) \simeq \frac{1}{3}[1 + 2\cos Q] + \frac{1}{3}i(D/Q^2)[1 - \cos Q]$, and the $f_i = -\frac{3}{2}(\partial/\partial Q_i)f_0$ can be evaluated in the $D \to 0$ limit as $f_a = \hat{Q}_a \sin Q$, $f_\alpha = iv_\alpha(1 - \cos Q) \equiv -ig_\alpha$, where $v_\alpha \equiv \Sigma_{ab\alpha} d_{ab\alpha} \hat{Q}_a \hat{Q}_b$. Note that all the matrix elements of $\overline{V}$

are real,

$$\overline{V}(t) = \tfrac{1}{3}[1 + 2\cos Q] + i\boldsymbol{\lambda}\cdot\hat{\omega}\sin Q - \sum_{\alpha}\lambda_{\alpha}v_{\alpha}(1-\cos Q). \tag{58}$$

Using the property $v^2 \equiv \Sigma_{\alpha} v_{\alpha}^2 = \frac{1}{3}$, also found via the Appendix, it is easy to verify unitarity; e.g., $1 = f_0^2 + \frac{2}{3}\Sigma_a f_a^2 + \frac{2}{3}\Sigma_{\alpha} g_{\alpha}^2$.

The effect of such a $\overline{V}$ unitary transformation, as in (47), is to make the unit vectors $\hat{\Phi}_i$ rotate with frequency proportional to $(1/t)\int_0^t dt'\,\omega(t')$, here assumed to be a reasonably large number, while leaving the magnitude unchanged, $\Phi = \phi$. One then tries to solve (46) under these conditions, and in a more explicit and useful form than that given by the formal solution,

$$\overline{W}(t) = \left(\exp\left[\int_0^t dt'\,\boldsymbol{\lambda}\cdot\boldsymbol{\Phi}(t')\right]\right)_+.$$

If one adopts the ansatz[10] $\overline{W}(t) = \exp[G_0(t) + \boldsymbol{\lambda}\cdot\mathbf{G}(t)]$, and forms the differential equation for $(G_0,\mathbf{G})$ using (51), one immediately obtains $G_0 \equiv 0$, and in a manner analogous to (54),

$$\Phi_i = \hat{G}_i\frac{dG}{dt} + G\int_0^1 d\mu\,\frac{d\hat{G}_j}{dt}\,(e^{2\mu\mathbf{A}\cdot\hat{G}G})_{ji}. \tag{59}$$

For large G, which we also assume and which turns out to be a more stringent assumption than $\int^t dt'\,\phi(t') \gg 1$, the exponent of the second rhs term of (59) cannot be treated as oscillatory, as in the analysis of (54), for the eigenvalues of $\mathbf{A}\cdot\hat{G}$ are real numbers, of both signs. It is still true, however, without approximation that

$$\frac{dG}{dt} = \hat{G}\cdot\boldsymbol{\Phi} = \xi(t)\phi(t),$$

or

$$G(t) = \int_0^t dt'\,\xi(t')\phi(t'), \tag{60}$$

where $\xi(t) = \hat{G}(t)\cdot\hat{\Phi}(t)$. Rewriting (59) in the form

$$\hat{\Phi}_i\phi - \hat{G}_i(\hat{G}\cdot\hat{\Phi})\phi = G\int_0^1 d\mu\,\frac{d\hat{G}_j}{dt}\,(e^{2\mu\mathbf{A}\cdot\hat{G}G})_{ji}, \tag{61}$$

and, multiplying both sides of (61) by $(\mathbf{A}\cdot\hat{G})_{il}$, one forms

$$\frac{1}{2}\frac{d\hat{G}_j}{dt}\,(e^{2\mathbf{A}\cdot\hat{G}G} - 1)_{jl} = \phi\Phi_i(\mathbf{A}\cdot\hat{G})_{il}, \tag{62}$$

with the quantity proportional to $\hat{G}_i(\mathbf{A}\cdot\hat{G})_i$ vanishing by symmetry. Just as one performed the analysis expressing the approximate form of $\overline{V}$ in terms of the eigenvalues of $\boldsymbol{\lambda}\cdot\mathbf{Q}$, as in (58), so the eigenvalues of $(\mathbf{A}\cdot\hat{G})$ can be invoked to rewrite (62). For large G, only the largest positive eigenvalue, $\eta_{\max}$, will be important, $e^{2(\mathbf{A}\cdot\hat{G})G} \sim e^{2G\eta_{\max}}\cdot O(1)$, and there follows from (62) the qualitative expression

$$\frac{d\hat{G}}{dt} \sim e^{-2G\eta_{\max}}\cdot\phi\hat{\Phi}\,(\mathbf{A}\cdot\hat{G}),$$

showing that, for large G, $d\hat{G}/dt$ is damped exponentially, i.e., after a certain time has been reached for which $G \gg 1$, $\hat{G}(t)$ is essentially constant. The unit vector $\hat{\Phi}(t)$, however, continues to oscillate wildly; and the result is that $\xi(t) = \hat{G}(t)\cdot\hat{\Phi}(t)$ oscillates to zero for sufficiently large t. For small t, $\hat{G}$ and $\hat{\Phi}$ are in phase, leading to $\xi(t) \sim 1$; but as time increases, they fall out of phase, and then in phase again, etc. Because G is supposed large and positive, $\phi(t)\xi(t)$ is more often positive than negative, of qualitative form given in Fig. 1. If this picture is correct, then $G(t)$ does not grow smoothly, but only in spurts, when $\xi(t)$ is reasonably positive. Since the onset of turbulence will be associated with $G \gg 1$, the behavior of $\xi(t)$ provides a possible mechanism for the fact of intermittency.[11] A detailed analysis depends upon a systematic resolution of (59). Finally, since $\overline{W} = \exp(\boldsymbol{\lambda}\cdot\hat{G}G)$, an eigenvalue analysis of $\boldsymbol{\lambda}\cdot\hat{G}$ will generate $\overline{W} \sim e^{G(t)}\cdot O(1)$, which is the only part of this discussion to be used below. In particular, for the nonunitary matrix $\overline{W}$,

$$\overline{W}^{\dagger}\overline{W} \sim e^{2G(t)}\cdot O(1). \tag{63}$$

To see how these SU(3) forms can be useful, we now insert them into a crude, dimensional model which examines the growth of enstrophy, $\Omega(t) = \frac{1}{2}\int d^3x\,\boldsymbol{\omega}^2(\mathbf{x},t)$. With (23a) and (19), Ω can be put into the form

$$\Omega = \tfrac{1}{2}\int d^3y\int d^3y' \times\delta(\mathbf{q}(t;\mathbf{y}) - \mathbf{q}(t;\mathbf{y}'))g^{\dagger}(\mathbf{y}') \cdot\overline{W}^{\dagger}(\mathbf{q}(t;\mathbf{y}'),t)\overline{W}(\mathbf{q}(t;\mathbf{y}),t)g(\mathbf{y}),$$

using an obvious matrix notation. It is simplest to change variables from $(\mathbf{y},\mathbf{y}')$ to $(\mathbf{q},\mathbf{q}')$, observing that the determinant of each transformation is unity, $d^3q = \det(\partial q)\cdot d^3y$; and with (42), $\det(\partial q) = \exp(\mathrm{tr}\ln\overline{U}) \Rightarrow 1$. Then, (64) can be written as

$$\Omega = \tfrac{1}{2}\int d^3q\,g^{\dagger}(y(q))\cdot\overline{W}^{\dagger}\,\overline{W}\cdot g(y(q)),$$

or with (63)

$$\Omega \sim \int d^3q\cdot g^{\dagger}\cdot e^{2G(t)}\cdot O(1)\cdot g. \tag{64}$$

We now invoke a crude, dimensional argument to determine the possible significance of the above assumptions, by transferring them to the behavior of a time-dependent length scale, $l(t)$. If only a single length scale is used, we assume we are studying the growth of "isotropic" turbulence, as represented by the growth of enstrophy.[12] What is fixed in the fluid is assumed to be its energy/density, $\kappa = E/\rho = \frac{1}{2}\int d^3x\,\mathbf{v}^2$, and we write the dimensional relation $\kappa \sim l^3v^2$, or $v \sim \kappa^{1/2}\cdot l^{-3/2}$. Then, $\Omega = \frac{1}{2}\int d^3x\,\boldsymbol{\omega}^2 \sim l^3(v/l)^2 \sim \kappa/l^2$. Comparing with the rhs of (64), the dimensions of $\int d^3q\,g^{\dagger}g$ are also $\sim L^{-2}$, but these coordinates must refer to the initial configuration of the velocity/vortex fields, since the latter are introduced by the source g at $t = 0$. Hence we write $\int d^3q\,g^{\dagger}g \sim \kappa/l_0^2$, where l_0 refers to a typical length scale at a

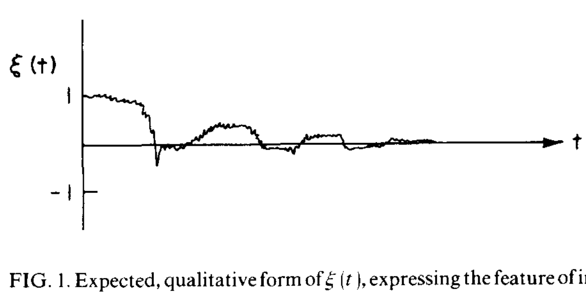


FIG. 1. Expected, qualitative form of $\xi(t)$, expressing the feature of intermittency.

pre-growth-of-turbulence time t_0, which for simplicity is set equal to zero, $t_0 = 0$. Finally, we estimate $G(t) = \int_0^t dt'\,\xi(t')\phi(t')$, using $\phi(t) \sim (v/l) \sim \kappa^{1/2}/l^{5/2}(t)$. In this way, a crude model of $l(t)$ can be extracted from (64),

$$(\kappa/l^2(t)) \sim (\kappa/l_0^2)\cdot\exp\left[2\kappa^{1/2}\int_0^t dt'\,\xi(t')l^{-5/2}(t')\right],$$

or

$$l(t) \sim l_0 \exp\left[-\kappa^{1/2}\int_0^t dt'\,\xi(t')\cdot l^{-5/2}(t')\right]. \tag{65}$$

The decrease of $l(t)$ as t increases presumably corresponds to the transfer of the original input energy into vortices of smaller and smaller spatial scale. To solve (65), it is only necessary to differentiate once and form the differential equation

$$\frac{dl}{dt} \sim -\kappa^{1/2}\xi(t)\cdot l^{-3/2}(t),$$

which can be integrated immediately,

$$l(t) = \left[l_0^{5/2} - \tfrac{5}{2}\kappa^{1/2}\int_0^t dt'\xi(t')\right]^{2/5},$$

generating

$$\Omega(t) \sim \left[l_0^{5/2} - \tfrac{5}{2}\kappa^{1/2}\int_0^t dt'\,\xi(t')\right]^{-4/5}. \tag{66}$$

It is clear that the type of growth of Ω, as t increases, now depends upon the falloff of $\xi(t)$, which we crudely model according to the following possibilities:

(i) If $\xi(t) \sim +t^{-(1-p)}$, $0 \leqslant p < 1$, then Ω has a finite-time singularity, $\Omega(t) \sim (t^* - t)^{-4/5}$.

(ii) if $\xi(t) \sim +t^{-(1+p)}$, $p > 0$, Ω may or may not have a finite-time singularity, depending on the specific constants involved. The singularity[13] of possibility (i) is less severe than that previously found in the quasinormal approximation[14] $\Omega \sim (t^* - t)^{-2}$, but it may well be that possibility (ii) is more accurate.

If it is initially assumed that there are two relevant length scales, rather than one, e.g., $L(t)$ corresponding to the length of typical vortex, and $l(t)$ corresponding to its radial dimension, an analysis similar to that leading to (66) yields

$$l^2(t) \simeq l_0^2 - 2\kappa^{1/2}\int_{t_0}^t dt'\,\xi(t')L^{-1/2}(t'),$$

and possibilities similar to (i) and (ii) can be developed. In any case, nonzero viscosity will smooth out an inviscid finite-time singularity.

V. SUMMARY

In this paper a formalism to study the properties of an inviscid NS fluid without rigid boundaries has been developed and applied to two classes of problems—relatively simple, exact examples in two dimensions—and to the extraction of a crude, dimensional model from an SU(3) analysis of vortex stretching in three dimensions. That the basic method is correct can be verified from these initial examples, but whether it will be useful in more complicated two- or three-dimensional problems is a different matter, which remains to be seen.

To the best of the author's knowledge the application of Fradkin's generic Green's function representation, and the use of an SU(3) description for the ensuing vortex stretching, have not previously appeared in the (voluminous!) NS literature; and it is hoped that these techniques will provide tools for calculations less crude than those presented in Sec. IV. One interesting feature of the SU(3) analysis of strong vortex stretching has been the automatic appearance of a certain degree of intermittency, characterized by the function $\xi(t)$ of (60), which may turn out to be a useful way of describing the properties of experimental, intermittent turbulence. This tentative identification of $\xi(t)$ with intermittency is here only suggested, rather than claimed; but it is, perhaps, a suggestion which may turn out to be at least partially true.

The formalism itself suggests various other calculational attempts, such as the approximate solution of (25) when vortex stretching is represented by $\partial_a q_b$, or the behavior of these estimates when vortices are continually fed into the fluid, with the aid of a source $\mathbf{g}(\mathbf{y},y_0)$ more general than the $\mathbf{G}(\mathbf{y})\cdot\delta(y_0)$ used above. Perhaps the most interesting modification would be the inclusion of viscosity corrections to all the calculations of this paper. A somewhat different resolution of Fradkin's original representation, not tied to the extremum calculation of Sec. II and valid for $\nu \neq 0$, will be presented separately.

It should be mentioned that hardly any methods exist to treat the nonabelian, vortex-stretching problem, when perturbation in $(\boldsymbol{\omega}\cdot\nabla)\mathbf{v}$ is improper, other than the strong-coupling approach of Sec. IV. Other possibilities are to treat the dimension D as very large, and search for simplifications in the large-D limit, in analogy with current work on the large-N limit of certain[15] quantum field theories; or for fixed D to replace the matrices λ_i, of the defining representation of SU(D), by semiclassical coordinates corresponding to higher-dimensional representation of SU(D). For example, for SU(3), the hypercharge and isospin quantum numbers present in this formalism could be treated as if they were continuous coordinates, following an approximation technique long known in nuclear and particle physics. But these are just stop-gap measures, which really do not get to the heart of the problem of how to find a useful, nonformal representation for the ordered bracket of (14).

The appearance of such SU(D) coordinates suggests that there may be an underlying SU(D) symmetry of the basic velocity/vorticity NS equations. Clearly, an analogy exists between certain aspects of hydrodynamics and nonabelian field theory; perhaps an analogy may also be drawn between the topological structure of bent, closed, or knotted hydrodynamical vortex tubes, in a semiturbulent situation, and the Copenhagen ("spaghetti") vacuum of interwoven flux tubes of QCD.[16] We do not know the answer to this, but point out that, among others, it is an interesting question to ask.

ACKNOWLEDGMENTS

It is a pleasure to thank colleagues already well versed in matters hydrodynamical for their encouragement and patient explanations of many topics quite obvious to them, if not to the novices who received and were enlightened by their remarks. In particular, we warmly thank H. Aref and

S. Libby of Brown and U. Frisch and J.-D. Fournier of the Observatory of Nice.

APPENDIX

All the material of this appendix has been taken from Ref. 4. The eight, traceless SU(3) matrices λ_i satisfy the relations

$$\mathrm{Tr}[\lambda_i \lambda_j] = 2\delta_{ij},$$

$$[\lambda_i, \lambda_j] = 2if_{ijk}\lambda_k,$$

and

$$\{\lambda_i, \lambda_j\} = \tfrac{4}{3}\delta_{ij} + 2d_{ijk}\lambda_k.$$

An explicit representation of the λ_i may be written as

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_2 = i\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

$$\lambda_5 = i\begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\lambda_7 = i\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \lambda_8 = \begin{pmatrix} 1/\sqrt{3} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & -2/\sqrt{3} \end{pmatrix}.$$

Nonzero elements of f_{ijk} and d_{ijk} are given below. The f_{ijk} are odd under permutation of any two indices, while the d_{ijk} are even.

ijk	f_{ijk}	ijk	d_{ijk}
123	1	118	$1/\sqrt{3}$
147	1/2	146	1/2
156	$-1/2$	157	1/2
246	1/2	228	$1/\sqrt{3}$
257	1/2	247	$-1/2$
345	1/2	256	1/2
367	$-1/2$	338	$1/\sqrt{3}$
458	$\sqrt{3}/2$	344	1/2
678	$\sqrt{3}/2$	355	1/2
		366	$-1/2$
		377	$-1/2$
		448	$-1/(2\sqrt{3})$
		558	$-1/(2\sqrt{3})$
		668	$-1/(2\sqrt{3})$
		778	$-1/(2\sqrt{3})$
		888	$-1/\sqrt{3}$

[1]H. Aref and N. Pomphrey, Phys. Lett. A **78**, 297 (1980), and other papers quoted therein.

[2]E. S. Fradkin, Nucl. Phys. **76**, 588 (1966); E. S. Fradkin, U. Esposito, and S. Termini, Nuovo Cimento, Ser. I, **2**, 498 (1970).

[3]See, for example, G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge U. P., Cambridge, 1967), Chap. 5.

[4]It should be noted that only the symmetric part of $\partial_b v_a$ can enter into (4), or (2), as long as $\boldsymbol{\omega} = \nabla \times \mathbf{v}$, although for simplicity and possible application to other problems, we keep to the SU(3) forms of the text. But when approximations are made, one must be careful to consider quantities manifestly independent of the vorticity components $\psi_{2,5,7}$ of (43), such as Ω-growth estimates of Sec. IV, which really depend on the strain $\boldsymbol{\phi}$ only. The quantity $\xi(t)$ of (60) then becomes $\hat{G}\cdot\hat{\phi}$, where $\boldsymbol{\Phi}$ is replaced by $\boldsymbol{\phi}$.

[5]A discussion of ordered exponentials can be found in any text on quantum field theory or many-body physics. One which contains, in addition, some useful functional methods is H. M. Fried, *Functional Methods and Models in Quantum Field Theory* (M.I.T. Press, Cambridge, MA, 1972).

[6]M. Gell-Mann and Y.Ne'eman, *The Eight-fold Way* (Benjamin, New York, 1964).

[7]See, e.g., P. Germain in *Les Houches Summer School Lectures*, edited by R. Balian and J.-L. Peube (Gordon and Breach, Paris, 1973).

[8]See, e.g., the discussion given by U. Frisch and R. Morf, Phys. Rev. A **23**, 2673 (1981), and other references quoted therein.

[9]These SU(3) techniques were used by H. M. Fried, Proceedings of the U.S.–France Scientific Seminar "Theoretical Aspects of QCD", Marseille, 1981; similar methods were used in an SU(2) context by H. F. Fried, Phys. Rev. D **16**, 1916 (1977). Of course, very many different methods have been used to approximate the effects of vortex stretching and the possible development of finite-time singularities, e.g., U. Frisch, in *Les Houches Summer School Lectures* (Gordon and Breach, Paris, 1981), and many references quoted therein.

[10]More precisely, G may have an imaginary component, $G \to G + iH$, where the components of G and H are real, and they multiply the real and imaginary Gell-Mann matrices, respectively. All the analysis which follows (62) refers only to the real part of the eigenvalues of $\boldsymbol{\lambda}\cdot(\mathbf{G} + i\mathbf{H})$, for that is the quantity entering into the crude estimates of enstrophy following (64).

[11]Starting from a scalar equation, similar to (44) with $\boldsymbol{\omega} = 0$, R. H. Kraichnan, [J. Fluid Mech. **64**, 737 (1974)] has extracted special forms of intermittency. It is, at present, an open and interesting question if anything as solid can be extracted from (59) for the NS problem.

[12]This is a big and, to our knowledge, unjustified assumption, which attempts to link the deterministic NS properties under discussion with quantities more properly defined in terms of statistical averages. By using only a single length scale, we envisage a situation in which certain regions of the fluid are locally dense with tightly coiled vortex rings; as the rings stretch, their radii decrease and they immediately coil and twist and could, as the analysis suggests, possibly fit into a smaller spatial region, of smaller length scale and higher vortex density. But there is no reason to treat this crude argument as anything more than an interesting perhaps oversimplified exercise in dimensional analysis. A survey of other, different approaches to the growth of enstrophy can be found in S. Orszag, in Ref. 7.

[13]$l(t) \sim (t^* - t)^{2/5}$ and $v(t) \sim (t^* - t)^{-6/5}$ produce a result with the same exponents as those found in the analysis of energy decay by P. G. Saffman, Phys. Fluids **10**, 1349 (1967).

[14]I. Proudman and W. H. Reid, Philos. Trans. R. Soc. London Ser. A **247**, 163 (1954). If it is assumed that energy *density* is conserved, then enstrophy density $\frac{1}{2}\omega^2 \sim [l(t)]^{-2}$, and the techniques leading to (66) will instead produce $\omega^2 \sim [l_0 - c\int_0^t dt'\,\xi(t')]^{-2}$, where c is a constant. If the function used to model $\xi(t)$ satisfies possibility (i), then $\omega^2 \sim (t^* - t)^{-2}$, which is the same as that found in the analysis of this reference. More recently, a similar growth of vorticity has been found to follow from a model of the NS equations by P. Vieillefosse, J. Phys. (Paris) **43**, 837 (1982). Other work on this subject may be found in J. Léorat, Thèse de Doctorat, Université de Paris -VII, 1975.

[15]G. 't Hooft, Nucl. Phys. B **72**, 461 (1974). A very modern reference is C-I. Tan, Zheng Xi-Te, and Chen Tian-lun, "Reduced Large N Lattice Theory Without Spontaneous Symmetry Breakdown," Phys. Letters (to be published).

[16]H. B. Nielsen in *Proceedings of the 3rd Adriatic Summer Meeting on Particle Physics*, Dubrovnik, 1980, edited by I. Andric, I. Padic, and N. Zorko (North-Holland, Amsterdam (1980). The original paper is H. B. Neilson and P. Olesen, Nucl. Phys. B **160**, 380 (1979).